Omid Noroozi
Bram De Wever *Editors*

The Power of Peer Learning

Fostering Students' Learning Processes and Outcomes

Springer

Social Interaction in Learning and Development

Studying social interaction in human mind and activities is highly relevant for different epistemological and theoretical approaches (e.g., individual constructivism, social constructivism, dialogical approach). Consequently, there is a growing number of social interaction studies in various contexts (family, educational, professional, clinical, institutional, social, political, and cultural settings) which are based on different theoretical perspectives and methodological approaches. This produces a multiplicity of findings which are highly relevant, both theoretically and practically—although weakly interrelated and seldom discussed together. The main aim of this book series is to create a space for continuous and systematic critical reflection of social interaction studies and their integration with a special focus on: (1) a detailed account of actors and processes involved in different types of situated social interaction, (2) situatedness of social interaction within sociocultural and sociomaterial contexts and how social interaction and contexts constitute and transform each other; (3) how properly designed social interactions can provide opportunities for learning and development (in formal, informal, non-formal education), and (4) how the individual person navigates within these social interactions.

The book series aims to support an argumentative and productive dialogue among different theoretical and methodological traditions, in order to enable a better understanding of their strengths and weaknesses.

For more information on how to submit your proposal, please contact the publisher: Marianna.Georgouli@springer.com

Omid Noroozi · Bram De Wever
Editors

The Power of Peer Learning

Fostering Students' Learning Processes and Outcomes

Springer

Editors
Omid Noroozi 🆔
Education and Learning Sciences
Wageningen University and Research
Wageningen, The Netherlands

Bram De Wever 🆔
Department of Educational Studies
Ghent University
Ghent, Belgium

Series Editors' Preface

Peer Learning as a Powerful Tool for Feedback and Assessment Between Students

The present book series on *Social Interaction in Learning and Development* has been established as a space for continuous and systematic critical reflection of theories and studies focusing on co-constructing learning and development throughout the process of interaction with others. As we consider that studying Social Interaction in Learning and Development (as well as how it might constitute human mind and activities) is highly relevant for different epistemological and theoretical approaches (e.g., individual constructivism, social constructivism, dialogical approaches), we recognize the relevance of a growing number of different studies on social interaction. Research in various contexts (family, educational settings, professional fields, clinical, institutional, social, political, and multicultural situations), based on different theoretical perspectives and methodological approaches, produced a multiplicity of perspectives and findings which are highly relevant for various theoretical and practical reasons. The diversity of available studies and findings makes a step further in the process of reflection and integration of different challenges, exactly because it creates a good opportunity for a deeper understanding of how social interaction and individual learning and development are interwoven.

By editing this book series, we are convinced that included volumes might serve as a meeting point of various perspectives on studying social interaction in learning and development. As one of our goals, we intend to propose the book series as a platform to support dialogical reflection of controversies and issues related to theories, research methods, findings, and practical applications related to the research on social interactions and learning.

The present volume is part of the book series because it makes a highly original contribution to the research field in educational psychology that bears on group learning in general, and more specifically to the experimental study of approaches to peer feedback, critique, and appraisal. In our opinion, the volume brings together theory, methodology, tools, and empirical evidence about peer learning in higher education. It helps the readers to grasp the cutting-edge developments in the field and is presented as a compendium of high-level research that does not yet exist.

For this reason, the specific attention on peer assessment/feedback as an educational approach in the field of group work, allied to a comprehensive coverage of methodological, experimental, educational, and technological aspects, is a valuable resource to transfer some of the work of evaluating students' production to the students themselves. The volume constitutes an effective possibility to exploit students' abilities and to figure out how students manage to explaining to each other in a powerful way. In this sense, this volume can be considered as a valuable resource not only for researchers in the field of educational psychology and for educators, but also for academics from diverse disciplines dealing with assessment and peer learning.

Biel/Bienne, Switzerland   Francesco Arcidiacono
Belgrade, Serbia   Aleksandar Baucal

Preface

This book deals with how peer learning processes in the broad sense (including written or oral peer feedback, peer assessment, peer review, peer interaction, peer dialogue, etc.) can stimulate learning processes and outcomes in educational settings. The aim of this book is to report the latest cutting-edge research in the field of peer learning. The orientation of the book, beyond the theoretical aspects, is empirical and practical. Special emphasis is given to integrate theory, research, and practice while making a clear link between educational technology, learning sciences, educational psychology, computer sciences, and learning analytics with peer learning processes and outcomes.

The context of most of the contributions in this book is related to settings in higher education, though some contributions focus on secondary education and much of the studied practices are transferable to other educational contexts. The lion's share of the chapters studies the core practices of getting students to interact with each other regarding their work and learning activities. This interaction can be labeled peer assessment or peer feedback, depending on the level and nature of the evaluative processes, or even broader. The idea of having students to interact and collaborate as part of their learning is built on broader ideas in the field of the learning sciences and learning and instruction. The general idea undergirding collaborative activities is that learning is not just a matter of transferring knowledge, but of having students engaged in cognitive activities to actively construct their knowledge. These ideas are grounded in cognitive constructivist developmental theories as well as more socio-cognitive and social constructivist theories in which interaction plays a central role. More specifically, within the field of computer-supported collaborative learning, attention has been paid to specific cognitive activities that are triggered during interaction, such as negotiating, arguing, asking questions, and providing feedback. This book holds a compilation of studies focusing on different aspects of peer interaction and is divided into four parts including conceptual, technological, methodological, and empirical contributions.

Part I, consisting of three chapters, covers conceptual aspects of peer learning. The first part of the book begins with the chapter by *Bhavani Sridharan, Jade McKay, and David Boud*, who offer a four-pillar framework for peer assessment for collaborative teamwork in higher education. This framework includes veracity, validity, volume, and literacy with overlapping and specific features. While

the veracity pillar and the validity pillars deal with the assessment design and the implementation considerations, respectively, the volume pillar is linked with the technology factors, and the literacy pillar is associated with the roles and responsibilities for peer assessment. This framework can be helpful for educators, policymakers, and scholars to overcome the challenges of peer assessment in the collaborative teamwork context. The second chapter of this part by *Kamila Misiejuk and Barbara Wasson* uses a scoping review to provide an overview of the role of learning analytics in understanding peer assessment. The authors systematically review relevant papers to elaborate on how to incorporate automated assessment and visualizations into peer assessment, how to apply data analysis methods to peer assessment, and how to evaluate different types of peer assessment. This review can serve as a helpful resource to find out how learning analytics can be used effectively to facilitate peer assessment. In the last chapter of this part, *Ya Ping (Amy) Hsiao and Kamakshi Rajagopal* discuss that for their undergraduate thesis, students are expected to co-supervise each other as a supplemental strategy for supervisor feedback. This study focuses on supporting feedback receivers through training materials and instructional activities offered by teachers. The authors state that the suggested and designed instructional activities can be used for improving students' multiple peer feedback performance. This review study contributes to the advancement of the current literature on peer feedback and how one can effectively use instruction and training in this regard.

Part II, consisting of three chapters, focuses on the methodological aspects of peer learning. In the first paper of the second part, *Tine van Daal, Mike Snajder, Kris Nijs, and Hanna Van Dyck* compare the effects of two assessment methods (namely comparative judgment and criteria list) on students' problem-solving in physics, writing, and performance. Results showed some differences between peer assessment conditions regarding the quantity and the content of the peer feedback; however, the peer assessment method did not impact students' performance. Results further showed that students in the comparative judgment condition gave more positive feedback on the syntax of the texts, while students in the criteria condition provided more positive feedback on the aspect of interpretation. In the next study, *Jasperina Brouwer and Carlos A. de Matos Fernandes* use stochastic actor-oriented models (SAOMs) to explain collaboration intentionality (CI) as a prerequisite for peer feedback and learning in networks. The chapter authors state that the model shows a homophily influence. This means that students favorably seek feedback from others who are similar in collaboration intentionality. Students who seek feedback from one another become more similar in terms of collaboration intentionality over time, and this similarity is driven by selection and influence mechanisms in peer feedback networks. In the last chapter of the second part, *Kyriaki Vakkou, Tasos Hovardas, Nikoletta Xenofontos, and Zacharias C. Zacharia* compare expert and peer assessment of pedagogical design in integrated Science, Technology, Engineering, Arts, and Mathematics education. Although the significant correlations are computed for global measures of validity (correlations between total scores of expert and peer assessors) and reliability (correlations between total scores of different peer assessors for the same pedagogical scenario),

the assessment criteria for which peer assessment failed to be valid and/or reliable should be considered carefully in future training sessions. Also, indications of participant preference of expert feedback over peer feedback exist in the study so peer assessors can give their feedback to peers at least once to rationalize their quantitative scores in each assessment criterion. The findings of this part can be useful for teachers who regularly provide learning opportunities for their students to engage in peer learning strategies.

Part III, consisting of four chapters, discusses the technological developments for the effective design of peer learning. In the first chapter of the third part, *Stan van Ginkel and Bo Sichterman* discuss how virtual reality (VR) can be best designed for constructing computer-mediated feedback for enhancing students' presentation skills. The authors discuss two recent VR experiments in presentation research that can be used to effectively construct feedback messages in VR for improving peer learning presentation. Recommendations and implications are provided for future studies on computer-mediated feedback for peer learning in presentation research. In the next chapter, *José Carlos G. Ocampo and Ernesto Panadero* examine web-based peer assessment platforms based on their characteristics and features that can potentially affect student learning, feedback exchange, and social interaction. The authors use nine peer assessment design elements and state that majority of the platforms offer features to facilitate peer assessment in different disciplines and in numerous ways and has the potential to affect student learning, feedback, and social interaction. The authors suggest that extensive training is needed for teachers and students to integrate the features provided by these platforms into educational contexts. In the following chapter, *Sebastian Strauß and Nikol Rummel* discuss that research on group awareness tools has not presented a comprehensive framework about the features underlying their effectiveness. Thus, they examine potential boundary conditions to find out whether groups take up the information from group awareness tools and transform it into actions that adjust the current ways of interaction to the group. The authors suggest that future research should focus more on the design of group awareness tools, on processes that are necessary to support group-level feedback, and on effective regulation of collaboration. In the last chapter of the third part, *Ellen Rusman, Rob Nadolski, and Kevin Ackermans* argue that text-based analytic rubrics offer limited capacity to convey contextualized, procedural, time-related, and observable behavioral aspects of a complex skill so they limit the construction of a rich mental model. Instead, they state that using video-enhanced rubrics followed by a technology-enhanced formative assessment method may produce a richer mental model, higher feedback quality, and more positive growth in three generic complex skills, namely presenting, collaborating, and information literacy. Hence, they suggest using the Viewbrics technology-enhanced formative assessment method with video-enhanced rubrics for developing skills' mastery levels of students.

Part IV, consisting of seven chapters, the largest part, presenting cutting-edge empirical research studies in the field of peer learning, begins with PeerTeach by *Soren Rosier*. This chapter has an experimental design to investigate PeerTeach online training for tutors to support students' implementation of learner-centered

teaching methods. The study results show that the trsaining increases the frequency of students using learner-centered teaching methods in both online and real-life tutoring scenarios. This suggests that training students to use learner-centered tutoring strategies can greatly improve the efficacy for peer tutoring in classrooms, and that technological solutions can scale this type of training. In the second chapter, *Julia Kasch, Peter van Rosmalen, and Marco Kalz* use a thematic analysis as part of an exploratory sequential mixed methods research design to explore personal factors that affect students' peer feedback orientation. The most important personal factors influencing their peer feedback orientation are found as the perceived usefulness of receiving and providing peer feedback, the social bond between students, fairness, and skills. This chapter offers a new conceptualization of peer feedback orientation and contributes to the theory development for peer feedback orientation. In the third chapter, *Natasha Dmoshinskaia and Hannie Gijlers* report an overview of the results of four (quasi-) experimental studies with secondary school students who give feedback on a small-scale product (concept map) in an online inquiry-learning environment. The authors put forward that giving feedback to peers can be a learning experience for a feedback provider. Also, such learning may allow students not only to be cognitively involved with the material, but also to be involved at a meta-level since evaluating others' products and providing appropriate feedback may require higher-order thinking. The authors suggest that since online platforms may provide flexibility during the feedback-giving process based on the learning goals, they can be helpful for giving feedback more natural and easier than traditional instruction. In the fourth study, *Emmeline Byl, Keith J. Topping, Katrien Struyven, and Nadine Engels* explore how different peer interaction types impact students' social and academic integration and institutional attachment. They collected both quantitative and qualitative data from undergraduate students in Psychology and Education Sciences through online surveys and interviews. The authors claim that peer mentoring is the most effective means to enhance social integration. While, for academic integration, peer tutoring is an effective peer interaction tool. In terms of institutional attachment, neither peer mentoring nor peer tutoring are found to be effective. Fifth, *Morgane Senden, Dominique De Jaeger, Tijs Rotsaert, Fréderic Leroy, and Liesje Coertjens* focus on designing an online training that creates a psychologically safe and trustworthy environment for peer feedback activities in a higher education context. The suggested training includes five stages as follows: (a) discovery of students' representations, (b) lecture on how to provide effective feedback, (c) peer feedback practice, (d) role-play and discussion in small groups, and (e) summary of key learning points. The authors use a questionnaire to explore students' perceptions of the training. The authors state that students' general impression of the training is positive and thus recommend to use such design for peer feedback settings. In the sixth chapter, *Nafiseh Taghizadeh Kerman, Seyyed Kazem Banihashem, and Omid Noroozi* explore the relationship among students' attitude toward peer feedback, peer feedback performance, and peer feedback uptake in the context of argumentative essay writing within an online learning environment. This exploratory study is built on an online module with three main tasks including an original essay, peer

feedback, and a revised essay. The authors report that if students perceive peer feedback as useful, they are willing to take it up. They also find that the quality of received peer feedback is related to perceived fairness and trustworthiness of peer feedback. If the received peer feedback entails justifications of the identified problems and suggestions for further improvements, students are more willing to perceive peer feedback as a useful, fair, and trustworthy to accept. The authors suggest that these findings can guide teachers in better adoption of peer feedback activities for essay writing. In the last chapter of the book, *Laura Ketonen, Pasi Nieminen, and Markus Hähkiöniemi* use a case study method to find out how lower-secondary science students exercise agency during formative peer assessment. The authors categorize agency in nine forms: initiating, echoing, judging work, avoiding criticism, seeking help, appraising feedback, rejecting feedback, revising work, and avoiding revision. They also identify it in three roles, namely group member, assessor, and assessee. The researchers suggest that peer assessment does not challenge students equally, so their agency needs to be reinforced to make them productive during the peer assessment process.

In sum, this book presents peer learning research studies that involve learning and interaction processes and outcomes that involve some kind of feedback activity and exchange between peers. Most of the studies are focusing on the core processes of feedback and assessment between students; however, the collection goes beyond peer feedback and peer assessment and also discusses broader issues such as peer collaboration, peer dialogue, and peer interaction. The full range of peer learning activities is tackled through conceptual, technological, methodological, and empirical contributions on how to best design effective peer learning in real educational settings. We hope this book will inspire further research and development in the field of peer learning.

Wageningen, The Netherlands   Omid Noroozi
Ghent, Belgium   Bram De Wever

Contents

About the Editors

Omid Noroozi (Ph.D., 2013) is Associate Professor in the Education and Learning Sciences group at Wageningen University and Research, the Netherlands. He has fostered an interest in understanding the relations among technology, pedagogy, and learning higher-order skills (e.g., critical thinking, reasoning, problem-solving, communication, collaboration, self-regulation, and entrepreneurial thinking) with a specific focus on students' argumentation competence development in higher education.

Omid has served as an editorial board member of various international peer-reviewed journals and also edited several special issues for JCR-indexed journals as guest editor. Omid is now serving as an executive board member of the International Society for Technology, Education, and Science (ICTES), Editor-in-Chief for *International Journal of Technology in Education (IJTE)*, and President and Scientific Chair of International Conference on Studies in Education and Social Science (ICSES). He has been a visiting scholar at University of Michigan, USA, University of Oulu, Finland, and Tarbiat Modares University, Iran.

Bram De Wever (Ph.D., 2006) is Associate Professor at the Department of Educational Studies at Ghent University, Belgium, and head of the research group TECO-LAB at that department. His research is focusing on technology-enhanced learning and instruction, peer assessment and feedback, computer-supported collaborative learning activities, inquiry learning, and argumentative writing. Research settings include mostly secondary, higher, and adult education.

Bram is currently Associate Editor of *Journal of the Learning Sciences* and the supervisor of the Flemish Research team of PIAAC (Programme for the International Assessment of Adult Competencies; OECD). He is serving on the board of *Learning and Instruction*, *Computers and Education*, and the *International journal of Computer-Supported Collaborative Learning*. He is active in the European Association for Research on Learning and Instruction (EARLI) and the International Society of the Learning Sciences (ISLS), where he respectively has been a SIG coordinator and program chair.

Part I

Conceptual Contributions on Peer Learning

The Four Pillars of Peer Assessment for Collaborative Teamwork in Higher Education

Bhavani Sridharan, Jade McKay, and David Boud

1.1   Introduction

Peer learning, in the form of various collaborative learning models, has become a dominant approach in higher education to foster learning, engagement, and development of well-rounded graduates. Peer learning refers to "the acquisition of knowledge and skills through active helping and supporting among status equals or matched companions" (Topping, 2005, p. 631). The popularity of peer learning is evident from the extant literature surrounding the adoption of a repertoire of nuanced strategies including peer mentoring, teaching, coaching, review, assessment and feedback, study-buddy support, team-based learning, collaborative learning, cooperative learning, reciprocal peer learning, amongst others (Boud et al., 2014).

Nevertheless, the challenges surrounding peer learning strategies, particularly those entailing formal assessment, are problematic and complex since assessment is pivotal to the success of higher education systems (Strijbos & Sluijsmans, 2010). Students are very sensitive to assessment strategies, affecting emotional well-being (Jones et al., 2021), learning experiences, satisfaction and learning outcomes (Li et al., 2020). Additionally, wide variation in peer learning practices and ambiguities

B. Sridharan (✉) · J. McKay
Faculty of Law and Business, Australian Catholic University, Melbourne, Australia
e-mail: Bhavani.Sridharan@acu.edu.au

D. Boud
Centre for Research in Assessment and Digital Learning, Deakin University, Geelong, Australia

Faculty of Arts and Social Sciences, University of Technology Sydney, Sydney, Australia

Work and Learning Research Centre, Middlesex University, London, UK

surrounding its effect on learning outcomes adds to implementation difficulties (Panadero, 2016).

In this context, peer learning models that combine peer assessment and peer feedback in collaborative teamwork (CTW) contexts embracing formal assessment methods provide a mechanism to fulfill a myriad of social, professional and educational goals (Planas-Lladó et al., 2021). Peer assessment refers to grading of peers while peer feedback entails giving, receiving and using qualitative comments by peers to support learning (Hoo et al., 2021). For the purposes of this chapter, peer assessment subsumes both peer rating and peer feedback. CTW is a structured form of collaborative learning requiring members to work together in small groups to achieve a common goal.

This combination cannot only strengthen the holistic development of knowledge, skills and abilities sought by students, employers and accrediting bodies (Planas-Lladó et al., 2021) but may also compensate for inherent limitations of individual strategies (Li et al., 2020). Peer assessment can influence the product quality from CTW tasks through leveraging individual accountability (Jacobs & Renandya, 2019), interdependent behaviour and strengthening learning (Planas-Lladó et al., 2021).

This chapter focuses on the peer assessment of process in producing a tangible artifact in both the formative and summative context. In CTW, this approach has been identified as more appropriate, as students are best positioned to assess their peers' behaviours and dispositions owing to the proximal working relationship with team members (Sridharan et al., 2019). Nevertheless, this approach faces distinct challenges such as marking bias, implementation difficulties, engagement issues, quality and usability of feedback, trust issues and others (Oakley et al., 2004).

These challenges point to the need for an effective peer learning model to have impactful outcomes. Yet, studies exploring such an arrangement in CTW are sparse. Panadero (2016) stresses the need for considering social and human factors on peer assessment research as it generates psychological and emotional reactions. Scholars have identified gaps between theory and practice, and superficial implementation of CTW (Lawlor et al., 2018). Moreover, existing models predominantly focus on peer assessment in a cognitive context and therefore its direct and nuanced applicability to CTW is limited (Adachi et al., 2018; Gielen et al., 2011; Topping, 1998). To this end, we propose a framework specifically focussing on CTW and orienting it to specific peer assessment challenges and resolutions.

In this chapter, we set the scene by establishing the key impediments of CTW and peer assessment as the potential solutions to the impediments based on existing studies. This is followed by distilling the range of peer assessment challenges articulated in the existing literature to determine key themes. Next, adopting a systematic approach to develop pragmatic solutions to overcome peer assessment challenges, we propose a four-pillar framework. Finally, we draw upon the findings to summarise the implications, practical recommendations and limitations of the framework.

1.2   Impediments and Solutions for CTW

Recognising the intertwined landscape of CTW and peer assessment, holistic understanding of CTW impediments is fundamental, without which solutions to peer assessment challenges may become ineffective. Several impediments to effectively transforming CTW are evident despite the growing adoption of group work in the higher education curriculum (Rubin & Dierdorff, 2009). Impediments affecting student satisfaction and experience arise from tensions surrounding cognitive, affective and behavioural dimensions (Salas et al., 2015).

1.2.1   Cognitive, Affective and Behavioural Impediments

Prior literature reveals an array of cognitive impediments in CTW around poor adoption of pedagogical approaches (Hansen, 2006; Marasi, 2019). Asking students to work in groups without adequately building teamwork skills will not guarantee desired outcomes (McKendall, 2000; Opdecam & Everaert, 2018). Oakley et al. (2007, p. 270) contend, "students are not born knowing how to work effectively in teams" and underscore the poor instructional model as a root cause of student dissatisfaction. Likewise, Loughry et al. (2014) claim poor peer learning experiences due to the teacher's adoption of a 'sink or swim' approach and lack of engagement or support, particularly during times of conflict (Moore & Hampton, 2015a). The potential harmful effects of CTW on learning can surface without instructor guidance, accountability processes and value propositions for students (Oosthuizen et al., 2021).

Impediments stemming from affective dimensions include lack of psychological safety (Salas et al., 2018), unfair grading (Stover & Holland, 2018), and lack of trust and conflict issues (O'Neill & Mclarnon, 2018). Salas et al. (2018) posit 'the license to speak up' is a critical factor to deter worries of being judged and ridiculed by team members. Student frustration and negative attitudes towards teamwork surface when all members get the same reward irrespective of their contribution or non-contribution (Mihelič & Culiberg, 2019). Lack of trust and conflict can also lead to knowledge hoarding, non-cooperation and conflict issues (Banihashem et al., 2012; Latifi & Noroozi, 2021; Latifi et al., 2021; Taghizadeh et al., 2022).

Behavioural impediments contributing to student dissatisfaction and negative attitudes towards CTW (El Massah, 2018) include free riding and social loafing (Oakley et al., 2004); lone wolf or silo working tendencies (Opdecam & Everaert, 2018); and dominant or inactive and uncooperative tendencies (Planas-Lladó et al., 2021). It is important to recognise the underlying causes of such behaviours to overcome these impediments. For example, non-contribution could arise from 'imposter syndrome' (doubting one's abilities) (Chapman, 2017), fear of criticism or the fear of becoming a 'sucker' (Sridharan et al., 2019). On the other hand, over or under-valuing one's own contribution can occur owing to the 'Dunning-Kruger'

effect (cognitive bias in estimation) (Schlösser et al., 2013) or inherent competitive tendency of individuals creating an imbalance in individual contributions.

1.2.2 Strategies to Overcome CTW Impediments

Scholars have proposed a range of strategies to address CTW impediments. To tackle the cognitive impediments, effectively considering pedagogical approaches to curriculum design covering training, task design and facilitating environment is imperative. Key learning and teaching strategies supporting CTW training include highlighting the importance and relevance of CTW; and embedding team build-ing activities; and team debriefing exercises (Hansen, 2006; McKendall, 2000). Critical task design strategies require assessment design that demands teamwork (work in collaboration) as opposed to group work (work independently) (Riley & Ward, 2017); application-based tasks; incentives to quality individual contributions (Bravo et al., 2019) and other context specific parameters such as cohort type, year level, task complexity and intended learning outcomes (Bravo et al., 2019). The provision of tools to collaborate and communicate can also foster a cohesive teamwork culture (Oosthuizen et al., 2021).

Mitigating the affective impediments, providing a conducive and psychologi-cally safe environment enabling open and honest communication is critical (Salas et al., 2018) to develop trust, resolve conflicts, and enhance performance (Fra-zier et al., 2017). Defining roles and responsibilities and setting ground rules and expectations can help shape a unified team ethos (Bell et al., 2018). Additionally, dynamic team configuration considering both similar traits (values, attitudes and abilities) and dissimilar (complementary skills) individual characteristics (Oakley et al., 2004) can pave the way for creating a cohesive environment.

Combating the behavioral impediments, peer assessment has the power to pre-vent unacceptable student behaviours, particularly when direct observation by instructors is not feasible (Sridharan et al., 2019). Peer assessment can enhance learning to address underlying causes of such behaviours through assessees receiv-ing feedback to take corrective actions, and assessors developing self-awareness, self-regulated learning and evaluative judgement capabilities (Dochy et al., 1999).

Nevertheless, prior research has identified limitations of peer assessment includ-ing variability (Willey & Gardner, 2009), student resistance (Topping, 2005), lack of honesty (Panadero et al., 2013), reliability and validity (Falchikov & Goldfinch, 2000), poor understanding and lack of knowledge and skills (Sridharan & Boud, 2019; Winstone et al., 2019) and lack of mutual respect (Zhou et al., 2020). While other studies posit various solutions to these challenges, they rarely attempt to address the broad scope of nuanced challenges relating to peer assessment in the CTW context.

1.3 Peer Assessment Challenges in CTW Context

Exploring the existing literature and evidence base, several peer assessment challenges have been identified. These are logically classified into four thematic clusters: quality and standards; validity and reliability; scalability and sustainability; and literacy.

1.3.1 Quality and Standards

Peers' capabilities, behaviours and attitudes in accurate, honest judgment of each other and genuine engagement are critical for guaranteeing the quality and standards of peer assessment, without which it is wasted effort and resources. However, prior studies indicate a number of challenges impacting accuracy, honesty, engagement and overall trustworthiness of peer marking (Sridharan et al., 2019). In terms of capability, evaluative judgements and providing effective and usable feedback to others are complex and must be learned (Boud et al., 2018). Behavioural concerns include: incentives to mismark (competition); giving low marks to high performing students; over-generous marking (particularly friends); sabotage (overrating self and underrating peers) to create self-advantage; collusion with a tendency to mark similarly to others (Sridharan et al., 2019). Moreover, psychological safety factors such as fear of disapproval, social pressure and discomfort in marking peers can negatively impact honest assessment of peers (Vanderhoven et al., 2015). This is even more problematic when the peer assessment process is not anonymous leading to assessees preconceived perceptions of the assessor and unwillingness to open disclosure of behavioural issues (Anson & Goodman, 2014). Attitude challenges include non-engagement or untruthful engagement with the peer assessment activity, particularly in the formative context (either non-completion or random or insincere completion) (Sridharan & Boud, 2019).

1.3.2 Validity and Reliability

Validity and reliability are central to enhancing peer assessment effectiveness. Validity refers to use of an accurate unbiased relevant and aligned instrument to gain process and stakeholder acceptance (Speyer et al., 2011). Reliability requires consistency in marking (avoidance of arbitrary marking and absence of measurement error) irrespective of who does the peer assessment. Factors affecting reliability include biased marking as a result of friendship, vindictiveness, reciprocity, poor understanding of quality and standards, amongst others (Sridharan et al., 2019). Reliability can be enhanced through adoption of effective calibration and moderation practices, however, it requires effort, time and positive disposition by stakeholders. Other challenges include thoughtful consideration of peer assessment design decisions surrounding: sufficient number of peer assessors, incentives

for taking it seriously, and anonymity to encourage honesty to ensure students trust in the system (Freeman & McKenzie, 2002).

1.3.3 Scalability and Sustainability

Scalable and sustainable practices through embedding formative and summative assessment with multiple exposures across the curriculum is vital for impactful outcomes. Stakeholder uptake is a challenge owing to administrative burdens of operationalising. This can be even more challenging in large classes owing to the time and effort-intensive nature of using traditional paper-based methods (Anson & Goodman, 2014). Technology can overcome these limitations, however, usability challenges surrounding stakeholder dispositions (perceived usefulness) and learning capabilities (perceived ease of use) can affect uptake (Salloum et al., 2019).

1.3.4 Assessment and Feedback Literacy

The two areas of literacy, namely, assessment and feedback literacy, are critical to ensure greater validity, reliability, consistency and to have a positive impact on learning. Assessment literacy is "the ability to design, select, interpret, and use assessment results appropriately for education decisions" (Quilter & Gallini, 2000, p. 116). Unpacking two types of assessment literacy are critical in CTW context: collaborative learning assessment (Meijer et al., 2020) and peer assessment. The former refers to appropriate choice of assessment methods to align with the goals of collaborative learning. Both entail the capacity of students and instructors to understand the purpose and processes of assessment, as well as to accurately determine 'quality' in their (and others') work (Smith et al., 2013). Evidence suggests lack of clear understanding of the purpose and value of the process by students and instructors (Meijer et al., 2020). Instructor-student partnership in co-creating assessment rubrics are found to be effective but are relatively uncommon in practice (Deeley & Bovill, 2017).

Feedback literacy refers to the abilities and dispositions to seek, generate, understand and utilise feedback towards learning benefit, and develop academic judgement capacities (Molloy et al., 2020). Poor feedback literacy can lead to lack of pedagogical consideration and poor engagement (Koh et al., 2021), emotional distress (Zhou et al., 2020), ineffective past-oriented feedback and poor implementation of feedback practices (Winstone et al., 2019). Koh et al. (2021) found that lack of authentic ownership and engagement of teachers can lead to poor educational outcomes. Likewise, the importance of a clear understanding of pedagogy, technology and content knowledge, and the need for unfolding the teacher's role are critical to mitigate assessment and feedback literacy limitations (Moore & Hampton, 2015b).

1.4 Framework Development

Analysis of the literature reveals a dearth of focused frameworks specifically addressing peer assessment challenges in CTW context. For example, Gielen et al.'s (2011) typology explores the diversity of peer assessment in a broader context by extending Topping's (1998) typology classifying 20 variables into five clusters (peer assessment decisions, link between assessment and learning environments, peer interaction, composition, and management of procedure) with a single reference to peer assessment of behaviour. Adachi et al. (2018) framework extends this, incorporating 19 contextual elements covering broader peer assessment context, with peer assessment of process cited once. Overall, existing frameworks fail to consider the complexities of peer assessment in the CTW context.

To fill this gap, this chapter proposes a framework which is designed to mitigate specific challenges surrounding peer assessment in the CTW context to enable deeper understanding of conditions for success, appropriate decisions by key stakeholders to derive best outcomes, and enhance enabling factors to facilitate successful learning. The framework is designed to aid educators and policymakers in determining how best to implement peer assessment which enhances student learning and outcomes.

The framework responds to the needs of key stakeholders: students by supporting peer learning through addressing accountability, engagement and emotional issues; accreditation bodies in authentic provision of assurance of learning evidence; employers by equipping students with work and life-ready skills, and educators, scholars and policymakers in facilitating effective operationalisation of peer learning strategies.

1.4.1 Design

Empowering students to understand quality and standards is imperative to transform learning through efficacious peer assessment design strategies including: demystifying assessment criteria (to ensure accuracy); anonymity (to promote honesty); and incentives (to enhance engagement). Demystifying assessment criteria has the potential to ensure students can more accurately judge the work of others and trust their peers to evaluate their work. Students understanding of assessment criteria/rubrics is critical given they have the power to reward or penalise their teammates (Sridharan et al., 2019). Learning activities entailing co-creation or discussion of rubrics along with examples may foster a shared understanding of quality and standards (Jopp, 2020). Ashton and Davies (2015) found that training students to assess improves their ability to differentiate quality between novice, intermediate, and advanced levels and provide quality feedback information. Likewise, assessor-training and calibration practices can diminish capability challenges (Li et al., 2020).

Anonymity in peer assessment offers advantages in terms of positive attitudes towards feedback, enhanced student learning, improved quality of feedback,

and prevention of undesirable social effects like peer pressure and favouritism (Panadero & Alqassab, 2019; Rotsaert et al., 2018). However, Rotsaert et al. (2018) contend that anonymity can prevent students from a two-way interactive feedback dialogue. On the other hand, anonymity can overcome the psychological safety challenges in truthful peer assessment (Vanderhoven et al., 2015). Besides, anonymity may help students to focus on the content of the feedback rather than the source, especially when there may be emotional tension arising from receiving and acting on feedback from a peer who is of equal status (Anson & Goodman, 2014). Indeed, while there are many positive features on feedback not being anonymous in situations without summative assessment, there are circumstances in which anonymity is needed.

Incentives to engage with both formative and summative practices is a critical aspect of successful peer assessment. To enhance student engagement, Gillanders et al. (2020) stress the need for detailed guidance for students, lecturer accessibility and exemplars. Stepanyan et al. (2009) identified four key components to engagement, including: supportive tutors; anonymity; accessing peer work; and the allocation of marks and in-class activities. Mark allocation can help students determine the value and overall importance of assessment tasks (Sridharan et al., 2019). While there can be no perfect breakdown/weight, the weighting allocation should: (a) reflect the goals for student learning and outcomes; and (b) seek to motivate students to produce high quality of work.

1.4.2 Implementation

Prior studies propose several strategies to tackle the validity and reliability concerns of peer assessment, classified into instrument validity, marking method validity and moderation process. Instrument validity refers to the choice of fit-for-purpose items with good measurement properties along with a well-defined rating scale. In this regard, Loughry et al. (2007) proposed an empirically tested and robust instrument comprising 87 items covering five dimensions based on extensive theoretical and empirical research. This has been integrated into the CATME tool, used extensively for practical implementation of self and peer assessment (Loughry et al., 2014). Similarly, Lejk and Wyvill (2001) reported the effectiveness of a holistic and category-based peer assessment instrument covering six dimensions.

Marking method validity refers to the appropriate choice of a marking calculation method that leads to consequential learning. To address integrity challenges, diverse calculation methods have been proposed such as weighted marks (Freeman & McKenzie, 2002), procedures to correct for marker biases (Li, 2001) and relative performance factors (Willey & Gardner, 2009) to deal with variation in marking standards and quality within and between groups.

Two popular choices are considering peer assessment of process and adjusting CTW product mark by individual process marks. Peer assessment of process has a number of benefits including tackling teamwork challenges and providing
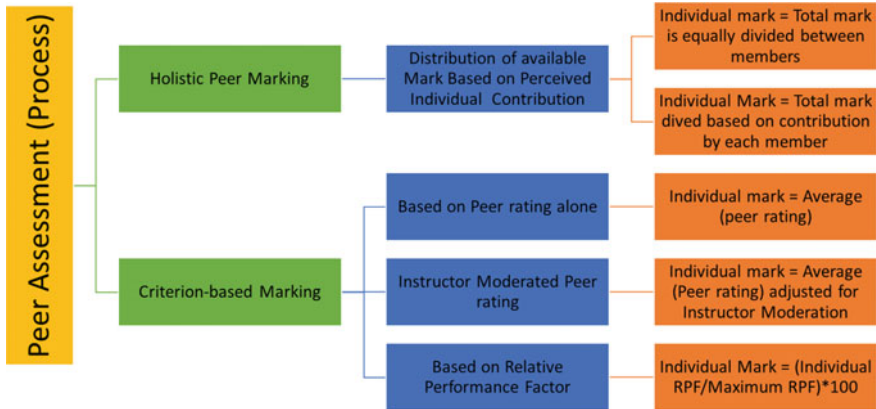
Fig. 1.1 Peer assessment of process: calculation options

assurance of learning evidence for accreditation bodies (Loughry et al., 2014). Figure 1.1 provides an overview of diverse calculation options with progressively increasing complexity and validity, adopting both holistic and criterion-based peer rating methods. While holistic marking is easy to implement, evidence suggests lack of mark differentiation compared to criterion-referenced approach (Lejk & Wyvill, 2001). Another limitation of holistic marking is the inability to provide information on specific areas for improvement. Criterion-based marking has the potential to reduce marking bias if implemented effectively and help identify weak areas. Calculations based on individual performance relative to the group performance can be more reliable as this addresses issues of variation in marking standards. Relative performance factor (RPF) is calculated as follows:

$$\text{RPF factor} = \text{Total ratings for individual team member}$$
$$\div \text{ Average of total rating for all team members}$$

Adjusting product marks by process mark enables allocation of individual marks for a CTW task based on individual contributions. Figure 1.2 provides more nuanced methods for adjusting product by process marks using types of calculation methods[1] with varying degrees of penalties for poor behaviours in working as a team. Specifically, the three methods for calculating RPF include: non-linear (square root of ratio of RPF); linear (simple ratio—RPF formula); and curvilinear method (linear formula for RPF scores below 1 and non-linear formula for RPF above 1). The non-linear model is less punitive than the linear model for under-contributors. The linear model is less punitive for over-contributors. The curvilinear model penalises both under and over-contributors. It might therefore

---

[1] https://sparkplus.com.au/using-sparkplus.php.
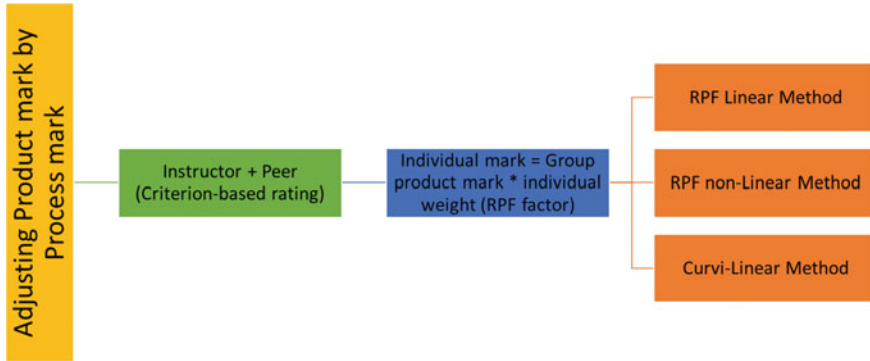
Fig. 1.2 Calculation methods for adjusting product grade with process results

be appropriate to adopt the non-linear method for first year students, the linear method for second year students and the curvilinear for final year and post graduate students.

Moderation process requires the shared understanding of quality and standards to address reliability concerns and instil confidence amongst students in peer rating. Sadler (2010) advocates the development of "appraisal expertise" to ensure students have the capacity to judge their own performance as well as that of their peers. Increased reliability can be realised through repeated exposure and provision of explicit rubric criteria (De Wever et al., 2011).

In this regard, three types of moderation activities are beneficial: pre-moderation (before marking commences), peri-moderation (during marking) and post moderation (after marking). Pre and peri-moderation activities require student engagement and post-moderation requires instructor engagement in adjusting the mark based on evidence provided by students. Pre-moderation activities include demystifying quality expectations, peer-rater calibration practices, and peer-rating training (Li et al., 2020). Peri-moderation could take the form of formative assessment by providing exposure to peer marking without penalty as well as developing self-awareness and taking corrective actions. Post-moderation requires instructors addressing marking variation within and between groups by using triangulation evidence from the system and students. Automated peer assessment tools such as CATME and FeedbackFruits have the power to provide additional information on students marking behaviours such as over-rating, colluding, and under-rating. This along with instructors' tacit knowledge and reflection activities could be used to moderate individual scores.

1.4.3   Technology

Embracing automation technology can alleviate scalability, sustainability and usability challenges (Anson & Goodman, 2014). Scalability relates to the capacity

to implement peer assessment in large classes and multiple units of study. Sustainability refers to maintaining initiatives across the curriculum continuously for long-term success. Usability refers to positive user experience and satisfaction to support sustained technology adoption.

A range of technologies and supporting functionalities need to be considered in choosing a system to mitigate these challenges. These include provision for: team formation, calibration exercises, peer assessment, giving and receiving feedback, feedback on feedback, team and individual reflection, and communities of inquiry activities. For example, SPARK[PLUS], CATME, FeedbackFruits, amongst other tools, have been used to support peer assessment and feedback activities (Loughry et al., 2014; Willey & Gardner, 2009). Institutional Learning Management Systems (LMS) tools such as discussion forums and Wikis can support communities of inquiry activities, brainstorming, exchanging ideas and information. Likewise, most LMS provide facilities for basic team formation such as self-selection, random allocation and teacher allocation for group formation.

Most self and peer assessment systems have advantages and disadvantages (See Fig. 1.4). CATME has unique functionality for dynamic team configuration enabling mixing homogenous and heterogeneous individual characteristics. Similarly, Feedbackfruit's unique feature is its ability to interact with institutional LMS. Both CATME and SPARK[PLUS] can automatically calculate a relative performance factor. Many of these technologies facilitate the automatic generation of results for individuals to compare their self-score against aggregate peer scores. 'Team charter' from CATME can support team meetings, setting out roles, expectations, and processes, and laying foundations for teamwork which have been identified to enhance teamwork effectiveness (Bell et al., 2018). Additionally, some of these technologies can classify students based on their marking pattern (such as overconfident, underconfident, manipulator, conflict, clique) using a powerful algorithm, which can be useful for instructor post-moderation processes.

These technologies also help develop lifelong skills; namely evaluative judgement (the ability to judge the quality of one's own and others' work) (Boud et al., 2018). However, effective use of these to derive benefit relies upon ease of use of the tool, stakeholder engagement, pedagogical underpinning and ownership of implementation. For example, it is crucial to consider the trade-off between usability and functionality of these systems for securing institutional licensing.

1.4.4   Roles and Responsibilities

The development of knowledge, skills and ability of both instructors and students, is critical to address peer assessment literacy challenges, to effectively fulfil their respective functions through partnership and shared roles and responsibilities. In particular, the two areas of literacy, namely, assessment and feedback literacy, are critical as the evidence suggests making evaluative judgements and providing effective feedback are complex and must be learned (Boud et al., 2018).

Assessment literacy is critical and viewed by some as a sine qua non for instructors as inadequate knowledge in assessment impacts the overall quality of education (Popham, 2009). According to Pastore and Andrade (2019), assessment literacy helps instructors use critical information about student learning to teach more effectively, enabling them to respond to students' learning needs. For students, assessment literacy relates to three key factors according to Smith et al., (2013, p. 1): (1) understanding the purpose of assessment and how it connects to their learning overall; (2) awareness of the process of assessment; and (3) the opportunity to practice making judgements about quality and areas for improvement.

To support peer assessment, Meijer et al. (2020) stress the importance of appreciating the rationale and purpose of collaborative learning and assessment between instructor-students and among students to develop assessment literacy. Deeley and Bovill (2017) argue the need for instructor-student partnership and its orientation for learning through engaging students as 'partners in assessment'. Peer assessment training has been found to increase perceptions of psychological safety which leads to increased confidence and trust in peer assessors (Cheng et al., 2015). Considering students' roles as assessee and assessor requires both emotional strength and resilience; training, monitoring and providing guidance in peer assessment is imperative (Gielen et al., 2011; Panadero, 2016).

Students need to be trained in assessment, feedback and evaluative judgement skills to improve peer assessment validity and reliability. Developing stakeholders skills in feedback provision to focus on task/process (not on person), orientation (forward-oriented) and specificity (areas for improvement) are critical to influence positive impact on learning and behaviour. The provision of exemplars, calibration and formative assessment tasks, co-designing evaluation tools are powerful mechanisms in developing evaluative judgements around what constitutes 'quality'. Carless and Boud (2018) highlight the teacher's role in modelling the uptake of feedback by encouraging students to seek, use, generate and act on feedback. Developing skills around peer feedback is critical to ensure effective elicitation, process and enaction by students (Malecka et al., 2020). Peer assessment skills could be further enhanced through reflecting on feedback and feedback on feedback.

In summation, the roles and responsibilities of both instructors and students broadly relate to: (a) capacity building and engagement with resources to develop peer assessment literacy; (b) engagement in calibration exercises, formative assessment, summative assessment, giving feedback, use of technology; (c) proactively seeking, engaging and acting on feedback; and (d) reflecting and taking actions for continuous improvement and lifelong learning.

Based on the above analysis of literature, we propose a a four-pillar framework by holistically considering complex and intertwined challenges of peer assessment in formal CTW assessment context. This is designed to provide guidance to educators and scholars for navigating various peer learning challenges and creating a
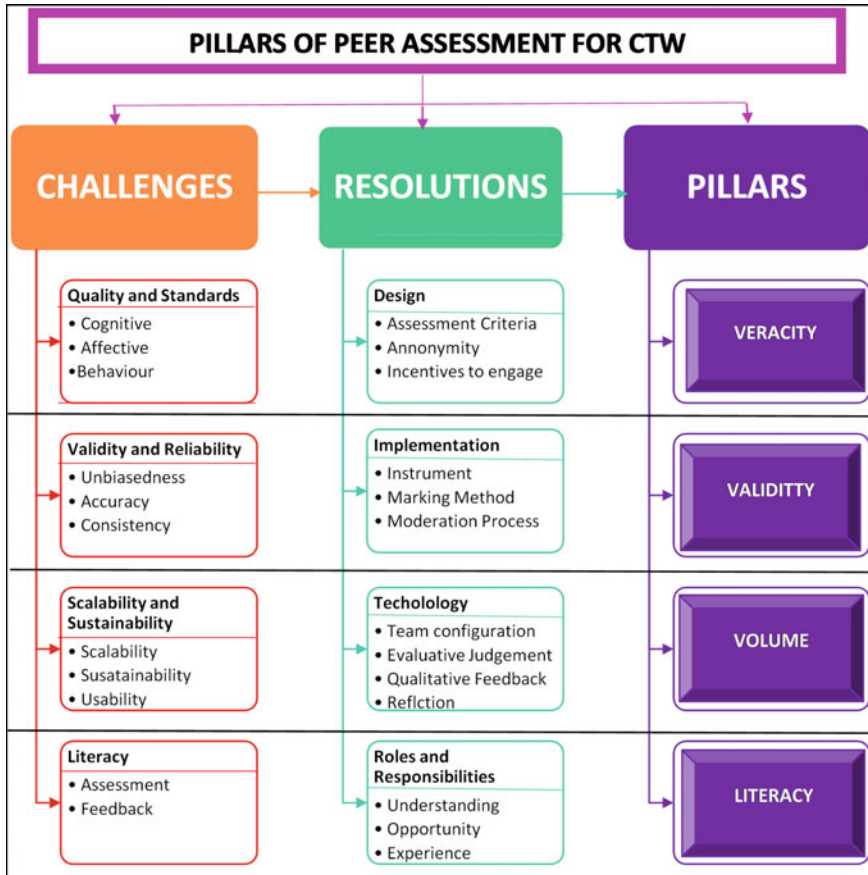
Fig. 1.3 The four pillars of peer assessment

stable and sustainable peer learning ecosystem model to have an impactful out-
come as shown in Fig. 1.3. However, we acknowledge the need for adaptation to
align with the context and purpose of the peer learning to effect change.

1.5   Discussion

The framework presented four key pillars (veracity, validity, volume and literacy)
based on themes emerged from a critical review of the literature contributing to
scholarship encompassing a broad scope of enabling strategies to mitigate chal-
lenges associated with peer assessment in CTW, which few existing models do.
We contend that when designed and implemented effectively, peer assessment in
CTW can become a powerful strategy to instil a range of soft skills including team-
work, leadership, negotiation, conflict resolution, amongst others. The framework

Activity Specification	CATME	SparkPLUS	FeedbackFruits	LMS*
Simple Group formation (self-selection, random allocation, manual allocation)	⊠	⊠	☐	⊠
Dynamic groups formation	⊠	☐	☐	☐
Calibration practice	⊠	☐	☐	☐
Self and peer rating	⊠	⊠	⊠	☐
Choice for anonymity in rating and feedback	⊠	⊠	⊠	☐
Auto Populated Criteria Option	⊠	⊠	☐	☐
Flexibility to add nuanced criteria	☐	⊠	⊠	☐
Qualitative feedback to peers – Overall	⊠	⊠	⊠	☐
Qualitative feedback to peers – Criterion level	☐	☐	⊠	☐
Feedback on Feedback	⊠	☐	⊠	☐
Post-moderation information	⊠	☐	⊠	☐
Flexibility in mark allocation for completing various tasks	☐	☐	⊠	☐
Automatic RPF calculation	⊠	⊠	☐	☐
Accommodate Self Reflection	☐	☐	⊠	☐
Accommodate Team reflection	⊠	☐	⊠	☐
Private information to instructor (for any poor teamwork behaviour)	⊠	☐	☐	☐
Pushing group formation from LMS	☐	☐	⊠	NA
LMS Learning Technology Integration	☐	☐	⊠	NA
Discussion Forum, Wikis, etc.	☐	☐	☐	⊠
*Moodle, BrightSpace, Blackboard				

Fig. 1.4 Comparison of key features from self and peer assessment technologies

has the potential to influence key stakeholders to advance deeper understanding of challenges and opportunities in embracing effective peer assessment practices in CTW. The key implications for pragmatic application of the framework are summarised below.

To mitigate the capabilities and behavioural challenges, intervention strategies in the veracity pillar include demystifying expectations, anonymity and incentives. However, there is no 'one solution fits all' strategy to tackle the challenges. For example, a partnership approach to co-creation as a mechanism for developing shared understanding of quality and standards demands shift in perceptions of stakeholders (Bovill et al., 2016). Anonymity can tackle inhibitions in honest marking and reduce anxieties of retaliation from peers, however, it prevents serious engagement and dialogic conversation, which are critical for learning (Rotsaert et al., 2018). Formative assessment is powerful to support peer learning, however, lack of incentives can impede engagement. Introducing it as a hurdle task may solve this challenge. On the other hand, incentives in the form of summative assessment may lead to competition instead of collaboration. Integrating criteria for collaboration and cooperation can address this issue.

Approaches proposed in the validity pillar include robust implementation decisions about assessment instrument, marking method and moderation process with careful consideration to context and constraints. For instance, instructors need to carefully consider several factors: alignment with learning outcomes, choice of

methods conducive for learning and adopting appropriate moderation practices. To impact consequential learning, a range of solutions are proposed including a diverse choice of instruments, calculation methods such as weighted marks (Freeman & McKenzie, 2002), procedures to correct for marker biases (Li, 2001), use of a relative performance factor (Willey & Gardner, 2009) and moderation activities (pre, peri and post). To avert students turning against peer assessment without exposure, use of lenient marking methods for first year students and a firmer approach for mature students can be considered.

Enabling scalability and sustainability, volume pillar considers a scaffolded approach and multiple exposures to peer assessment. Effective practices can be achieved through technology affordances and instructors' ownership for successful implementation (Koh et al., 2021). A comparison of functionalities of three popular technologies namely Spark^{PLUS}, the CATME, FeedbackFruits is provided to make informed decisions in choosing a tool. Even with technology support, peer assessment can be a time-consuming task for novice instructors (Anson & Goodman, 2014). Recognition of this in workload models and capacity building sessions can pave the way for change. Additional program level policy decisions to scaffold across the course will enable authentic transformation of CTW skills and genuine uptake of peer assessment activities.

Developing a deeper understanding of formative and summative functions of assessment by key stakeholders is underscored in the literacy pillar. This requires both cognisance and application of the formative and summative assessment tasks and feedback practices to avert harmful effect on learning (Boud et al., 1999). Strategies to achieve this include assessment bootcamp sessions to explicate the purpose and processes; integrative assessment practices which requires actioning on feedback before attempting follow-on task; reflective writing on how they used the feedback; post-feedback proforma activities on the value and use of feedback; feedback on feedback to encourage deep engagement; developing students' capacity to give, receive and act on feedback; and mindful growth mindset feedback practices without invoking self-esteem issues. Developing appropriate institutional policies around reframing effective assessment, feedback and professional development practices can significantly resolve these challenges.

1.5.1 Usage of the Framework

The functioning of the framework has implications for a range of stakeholders including educators, policy makers and scholars. For educators, the framework offers a distilling of the extant research on the tensions, possible ways to overcome challenges, and purpose-fit approach to effective adoption of peer assessment in the CTW context. A critical factor in the effective use of the framework is building the capacity of both educators and students in understanding the complexities and pedagogical underpinnings of peer assessment. Once educators are equipped with

the necessary skills, they need to ensure students are also sufficiently trained in the skills required to effect change. Educators need to develop clear procedures and processes for students, and the framework may assist by functioning as an overview and checklist of critical points. In its comprehensive insights into the complex and multifaceted components, the framework may serve as a useful aid for educators in determining how best to implement peer assessment to enhance student learning and outcomes.

For institutional policy makers, the framework presents a pathway for addressing the tensions and developing policies and institutional support for mainstream adoption of best practices in peer assessment. Policy makers are often the way to ensure impactful outcomes at an institutional level. The framework proposes a comprehensive overview of challenges and resolutions around peer assessment, which may help inform best practices.

For researchers, the framework offers a useful distilling of the extensive body of extant literature around peer learning and assessment in the CTW context. It may prove useful in informing considerations of innovative initiatives and approaches in peer assessment moving forward, as well as serving as a springboard to future research.

1.5.2   Limitations

The proposed framework is not without its limitations. Firstly, it has emerged from work in a CTW context, which may mean it may not apply to all peer assessment contexts. Secondly, while it traverses a spectrum of significant challenges and mitigating factors, the framework may not address them all. Finally, successful implementation requires attention to the context in which peer assessment is being implemented.

1.5.3   Further Research

Further research has the potential to refine the framework, empirically test the effectiveness of the proposed strategies to support pragmatic application. Implementation and monitoring will help flesh out its parameters and limitations and assist in its finessing. Another consideration is to elaborate on the capacity building of students in peer assessment and optimal conditions under which they can be supported to develop their feedback and assessment literacy.

1.6 Conclusion

This chapter offers guidance for the multitude of challenges of peer assessment in the CTW context. It does so by identifying the various tensions within CTW and challenges from each of the pillars along with proposing recommendations and fit-for-purpose approaches to tackle the issues to support an effective peer assessment ecosystem. This requires holistically considering its multifaceted aspects through a seamless integration of all four pillars: veracity, validity, volume, and literacy. We underscore the aligned roles of students, instructors, technology and institutional support as catalysing agents of change for transformational learning. Additionally, a significant cultural shift in reimagining assessment and feedback practices, renewal of institution policies and capacity building of key stakeholders will go a long way to effect positive change. Considering the complexities and multifaceted requirements of CTW, more research is required to deal with the challenges of practical implementation for each of the pillars.

References

Adachi, C., Tai, J., & Dawson, P. (2018). A framework for designing, implementing, communicating and researching peer assessment. *Higher Education Research and Development, 37*(3), 453–467. https://doi.org/10.1080/07294360.2017.1405913

Anson, R., & Goodman, J. A. (2014). A peer assessment system to improve student team experiences. *Journal of Education for Business, 89*(1), 27–34. https://doi.org/10.1080/08832323.2012.754735

Ashton, S., & Davies, R. S. (2015). Using scaffolded rubrics to improve peer assessment in a MOOC writing course. *Distance Education, 36*(3), 312–334. https://doi.org/10.1080/01587919.2015.1081733

Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review,* 100489. https://doi.org/10.1016/j.edurev.2022.100489

Bell, S. T., Brown, S. G., Colaneri, A., & Outland, N. (2018). Team composition and the ABCs of teamwork. *American Psychologist, 73*(4), 349–362. https://doi.org/10.1037/amp0000305

Boud, D., Ajjawi, R., Dawson, P., & Tai, J. (2018). *Developing evaluative judgement in higher education*. Routledge.

Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment and Evaluation in Higher Education, 24*(4), 413–426. https://doi.org/10.1080/0260293990240405

Boud, D., Cohen, R., & Sampson, J. (2014). *Peer learning in higher education: Learning from and with each other*. Routledge.

Bovill, C., Cook-Sather, A., Felten, P., Millard, L., & Moore-Cherry, N. (2016). Addressing potential challenges in co-creating learning and teaching: Overcoming resistance, navigating institutional norms and ensuring inclusivity in student–staff partnerships. *Higher Education, 71*(2), 195–208. https://doi.org/10.1007/s10734-015-9896-4

Bravo, R., Catalán, S., & Pina, J. M. (2019). Analysing teamwork in higher education: An empirical study on the antecedents and consequences of team cohesiveness. *Studies in Higher Education, 44*(7), 1153–1165. https://doi.org/10.1080/03075079.2017.1420049

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment and Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Chapman, A. (2017). Using the assessment process to overcome Imposter Syndrome in mature students. *Journal of Further and Higher Education, 41*(2), 112–119. https://doi.org/10.1080/0309877X.2015.1062851

Cheng, K.-H., Liang, J.-C., & Tsai, C.-C. (2015). Examining the role of feedback messages in undergraduate students' writing performance during an online peer assessment activity. *The Internet and Higher Education, 25*, 78–84. https://doi.org/10.1016/j.iheduc.2015.02.001

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2011). Assessing collaboration in a wiki: The reliability of university students' peer assessment. *The Internet and Higher Education, 14*(4), 201–206. https://doi.org/10.1016/j.iheduc.2011.07.003

Deeley, S. J., & Bovill, C. (2017). Staff student partnership in assessment: Enhancing assessment literacy through democratic practices. *Assessment and Evaluation in Higher Education, 42*(3), 463–477. https://doi.org/10.1080/02602938.2015.1126551

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education, 24*(3), 331–350. https://doi.org/10.1080/03075079912331379935

El Massah, S. S. (2018). Addressing free riders in collaborative group work: The use of mobile application in higher education. *International Journal of Educational Management*. https://doi.org/10.1108/IJEM-01-2017-0012

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287–322. https://doi.org/10.3102/00346543070003287

Frazier, M. L., Fainshmidt, S., Klinger, R. L., Pezeshkan, A., & Vracheva, V. (2017). Psychological safety: A meta-analytic review and extension. *Personnel Psychology, 70*(1), 113–165. https://doi.org/10.1111/peps.12183

Freeman, M., & McKenzie, J. (2002). SPARK, a confidential web–based template for self and peer assessment of student teamwork: Benefits of evaluating across different subjects. *British Journal of Educational Technology, 33*(5), 551–569. https://doi.org/10.1111/1467-8535.00291

Gielen, S., Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment and Evaluation in Higher Education, 36*(2), 137–155. https://doi.org/10.1080/02602930903221444

Gillanders, R., Karazi, S., & O'Riordan, F. (2020). Loss aversion as a motivator for engagement with peer assessment. *Innovations in Education and Teaching International, 57*(4), 424–433. https://doi.org/10.1080/14703297.2020.1726203

Hansen, R. S. (2006). Benefits and problems with student teams: Suggestions for improving team projects. *Journal of Education for Business, 82*(1), 11–19. https://doi.org/10.3200/JOEB.82.1.11-19

Hoo, H. -T., Deneen, C., & Boud, D. (2021). Developing student feedback literacy through self and peer assessment interventions. *Assessment and evaluation in higher education*, 1–14.https://doi.org/10.1080/02602938.2021.1925871

Jacobs, G. M., & Renandya, W. A. (2019). *Student centered cooperative learning: Linking concepts in education to promote student learning.* Springer.

Jones, E., Priestley, M., Brewster, L., Wilbraham, S. J., Hughes, G., & Spanner, L. (2021). Student wellbeing and assessment in higher education: The balancing act. *Assessment and Evaluation in Higher Education, 46*(3), 438–450. https://doi.org/10.1080/02602938.2020.1782344

Jopp, R. (2020). A case study of a technology enhanced learning initiative that supports authentic assessment. *Teaching in Higher Education, 25*(8), 942–958. https://doi.org/10.1080/13562517.2019.1613637

Koh, E. R., Tan, J. P. -L., Hong, H., Suresh, D., & Tee, Y. -H. (2021). Infusing the teamwork innovation my groupwork buddy in schools: Enablers and impediments. In *Scaling up ICT-based innovations in schools* (pp. 151–171). Springer.

Latifi, S., & Noroozi, O. (2021). Supporting argumentative essay writing through an online supported peer-review script. *Innovations in Education and Teaching International, 58*(5), 501–511. https://doi.org/10.1080/14703297.2021.1961097

Latifi, S., Noroozi, O., & Talaee, E. (2021). Peer feedback or peer feedforward? Enhancing students' argumentative peer learning processes and outcomes. *British Journal of Educational Technology, 52*(2), 768–784. https://doi.org/10.1111/bjet.13054

Lawlor, J., Conneely, C., Oldham, E., Marshall, K., & Tangney, B. (2018). Bridge21: Teamwork, technology and learning. A pragmatic model for effective twenty-first-century team-based learning. *Technology, Pedagogy and Education*, *27*(2), 211–232. https://doi.org/10.1080/1475939X.2017.1405066

Lejk, M., & Wyvill, M. (2001). Peer assessment of contributions to a group project: A comparison of holistic and category-based approaches. *Assessment and Evaluation in Higher Education, 26*(1), 61–72. https://doi.org/10.1080/02602930020022291

Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment and Evaluation in Higher Education, 45*(2), 193–211. https://doi.org/10.1080/02602938.2019.1620679

Li, L. K. (2001). Some refinements on peer assessment of group projects. *Assessment and Evaluation in Higher Education, 26*(1), 5–18. https://doi.org/10.1080/0260293002002255

Loughry, M. L., Ohland, M. W., & DeWayne Moore, D. (2007). Development of a theory-based assessment of team member effectiveness. *Educational and Psychological Measurement, 67*(3), 505–524. https://doi.org/10.1177/0013164406292085

Loughry, M. L., Ohland, M. W., & Woehr, D. J. (2014). Assessing teamwork skills for assurance of learning using CATME team tools. *Journal of Marketing Education, 36*(1), 5–19. https://doi.org/10.1177/0273475313499023

Malecka, B., Boud, D., & Carless, D. (2020). Eliciting, processing and enacting feedback: Mechanisms for embedding student feedback literacy within the curriculum. *Teaching in Higher Education*, 1–15.https://doi.org/10.1080/13562517.2020.1754784

Marasi, S. (2019). Team-building: Developing teamwork skills in college students using experiential activities in a classroom setting. *Organization Management Journal, 16*(4), 324–337. https://doi.org/10.1080/15416518.2019.1662761

McKendall, M. (2000). Teaching groups to become teams. *Journal of Education for Business, 75*(5), 277–282. https://doi.org/10.1080/08832320009599028

Meijer, H., Hoekstra, R., Brouwer, J., & Strijbos, J.-W. (2020). Unfolding collaborative learning assessment literacy: A reflection on current assessment methods in higher education. *Assessment and Evaluation in Higher Education, 45*(8), 1222–1240. https://doi.org/10.1080/02602938.2020.1729696

Mihelič, K. K., & Culiberg, B. (2019). Reaping the fruits of another's labor: The role of moral meaningfulness, mindfulness, and motivation in social loafing. *Journal of Business Ethics, 160*(3), 713–727. https://doi.org/10.1007/s10551-018-3933-z

Molloy, E., Boud, D., & Henderson, M. (2020). Developing a learning-centred framework for feedback literacy. *Assessment and Evaluation in Higher Education, 45*(4), 527–540. https://doi.org/10.1080/02602938.2019.1667955

Moore, P., & Hampton, G. (2015a). 'It's a bit of a generalisation, but…': Participant perspectives on intercultural group assessment in higher education. *Assessment and Evaluation in Higher Education, 40*(3), 390–406. https://doi.org/10.1080/02602938.2014.919437

Moore, P., & Hampton, G. (2015b). 'It's a bit of a generalisation, but…': Participant perspectives on intercultural group assessment in higher education. *Assessment and Evaluation in Higher Education, 40*(3), 390–406.

O'Neill, T. A., & Mclarnon, M. J. (2018). Optimizing team conflict dynamics for high performance teamwork. *Human Resource Management Review, 28*(4), 378–394. https://doi.org/10.1016/j.hrmr.2017.06.002

Oakley, B., Felder, R. M., Brent, R., & Elhajj, I. (2004). Turning student groups into effective teams. *Journal of Student Centered Learning, 2*(1), 9–34.

Oakley, B. A., Hanna, D. M., Kuzmyn, Z., & Felder, R. M. (2007). Best practices involving teamwork in the classroom: Results from a survey of 6435 engineering student respondents. *IEEE Transactions on Education, 50*(3), 266–272. https://doi.org/10.1109/TE.2007.901982

Oosthuizen, H., De Lange, P., Wilmshurst, T., & Beatson, N. (2021). Teamwork in the accounting curriculum: Stakeholder expectations, accounting students' value proposition, and instructors' guidance. *Accounting Education, 30*(2), 131–158. https://doi.org/10.1080/09639284.2020.1858321

Opdecam, E., & Everaert, P. (2018). Seven disagreements about cooperative learning. *Accounting Education, 27*(3), 223–233. https://doi.org/10.1080/09639284.2018.1477056

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment. In *Handbook of human and social conditions in assessment* (pp. 247–266).

Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment and Evaluation in Higher Education, 44*(8), 1253–1278. https://doi.org/10.1080/02602938.2019.1600186

Panadero, E., Romero, M., & Strijbos, J.-W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation, 39*(4), 195–203. https://doi.org/10.1016/j.stueduc.2013.10.005

Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education, 84*, 128–138. https://doi.org/10.1016/j.tate.2019.05.003

Planas-Lladó, A., Feliu, L., Arbat, G., Pujol, J., Suñol, J. J., Castro, F., & Martí, C. (2021). An analysis of teamwork based on self and peer evaluation in higher education. *Assessment and Evaluation in Higher Education, 46*(2), 191–207. https://doi.org/10.1080/02602938.2020.1763254

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice, 48*(1), 4–11. https://doi.org/10.1080/00405840802577536

Quilter, S. M., & Gallini, J. K. (2000). Teachers' assessment literacy and attitudes. *The Teacher Educator, 36*(2), 115–131. https://doi.org/10.1080/08878730009555257

Riley, J., & Ward, K. (2017). Active learning, cooperative active learning, and passive learning methods in an accounting information systems course. *Issues in Accounting Education, 32*(2), 1–16. https://doi.org/10.2308/iace-51366

Rotsaert, T., Panadero, E., & Schellens, T. (2018). Anonymity as an instructional scaffold in peer assessment: Its effects on peer feedback quality and evolution in students' perceptions about peer assessment skills. *The European Journal of Psychology of Education, 33*, 75–99. https://doi.org/10.1007/s10212-017-0339-8

Rubin, R. S., & Dierdorff, E. C. (2009). How relevant is the MBA? Assessing the alignment of required curricula and required managerial competencies. *Academy of Management Learning and Education, 8*(2), 208–224. https://doi.org/10.5465/amle.2009.41788843

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. In *Approaches to assessment that enhance learning in higher education* (pp. 55–70). Routledge.

Salas, E., Reyes, D. L., & McDaniel, S. H. (2018). The science of teamwork: Progress, reflections, and the road ahead. *American Psychologist, 73*(4), 593. https://doi.org/10.1037/amp0000334

Salas, E., Shuffler, M. L., Thayer, A. L., Bedwell, W. L., & Lazzara, E. H. (2015). Understanding and improving teamwork in organizations: A scientifically based practical guide. *Human Resource Management, 54*(4), 599–622. https://doi.org/10.1002/hrm.21628

Salloum, S. A., Alhamad, A. Q. M., Al-Emran, M., Monem, A. A., & Shaalan, K. (2019). Exploring students' acceptance of e-learning through the development of a comprehensive technology acceptance model. *IEEE Access, 7*, 128445–128462. https://doi.org/10.1109/ACCESS.2019.2939467

Schlösser, T., Dunning, D., Johnson, K. L., & Kruger, J. (2013). How unaware are the unskilled? Empirical tests of the "signal extraction" counter explanation for the Dunning-Kruger effect in self-evaluation of performance. *Journal of Economic Psychology, 39*, 85–100. https://doi.org/10.1016/j.joep.2013.07.004

Smith, C. D., Worsfold, K., Davies, L., Fisher, R., & McPhail, R. (2013). Assessment literacy and student learning: The case for explicitly developing students 'assessment literacy.' *Assessment and Evaluation in Higher Education, 38*(1), 44–60. https://doi.org/10.1080/02602938.2011.598636

Speyer, R., Pilz, W., Van Der Kruis, J., & Brunings, J. W. (2011). Reliability and validity of student peer assessment in medical education: A systematic review. *Medical Teacher, 33*(11), e572–e585.

Sridharan, B., & Boud, D. (2019). The effects of peer judgements on teamwork and self-assessment ability in collaborative group work. *Assessment and Evaluation in Higher Education, 44*(6), 894–909. https://doi.org/10.1080/02602938.2018.1545898

Sridharan, B., Tai, J., & Boud, D. (2019). Does the use of summative peer assessment in collaborative group work inhibit good judgement? *Higher Education, 77*(5), 853–870. https://doi.org/10.1007/s10734-018-0305-7

Stepanyan, K., Mather, R., Jones, H., & Lusuardi, C. (2009). Student engagement with peer assessment: A review of pedagogical design and technologies. In *International Conference on Web-Based Learning*, Berlin, Heidelberg.

Stover, S., & Holland, C. (2018). Student resistance to collaborative learning. *International Journal for the Scholarship of Teaching and Learning, 12*(2), 8.

Strijbos, J.-W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*(4), 265–269. https://doi.org/10.1016/j.learninstruc.2009.08.002

Taghizadeh, K. N., Noroozi, O., Banihashem, S. K., Karami, M. & Biemans, H. J. A. (2022). Online peer feedback patterns of success and failure in argumentative essay writing. *Interactive Learning Environments*, 1–10. https://doi.org/10.1080/10494820.2022.2093914

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276. https://doi.org/10.3102/00346543068003249

Topping, K. J. (2005). Trends in peer learning. *Educational Psychology, 25*(6), 631–645. https://doi.org/10.1080/01443410500345172

Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers and Education, 81*, 123–132. https://doi.org/10.1016/j.compedu.2014.10.001

Willey, K., & Gardner, A. (2009). Improving self-and peer assessment processes with technology. *Campus-Wide Information Systems, 26*(5), 379–399.

Winstone, N. E., Mathlin, G., & Nash, R. A. (2019). Building feedback literacy: Students' perceptions of the developing engagement with feedback toolkit. *Frontiers in Education, 4*, 39. https://doi.org/10.3389/feduc.2019.00039

Zhou, J., Zheng, Y., & Tai, J. H. M. (2020). Grudges and gratitude: The social-affective impacts of peer assessment. *Assessment and Evaluation in Higher Education, 45*(3), 345–358. https://doi.org/10.1080/02602938.2019.1643449

Learning Analytics for Peer Assessment: A Scoping Review

2

Kamila Misiejuk and Barbara Wasson

2.1 Introduction

Learning analytics (LA) is a research field that focuses on analysing educational data, with the goal of understanding and/or improving learning. LA is identified as having the potential to change assessment practices and support "the holistic process of learning" (Ferguson et al., 2016). Knight (2020) argues that LA can be used to move the focus from the summative assessment of products produced to facilitate more process-oriented assessments. Similarly, Archer and Prinsloo (2020) write that LA supports the assessment of and for learning and can help in understanding student learning, analysing learning behavior, predicting student learning needs, and prescribing interventions that may promote more effective teaching and learning; however, the ethics of student surveillance and privacy issues must be considered.

Some LA researchers note that assessment data are not commonly considered "an integral part of the analytics data cycle," but, rather, as an outcome measurement, which leads to assessment analytics being "still under-explored and largely under-developed" (Saqr, 2017, 1). Other reasons for not including assessment data in LA datasets are related to the strong emphasis on behavioural data rather than traditional assessment data, which may be more meaningful; the fact that LA is not led by pedagogy; and the fact that assessment data are not granular enough to allow

K. Misiejuk (✉) · B. Wasson
Centre for the Science of Learning and Technology (SLATE), University of Bergen, Bergen, Norway
e-mail: kamila.misiejuk@uib.no

Department of Information Science and Media Studies, University of Bergen, Bergen, Norway

O. Noroozi and B. de Wever (eds.), *The Power of Peer Learning*,
Social Interaction in Learning and Development,
https://doi.org/10.1007/978-3-031-29411-2_2

a detailed analysis of student behaviour (Ellis, 2013). There are also some concerns about implementing LA around removing human mentors from the feedback loop and students gaming the analytics (Buckingham Shum & Ferguson, 2012). On the other hand, the inclusion of assessment data, especially feedback data, has the potential to close the gap between data and education, increase LA usefulness, and broaden LA's scope (Ellis, 2013; Pardo, 2018; Saqr, 2017). Assessment data also have the advantage of being relatively easy to capture because students expect to be assessed based on their performance (Ellis, 2013). Knight (2020) highlights the fact that the "development of assessments based on novel process-based data is challenging (…) Thus, this development is likely to be time-consuming, expensive, and require systemic changes" (p. 133), and he also argues that the data should be used to support, not supplant, humans in their assessment practices.

Cope and Kalantzis (2016) mapped new assessment models that emerged alongside the increased prevalence of educational big data, including embedding assessment in learning, an increased focus on formative assessment, and a new conceptualization of summative assessment as a *progress view* rather than an *end view* of learning. Knight (2020) described three ways of transforming formative assessment with the help of LA: (1) developing new assessment techniques, (2) automating existing assessment techniques, or (3) augmenting existing assessment techniques. Moreover, he presented some potential augmentation scenarios, such as using LA to automatically allocate peers or automate feedback on the quality of the student feedback provided (backward evaluation).

One form of formative assessment is *peer assessment* (PA). Liu and Carless (2006) distinguish between *peer assessment* as "students grading the work or performance of their peers using relevant criteria" (p. 280) and *peer feedback* as "a communication process through which learners enter into dialogues related to performance and standards" (p. 280). In this chapter, we use PA as an umbrella term for all forms of PA, including peer feedback, peer grading, and peer review. Early LA research identified the potential of using LA techniques for constructionist learning activities, such as PA (Berland et al., 2014). Some potential practical implementations of LA in PA included feedback classification, a text analysis of rubric answers, combining peer and automated assessment, predicting the accuracy of peer raters, text analysis to monitor feedback quality and appropriateness, and clustering and visualisation techniques to optimise the feedback process (Ryan et al., 2019; Wahid et al., 2016).

2.2 Purpose of the Present Study

As the LA field is heading toward maturity, there is a need to examine how LA has been implemented in the field of PA. To date, there have been no literature reviews conducted on the broad topic of using LA in PA research, although there are reviews of LA and formative feedback (see Banihashem et al., 2022) and one review on an aspect of LA and PA (see Fig. 2.1). In a systematic literature review that included 28 papers, Nyland (2018) identified tools and techniques for
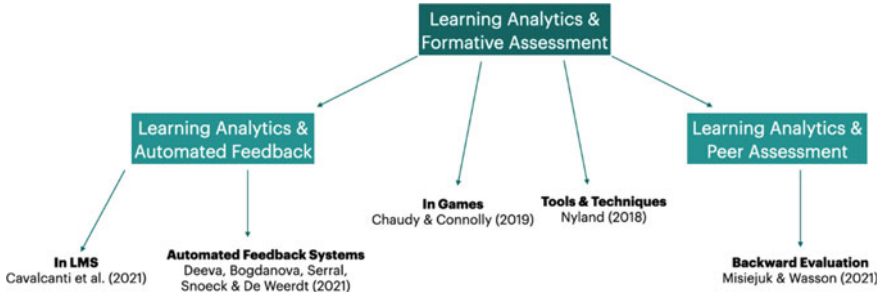
Fig. 2.1 Review studies of learning analytics and formative assessment

data-enabled formative assessment. Cavalcanti et al. (2021) conducted a systematic literature review that included 63 papers on automatic feedback generation in learning management systems. Chaudy and Connolly (2019) explored 34 relevant studies to identify the various approaches to integrating assessment in educational games and their associated empirical evidence. Deeva et al. (2021) classified and described 109 automated feedback systems. Misiejuk and Wasson (2021) focused on backward evaluation in PA, which is students receiving feedback on the quality of the feedback that they have given.

To fill this gap and help understand how LA is being used in PA, this chapter reports on a scoping review that focused on three research questions:

(1) Where in the peer assessment process are the analytics employed? What is the role of learning analytics in peer assessment research?
(2) What are the reported peer assessment challenges the research addressed with learning analytics? And how are they addressed?
(3) What insights into peer assessment can we gain from learning analytics?

2.3   Methodology

2.3.1   Scoping Review

As no studies analysing the broad use of LA in PA research have been conducted, a scoping review exploring the "the breadth and depth of a field" is an appropriate method with which to close this gap (Levac et al., 2010, 1). In this study, the scoping review approach, as described by Levac et al. (2010) was used. This included discussions between two researchers on the inclusion/exclusion of some of the papers, an iterative process of refining the coding criteria and research questions, and a report on the methodological details of the scoping review process.

2.3.2 Search

The search was conducted in December 2021 and resulted in 1534 papers (duplicates removed), which were screened for inclusion over three rounds (see search details in Fig. 2.2). Papers not written in English, those that were not peer-reviewed, and those published before 2011—the year of the first Learning Analytics and Knowledge (LAK) Conference—were excluded. Due to a large number of papers found during the search, the first screening focused on a detection of the phrase "learning analytics" in the abstract, title, or keywords and full text of the papers; in this way, "learning analytics" served as a proxy for authors centering themselves in the field of LA. If "learning analytics" was only found in the references, the paper was excluded. Papers published at the LAK Conference or in the Journal of Learning Analytics were allowed to bypass this rule, with the assumption that publishing in these places automatically establishes a link to LA. After the first round, 598 papers remained. The second screening tackled the "peer assessment" aspect of the review by checking whether some form of PA was described in the methods section of the article. After two rounds of screenings, the full text of 166 papers was examined for their relevance to the research questions.

After the exclusion of the non-relevant papers, the final review included 27 papers: fourteen journal articles and 13 conference papers. While most had an
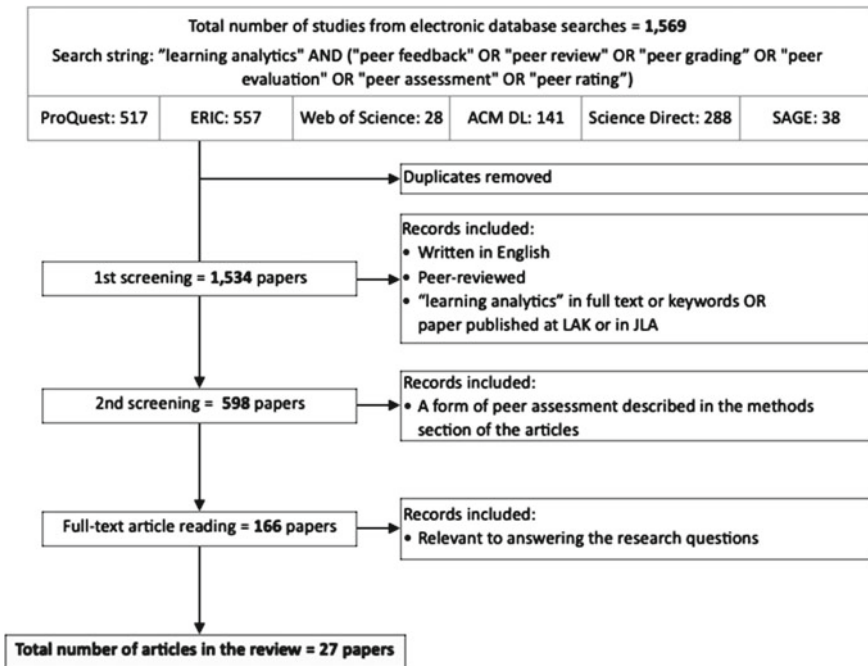
Fig. 2.2 Inclusion/exclusion process

overall focus on PA, for some, PA was secondary. For example, some papers used *PA data* for LA and delivered new insights into PA research, although the focus of the paper was not on PA. Most papers (22 of 27) conducted their studies in the context of higher education, except for Misiejuk et al. (2021), whose dataset included data from both higher education and K-12; Koh et al. (2016), Mørch et al. (2017), who used K-12 data; Hunt et al. (2021), who focused on professional development; and Babik et al. (2019), who simulated a dataset. Table 2.1 provides an overview of the 27 included papers.

2.4 Results

RQ1: Where in the peer assessment process are the analytics employed? What is the role of learning analytics in peer assessment research?

In the application of LA, eleven papers used LA to improve PA activity (*LA for PA*), while 15 papers used LA to analyse PA data (*LA on PA data*). We identified three main roles on the part of LA in improving PA: *tools*, *automated feedback*, and *visualizations*. For the papers that used LA to analyse PA data, four main application areas were mapped: *student interaction*, *feedback characteristics*, *comparison,* and *design.* Although some papers apply LA in more than one role, the paper categorization discussed below focused on the main use of LA in PA research described in the paper. Only one paper included both. Cheng and Lei (2021) both analysed PA data and developed visualizations for PA that showed students the social networks of their blogging and PA activities. Then, they examined the visualisation's influence on their engagement and group cohesion.

Tools. Four papers presented or developed *tools* with LA to help in facilitating PA. Using a novel quantitative approach, Nalli et al. (2021) developed and validated a Moodle plugin to facilitate the creation of heterogenous groups for a PA activity based on Moodle activity data. Chaparro-Peláez et al. (2020) developed a Moodle application, Workshop Data EXtractor (MWDEX), that can be used to extract, process, analyse, and visualize PA data in Moodle Workshops, and they conducted a short survey with instructors to validate the tool and inquire into how they implement PA. Vozniuk et al. (2014) presented an extension to a social media platform, GRAASP, that facilitates rating-based PA. The extension was evaluated in two analyses: (1) the validity of the PA in relation to the instructor's grade was calculated, and (2) the level of agreement between a group of children who cannot read and a group of university students was compared. Balderas et al. (2018) introduced a scalable framework for conducting qualitative assessments of collaborative Wiki assignments using AssessMediaWiki (AMW), a tool to facilitate PA in Wikis, and StatMediaWiki (SMW), a monitoring tool for Wikis. Both tools provide the instructor with fine-grained assessment information about student's collaborative work.

Table 2.1 An overview of selected papers

Paper	Discipline/subject	Sample size	Learning tools and platforms*	Application of LA	LA role	PA challenges	PA insights
Andriamiseza et al. (2021)	Multiple	616 students 21 instructors	Elaastic	LA for PA	Tools PA design	Scaling	PA design Student learning Validity and reliability
Babik et al. (2019)	–	–	–	LA of PA data	Comparison	PA evaluation	PA design
Balderas et al. (2018)	Computer science	43 students	MediaWiki, SMW, AMW	LA for PA	Tools PA design	Scaling	PA design
Bjelde and Lindberg (2018)	Science teaching	Not reported	–	LA for PA	PA design	–	PA design
Bridges et al. (2020)	Multiple	13 students	LAMS	LA of PA data	Interaction	–	Student interaction
Chaparro–Peláez et al. (2020)	Multiple	23 instructors	Moodle Workshops; MWDEX	LA for PA	Tools	Lack of tools Scaling	PA design
Cheng and Lei (2021)	Educational technology	24 students	Blackboard	LA for PA LA of PA data	Visualizations Interaction	Facilitating interaction	PA design Student interaction
Chiu et al. (2019)	Surgical education	50 students	dVSS	LA of PA data	Interaction	–	Student learning
Choi et al. (2019)	Data science	456 students (2 studies)	–	LA of PA data	Feedback characteristics	Feedback perception	PA design Feedback perception

(continued)

Table 2.1 (continued)

Paper	Discipline/subject	Sample size	Learning tools and platforms*	Application of LA	LA role	PA challenges	PA insights
Divjak and Maretić (2015)	Enterpreneurship	62 students (2 studies)	Moodle Workshop	LA of PA data	Feedback characteristics	Scaling	Feedback perception Validity and reliability
Djelil et al. (2021)	Multiple	422 students (7 courses)	Sqily	LA of PA data	Interaction	Facilitating interaction	PA design Student interaction
Er et al. (2021a)	Technology enhanced distance learning	14 students	Synergy	LA for PA	Tools Visualizations	Scaling	PA design
Er et al. (2021b)	Primary education	30 students	Synergy	LA of PA data	Interaction	Facilitating interaction	PA design Student interaction
Gunnarsson and Alterman (2014)	Internet and society, human–computer interaction	157 students (2 studies)	Blogosphere by Brandeis University	LA of PA data	Feedback characteristics	Scaling	PA design Student learning Validity and reliability
Huang et al. (2019)	Information management	96 students	Moodle	LA of PA data	Feedback characteristics Interaction	–	PA design Student interaction
Hunt et al. (2021)	Education	76 teachers (2 studies)	EPASS	LA for PA	Automated assessment Visualizations	PA evaluation	PA design Feedback perception

(continued)

Table 2.1 (continued)

Paper	Discipline/subject	Sample size	Learning tools and platforms*	Application of LA	LA role	PA challenges	PA insights
Khosravi et al. (2020)	Relational databases	384 students	RiPPLE	LA of PA data	Feedback characteristics	Lack of tools	PA design Validity and reliability PA comparison
Koh et al. (2016)	Teamwork competency awareness	272 students	ChatAdmin	LA for PA	Visualizations	–	PA design Feedback perception
Lárusson and White (2012)	Computer science	22 students	–	LA for PA	Automated assessment	–	Student learning
Lin (2019)	Language learning	57 students	1Know; Zuvio	LA of PA data	Comparison	PA evaluation	PA design Student learning Student interaction
Misiejuk et al. (2021)	Multiple	7660 records	Peergrade	LA of PA data	Feedback characteristics	Feedback perception	Feedback perception
Mørch et al. (2017)	Language learning	125 students	EssayCritic	LA of PA data	Comparison	PA evaluation	Student learning Student interaction
Nalli et al. (2021)	Physics	54 students	Moodle	LA for PA	Tools	Feedback perception	–
Sedrakyan et al. (2014)	Management information systems	86 students	JMermaid tool	LA of PA data	Interaction	–	Student interaction

(continued)

Table 2.1 (continued)

Paper	Discipline/subject	Sample size	Learning tools and platforms*	Application of LA	LA role	PA challenges	PA insights
Shibani et al. (2019)	Law, accounting	293 students (2 studies)	AcaWriter	LA for PA	Automated assessment	–	Student learning Feedback perception
Vogelsang and Ruppertz (2015)	Cyber security and human rights	467 students	iversity	LA of PA data	Comparison	Scaling	Validity and reliability
Vozniuk et al. (2014)	Social media	100 participants (2 studies)	GRAASP	LA for PA	Tools	Lack of tools Scaling	Validity and reliability

*Only learning tools and platforms reported by name in a paper listed

Automated Feedback. Three papers either compared PA with automated feedback or augmented a PA activity with automated feedback. Hunt et al. (2021) conducted a PA activity with teachers who were divided into two groups that used either an e-portfolio without LA or an e-portfolio enhanced with automated feedback and an activity dashboard. The analysis focused on feedback perceptions among feedback receivers and feedback providers. Lárusson and White (2012) developed a tool with which to automatically measure and visualize an originality score (*Point of Originality)* for students' contributions to the teacher to help with monitoring and evaluation in a student co-blogging activity that included PA. The score was validated in the study. Shibani et al. (2019) showcased an implementation of the *Contextualizable Learning Analytics Design* (CLAD) model with the help of an automated feedback tool, AcaWriter, in two contexts: law essay writing and business report writing. In both contexts, the students engaged in a PA activity and were divided groups that either received additional automated feedback from AcaWriter and did not receive automated feedback. An additional usefulness survey was conducted to compare both groups.

Visualization. Three papers focused on data visualization. Koh et al. (2016) presented a *Team and Self Diagnostic Learning* (TSDL) framework aimed at the teamwork competencies and collaboration skills of students. The framework was implemented during a PA activity in which students rated themselves and other team members in an online survey. The similarity scores between self- and peer-ratings were calculated. The results were visualized as student micro-profiles in a radar chart and shown to the students and teachers for their reflection. Er et al. (2021a) presented an open-source platform, Synergy, designed to support PA based on a *Theoretical Framework of Collaborative Peer Feedback*. One of the platform's features is the visualization of students' activity data for the instructor.

Student interaction. Six papers used LA to analyse PA data and explore topics such as student interaction and engagement. Bridges et al. (2020) combined PA data with video and discourse analyses to examine interprofessional team-based learning. Chiu et al. (2019) used peer observation and assessment data as a proxy for active engagement and evaluated their effects on student progress in surgical training using the da Vinci Skills Simulator (dVSS) platform. Djelil et al. (2021) analysed student interaction data from the learning platform Sqily, which included PA, to detect their engagement patterns, roles, and temporal dynamics. Huang et al. (2019) focused on the effects of gamification and quantity- and quality-based badges on peer feedback quality and student engagement in an online discussion forum. The gamification design was based on the Theory-driven Gamification model (GAFCC: G*oal, Access, Feedback, Challenge, Collaboration*), while the PA data were analysed using content analysis and social network analysis. Er et al. (2021b) applied process mining to identify and interpret engagement patterns in data from the PA platform Synergy. Sedrakyan et al. (2014) examined group interaction data during a conceptual modeling process that included PA.

Feedback characteristics. Five papers focused on peer feedback characteristics, such as perception and quality. Gunnarsson and Alterman (2014) conducted a study on peer promotion, a type of PA, in which students assessed other students work by liking other students' posts or awarding badges. Moreover, students were required to engage weekly in more traditional PA assignments by giving feedback using a 3-point scale on a questionnaire form and commenting on two posts. Khosravi et al. (2020) presented an adaptive platform, RiPPLE, that aims to support evaluative judgement skills and conducted a study in which students created multiple choice questions (MCQs) and gave each other peer feedback on the platform. Both the validity of the peer feedback and the development of peer feedback quality over time were explored in this study. Misiejuk et al. (2021) used a variety of LA methods to analyse the backward-evaluation big data to gain new insights into student perceptions of feedback and its relationship to rubrics. Choi et al. (2019) used natural learning processing to code and analyse the PA text data to determine the influence of the social economic status of students on the perceptions of the PA. Divjak and Maretić (2015) developed and tested a novel method via which to measure PA and self-assessment reliability using modified Manhattan metrics.

Comparison. Four papers compared different types of PA. Vogelsang and Ruppertz (2015) analysed MOOC data derived from the innovative integration of teaching assistants into assessment activities to determine student performance and the validity of this method in relation to PA, automated assessment, and instructor grading. Lin (2019) compared online and paper-based PA to explore the differences in learning achievement, learning involvement (measured using log data from a learning management system), learning autonomy, and student learning reflections. Mørch et al. (2017) generated automated feedback in the EssayCritic system for one group in a language learning scenario and compared their learning performance and writing process with a group that engaged in PA without EssayCritic. Babik et al. (2019) simulated datasets using LA methods to compare ranking-based and rating-based PA with a focus on structural effects.

Design. Two papers focused on designing PA. Bjælde and Lindberg (2018) reported on course design examples incorporating continuous feedback, including PA, and LA. Andriamiseza et al. (2021) explored a two-votes-based process, a form of peer instruction with embedded PA. The results of a learning activity that was conducted on the web platform Elaastic were analysed and presented to the instructors to inform their practice. This study not only provided instructors with recommendations for orchestration but also system designers with recommendations when designing a formative assessment system.

RQ2: What are the reported peer assessment challenges the research addressed with learning analytics? And how are they addressed?

Only 18 papers reported on challenges facing PA *that may be mitigated through LA*, while three papers reported on more than one issue. We identified five main

challenges: *scaling*, *PA evaluation*, *lack of tools*, *feedback perception*, and *facilitating interaction*. In this section, we describe the challenges and their potential mitigation.

Scaling. The scaling of PA was the challenge LA had the most potential to help, as reported in eight papers. As noted by Andriamiseza et al. (2021), the scaling of assessment activities generates rich datasets that may be used to help inform instructor practice. In their study, the data from a two-votes-based process with embedded PA was analysed to inform classroom orchestration. Chaparro-Peláez et al. (2020) noted the need to support MOOCs with efficient student-centered assessment methods, such as PA, which can be made scalable by using LA. To encourage the adoption of PA as a scalable assessment solution for large courses, Vozniuk et al. (2014) used LA to validate PA use on a social media platform, GRAASP, which can be used to set up a PA activity. A PA platform, Synergy, with integrated LA, was presented to facilitate the scaling of dialogic peer feedback in Er et al. (2021a). Gunnarsson and Alterman (2014) noted that students' content production in blogging environments may overwhelm instructors and lead them to not being able to identify and highlight high-quality contributions to the class. This was mitigated by the implementation of peer promotion, a type of PA that uses likes and badges. Although Wikis provide rich data that may be used to evaluate various skills, the assessment of Wikis is very complex and difficult to scale. To address this, Balderas et al. (2018) gave teachers information from qualitative and quantitative LA-supported assessment during a PA activity using Wikis. Divjak and Maretić (2015) described the need to use LA data to explore PA and assess the reliability and validity of PA, especially in large classrooms. For example, they noted that LA could help with equalizing in a PA activity—such as students giving all their peers the same marks—by discovering assessment patterns. A second example addresses students' lack of the metacognitive skills needed to perform PA, which may be mitigated by using LA to calculate PA reliability, which would enable teachers to identify students who needs help.

PA evaluation. Four papers identified the challenge of evaluating PA as an activity. Because online PA has the potential to facilitate higher-order thinking, such as improving writing abilities in language learning, and can be used as an effective flipped-classroom strategy, Lin (2019) studied the differences between online and paper-based PA. LA data from a learning management system were used as a proxy for student's learning involvement in both scenarios. Mørch et al. (2017) noted that LA-generated automated feedback may be as accurate and reliable as PA. At the same time, these systems could lead to conformity and less creativity in writing. To explore these issues, a study was conducted that compared the learning performance and writing processes of students who received automated feedback with those of students who only received feedback from their peers. Babik et al. (2019) observed that comparing different PA methods using real-life assessment data may conflate the analysis with cognitive and behavioural effects. To mitigate this phenomenon and focus on structural effects, a simulation model of PA was developed using a Monte-Carlo simulation, and network typology and aggregation

methods were used to compare ranking-based and rating-based PA. Hunt et al. (2021) report a potential advantage derived from adding LA to e-portfolios used in a PA activity: providing more tailored and timely feedback.

Lack of tools. Four papers described the lack of PA tools. Nalli et al. (2021) described a lack of tools that support the formation of heterogenous groups of students for PA. To address this, a variety of clustering algorithms using Moodle activity data were evaluated, and a Moodle plugin for group formation was developed and validated. Chaparro-Peláez et al. (2020) reported that there are few software tools to support PA. Moreover, the current Moodle Workshops version has many limitations in terms of data visualization, extraction, and exporting. As a solution, a new tool with LA functionalities, Moodle Workshop Data Extractor (MWDEX), was presented. The development of a PA extension for the social media platform GRAASP by Vozniuk et al. (2014) was motivated by the lack of ready-to-use PA platforms and PA validity issues. Many tools do not enable data harvesting, so the impact of implemented strategies cannot be evaluated. Khosravi et al. (2020) presented an adaptive tool, RiPPLE, that enables data extraction and fosters evaluative judgements. In an empirical study focusing on PA validity, students developed and peer-assessed multiple-choice questions (MCQs).

Feedback perception. Three papers recognized improving peer feedback perceptions as an important PA challenge. The current application of the Moodle Workshop randomly forms student groups for a PA activity, which negatively influences student satisfaction with the assessment activity. This motivated Nalli et al. (2021) to propose a sophisticated quantitative LA method and a Moodle plugin to form heterogenous groups, the implementation of which may lead to more positive perceptions of PA and higher success rates for all students in a class. Misiejuk et al. (2021) identified a challenge in understanding student perceptions of the feedback they received with regards to being able to use it effectively. To address this challenge, an extensive study that used a large dataset and applied a variety of LA methods (ENA, regression, and other methods) was conducted. Choi et al. (2019) described a need to understand the impact of socio-economic status on how student feedback is perceived. As a part of their analysis intended to gain more insights into this problem, they used automated text classification, an LA technique, to detect feedback characteristics.

Facilitating interaction. Three papers identify facilitating student interaction in PA as a problem and suggest that LA could help. Cheng and Lei (2021) identified the need to facilitate student interactions in blogging activities that include PA. Social network analysis (SNA), an LA technique, was used to analyse and visualize student engagement and group cohesion. The SNA graphs were shown to students, and their effect on student behaviour was explored. Djelil et al. (2021) noted that engaging students in PA is difficult and that PA itself is prone to biases. To gain more insights into student interactions, social network analysis, specifically a graphlet-based method, and clustering were used to analyse the PA data. Er et al. (2021b) noticed a challenge in understanding student engagement patterns

that could be used to improve PA. To identify these patterns, log data from a PA platform, Synergy, was analysed using process mining.

RQ3: What insights into peer assessment can we gain from learning analytics?
Only one paper did not report any insights into PA. We found five types of PA insights, which were *PA design*, *student learning*, *PA validity and reliability*, *student interaction*, and *feedback perception.*

PA Design. Most papers contributed new or improved designs for a PA activity with the help of LA, or their insights could inform more effective PA designs. The adaptive platform RiPPLE, presented by Khosravi et al. (2020), provides a learning environment that supports evaluative judgement and PA. Moreover, the tool enables the measurement and evaluation of such interventions. The theory-oriented design of the PA platform Synergy, presented in Er et al. (2021a), was evaluated positively by a group of students. While comparing online and paper-based PA, Lin (2019) noted the students' frustration with small screens in the online PA group when engaged with PA on mobile devices. An evaluation of the Workshop Data EXtractor (MWDEX), developed by Chaparro-Peláez et al. (2020), indicated that instructors typically do not use any software tool to facilitate PA. Moreover, although most instructors use Moodle in their day-to-day practice, they choose Blackboard's PA application rather than Moodle Workshops when they decide to use software to support PA, which may indicate their dissatisfaction with the Moodle Workshop module for PA.

A radar chart visualizing the similarity scores between self- and peer-ratings in a team awareness activity, presented in Koh et al. (2016), was perceived positively as a visualization tool, although the students had difficulties interpreting the similarity scores. The need for a more user-friendly dashboard was emphasized, and because some students and teachers found the PA ratings dishonest in the team awareness study, more training in PA was recommended. Cheng and Lei (2021) found that, when an interaction graph of within-group interactions was shown to students after the first PA activity, this had the undesired effect of generating fewer cross-group comments in the following cycles. This indicates that a clearer explanation of performance expectations is needed to help students interpret the visual analytics of their behaviour.

The finding that PA rating scales outperformed PA ranking scales according to a study conducted by Babik et al. (2019) can be used to design PA activities and systems because choosing either scale must be considered together with other design choices that they may influence, positively or negatively, either scale's validity and reliability.

Bjælde and Lindberg (2018) presented a course design that integrated PA and LA to facilitate assessment as learning and continuous feedback as an early intervention method. Student feedback perceptions after the PA activity guided the future course design. A scalable qualitative assessment framework that uses LA, developed by Balderas et al. (2018), can help teachers with the large-scale assessment, including PA, of collaborative Wiki contributions. Andriamiseza et al. (2021)

recommended that formative assessment systems based on a two-votes-based process show teachers the proportion of correct answers at the first vote, as well as the correlation between the correctness of a student's rating and their confidence level. In addition, they recommend that PA activities not include self-ratings, and that the system should be flexible in terms of how many peers assess one another.

Gamification is cost effective and relatively easy to implement and likely increases PA engagement in online discussion forums, as shown in Huang et al. (2019). Peer promotion using badges and likes may be considered as an addition to traditional PA to reduce instructors' workload, as described in Gunnarsson and Alterman (2014).

Er et al. (2021b) found that high-performing students had many bidirectional transitions between self-regulated learning and socially regulated learning, as well as between self-regulated learning and co-regulated learning on a PA platform, Synergy. One implication of this behaviour that may lead to better student performance is that additional support for engaging students in self-regulated learning, socially regulated learning, and co-regulated learning should be provided. Hunt et al. (2021) compared a group using an e-portfolio with LA visualizations and a group using an e-portfolio without LA. Both groups indicated a need for a face-to-face discussion as a part of the feedback process. Teachers in the e-portfolio with the LA group indicated that they need more support in dealing with analytics due to a lack of digital skills. Furthermore, they expressed a need to have more control over the visual analytics of their activities because the teachers felt overwhelmed at times. Djelil et al. (2021) used social network analysis (a form of LA) with data from the learning platform Sqily and found that teacher presence was significant across courses and crucial to initiating initial PA activities, suggesting that students may need support and direct guidance from a teacher to begin interacting with peers. The finding by Choi et al. (2019) that students reacted differently to feedback provided by students with different socioeconomic statuses (i.e., based on the nationality of the peer feedback provider) has design implications. Instructors must pay attention to which information about learners is visible to others, including indirect information that may indicate socioeconomic status, such as a name or profile picture. On the other hand, socioeconomic information may help instructors pair students with different socioeconomic statuses and thus ensure exposure to different perspectives.

Student learning. Six papers reported insights into student learning. Lin (2019) found no learning performance difference between online and paper-based PA in a flipped language-learning class. However, the online PA group expressed more ideas in their work, expressed more interest in the flipped learning environment, and showed higher learner autonomy during previewing before the class. Mørch et al. (2017) found no significant difference in learning performance between a group using automated feedback and a group engaged in PA. The group that used automated feedback, however, used significantly more subthemes and showed more ideas inspired by the automated feedback in their writing. It was difficult for students in the PA group to give content-oriented feedback, and they preferred to

comment on the essay structure. A group with additional automated feedback used significantly more rhetorical moves in their essays in the first context in Shibani et al. (2019). Furthermore, PA helped students with sense-making regarding the automated feedback.

Lárusson and White (2012) found a statistically significant positive correlation between the number of contributions that included comments on other students' blogs and students' final performance and the originality of their contributions. Chiu et al. (2019) found that implementing peer observation with PA during surgical student practice on a da Vinci Skills Simulator (dVSS) facilitated the improved performance of intermediate-level surgical tasks but not basic or advanced tasks. In a study of the two-votes-based PA process, Andriamiseza et al. (2021) found that benefits of formative assessment sequences increased when (1) the proportion of correct answers is close to 50% during the first vote or (2) the written rationales from students who gave correct answers are better rated than those from students with incorrect answers. However, the number of peer ratings made no significant difference in terms of the benefits of the formative sequences' benefits.

Reliability and Validity. Four papers described findings about PA reliability and validity. Vogelsang and Ruppertz (2015) found peer and teaching assistants' grading invalid as compared with the expert's grading. However, peer grading was valid, assuming that the teaching assistants' grading was accurate. Andriamiseza et al. (2021) established that peer ratings were consistent when correct learners were more confident than incorrect ones, while self-ratings were inconsistent in the peer rating context. Khosravi et al. (2020) established a strong and positive correlation between student and domain expert ratings on multiple choice questions (MCQs) on an adaptive platform, RiPPLE. Furthermore, the difference between the domain expert ratings and peer ratings decreased with time and practice. Gunnarsson and Alterman (2014) found that peer promotion helped identify higher quality posts. Some students could be identified as more reliable in evaluating post quality than others. Moreover, badges given before or after the traditional PA activity were found to be more reliable than those given during the PA activity. The evaluation of the GRAASP extension developed by Vozniuk et al. (2014) showed a strong agreement between the grades assigned by students and instructors in rating-based PA. To confirm that students did not grade the reports based on appearance, a second experiment was conducted with children, who rated the reports only based on their appearance, without reading the reports' content. Little agreement between the grades assigned by the students and children was found, and this result confirmed that students engaged with the content of the reports before grading them.

Divjak and Maretić (2015) developed a reliability measurement based on the modified Manhattan metrics (based on taxicab norm) indicating that reliable peer grading should be within 2 points (i.e., less than or equal to 2), while peer grading would be unreliable if it exceeded 2 points. In their case study, the PA grades were reliable.

Student interactions. Eight papers provided new insights into student interactions and behaviour. Students in the online PA group demonstrated higher learning involvement during flipped learning than the paper-based PA group in the study by Lin (2019). Mørch et al. (2017) noted that students in the automated feedback group were more motivated and worked harder on their essays than the PA group. The gamification-based group posted more, engaged more in PA, and gave higher-quality peer feedback in an online discussion forum than the control group in the study reported in Huang et al. (2019). A larger group of students in the gamification-based group provided feedback in comparison to the control group. After showing students their social network graphs on their intra-group blogging and PA behaviour, in the study reported in Cheng and Lei (2021), the interactions within the same group increased and the exploration of outside-group blogs decreased. This resulted in a clear subgroup structure.

Sedrakyan et al. (2014) found that both the best- and the worst-performing students were more engaged in their modeling activities just before the activity deadlines, including the PA deadline. However, the best performing groups were also very active between the deadlines. Moreover, the best-performing groups implemented more peer feedback in their models in comparison to the worst-performing groups during the conceptual modeling activity. Er et al. (2021b) found that high-performing students were more likely to engage as described in the theoretical framework for collaborative peer feedback, while medium-performing students deviated from the theory.

Djelil et al. (2021) found a positive trend in terms of learners engaging in PA activities on the learning platform Sqily. Furthermore, it was found that students may need some time to feel comfortable providing feedback to new peers. Bridges et al. (2020) compared the video, discourse, and PA data of two groups during an interprofessional team-based learning activity. According to an analysis of the PA data, the first group did not identify a leader, and their physical orientation was spatially and interactionally cohesive. The second group identified a strong leader, both in their PA and in their spatial composition.

Feedback perception. Five papers reported findings on feedback perception. Feedback providers, reported on in Hunt et al. (2021), found that giving feedback to others helped them reflect on their own work. At the same time, they felt uncomfortable being critical toward their colleagues. This influenced feedback receivers and their perception of feedback providers as not always being honest. Another finding in this study was that the group using an e-portfolio with LA had significantly more positive perceptions of the entire feedback experience than the group using an e-portfolio without LA. However, there were no significant differences in the perceptions of the quantity, quality, and use of feedback between the two groups. The PA activity in Divjak and Maretić (2015) was overwhelmingly perceived as motivating. In the team awareness study of Koh et al. (2016), some students and teachers disagreed with PA ratings and perceived them as dishonest. Students who perceive feedback as useful acknowledged their errors, expressed the intention to revise their text, and/or gave praise regarding the feedback in their

backward-evaluation comments, as shown in a study by Misiejuk et al. (2021). Students who evaluated the feedback that they received as not useful or showed confusion about the feedback were critical toward it and/or disagreed with it in their backward-evaluation comments. In general, students wanted feedback to be more specific, just, and constructive, rather than kind. No significant relationship between backward-evaluation and the structure of the PA rubric was found. Two studies by Shibani et al. (2019) found that the students in both studies perceived the writing activity with added automated feedback to be more useful than only PA without automated feedback. High-socioeconomic status students reacted differently to the feedback from medium- and low-socioeconomic status students in terms of feedback agreement and formality when status information was disclosed in a study reported in Choi et al. (2019).

2.5   Discussions and Conclusions

This chapter presents the first scoping review mapping of the LA applications in PA research. The review included 27 very diverse papers, which made reporting on the results challenging. Our research questions focused on the PA challenges that the papers identified and how were they addressed using LA, the role of the LA application in the PA activity, and, finally, the kinds of PA insights reported. We found two main areas in which learning analytics was used for PA: using LA to improve PA activity and using LA to analyse PA data.

We found that most research focused on addressing the challenge of scaling PA, developing new PA tools enhanced by analytics, or attempting to inform PA theory by evaluating different types of PA. Many insights from the research reported in the included papers may inspire new PA designs or improve existing ones by paying heed to reports of successful and unsuccessful implementations of LA in PA activities. In addition to the traditional PA research focus, such as validity and reliability and student learning, interesting studies were conducted on student interaction in PA, in which self-regulation, group building, and student interaction are analysed. Moreover, rich data from gamification-enhanced PA and collaborative writing in blogs or Wikis are utilised to gain more dynamic insights into students' development of feedback skills and learning.

This study has certain limitations. First, the inclusion/exclusion process was difficult, and perhaps, some papers that should have been included were excluded. It was challenging to define which papers were actually using LA because some papers used LA methods without the authors describing them in their papers. Thus, instead of evaluating the "LA-ness" of the papers, we used a proxy that defined a paper as being about using LA in PA research if that paper described LA or was published at the LAK conference or in the Journal of Learning Analytics. Furthermore, a significant group of papers were excluded because they focussed on insights into LA, rather than PA. For example, PA data contributed to the final grade, which was a part of a dataset analysed using LA to identify patterns of epistemic emotions in MOOCs (Han et al., 2021) or to predict time-on-task estimation

strategies (Kovanović et al., 2015). Future reviews might focus on PA data as a part of big datasets and how are they explored using LA; this topic was outside the scope of this chapter, but we found many papers addressing this issue.

Second, PA may be a part of many learning activities, such as student interactions in a discussion forum, but it is not always conceptualised as PA or analysed as such. We tried our best to include a variation of PA implementations, but with the large number of papers found in the search, this was not a trivial task. Finally, the diversity of the papers made the analysis challenging because it was difficult to identify the same issue across them, which may have led to some simplifications in our analysis of the papers and their insights into PA.

Several areas for further research were identified in this review. First, more work is needed to use insights from LA to improve the PA activity before once again using LA to see if there has been improvement (*cf.* Clow, 2012). Some of the papers in our review included two studies, however, the results from first study which gave insights into PA did not lead to a second study that used those insight to improve PA. Second, the automatization of aspects of PA (e.g., feedback classification; Wahid et al., 2016; Ryan et al., 2019) was identified as a potential application for LA, and though there are papers in our review that attempted to automate, the examples are few. Moreover, it seems that the automated methods, such as automated assessment, were used to compare with PA without an automated method, rather than using it as part of a PA process. Thus, more research is needed to improve the automation of aspects of PA through, for example, additional text quality measurements or group formation, as well as empirical studies of their implementations in teaching practice. Third, we found that the focus on either analysing PA data to gain insights into PA or trying to improve PA in practice is limiting, although necessary in some cases. Future research should investigate combining the two and using the LA insights directly to improve PA activity and tools. Fourth, as found in this review, showing analytics to students influences their behaviour, which in turn may be used as a powerful pedagogical tool within PA; however, more work in this area is needed both to understand the effect of analytics on students and teachers/instructors and how the analytics could be integrated into a PA activity. Finally, the analytics used to analyse data in the studies reported in the papers are significantly more advanced than the analytics currently available in the PA tools. A sensible integration of advanced analytics in the PA tools is another promising research area that would include not only technical aspects, but also would include examining the perception and understanding of the analytics by both students and instructors.

This review has shown that LA has the potential better understand and improve PA activity through new insights into student behaviour and the artefacts that they produce, interpersonal and intergroup interactions, or tool improvement. However, the research is still emerging and scattered. LA gives access to hidden data and finding patterns and insights in data from PA activity that is not easily accessible to humans. We hope that this review will act as a starting point for future work on using learning analytics to improve peer assessment activity.

References[1]

Andriamiseza, R., Silvestre, F., Parmentier, J. F., & Broisin, J. (2021). Recommendations for orchestration of formative assessment sequences: A data-driven approach. In *Proceedings of the EC-TEL'21* (pp. 245–259).*

Archer, E., & Prinsloo, P. (2020). Speaking the unspoken in learning analytics: Troubling the defaults. *Assessment and Evaluation in Higher Education, 45*(6), 888–900.

Babik, D., Stevens, S., & Waters, A. E. (2019). Comparison of ranking and rating scales in online peer assessment: Simulation approach. In *Proceedings of the LAK'19* (pp. 205–209).*

Balderas, A., Palomo-Duarte, M., Dodero, J. M., Ibarra-Sáiz, M. S., & Rodríguez-Gómez, G. (2018). Scalable authentic assessment of collaborative work assignments in wikis. *International Journal of Educational Technology in Higher Education, 15*(1), 1–21.*

Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. A. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review, 37*, 100489. https://doi.org/10.1016/j.edurev.2022.100489.

Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning, 19*(1–2), 205–220.

Bjælde, O. E., & Lindberg, A. B. (2018). Using continuous assessment with feedback loops to generate useful data for learning analytics. In *Proceedings of the ASCILITE'18* (pp. 53–62).*

Bridges, S. M., Chan, L. K., Chen, J. Y., Tsang, J. P., & Ganotice, F. A. (2020). Learning environments for interprofessional education: A micro-ethnography of sociomaterial assemblages in team-based learning. *Nurse Education Today, 94.*

Buckingham Shum, S., & Ferguson, R. (2012). Social learning analytics. *Journal of Educational Technology and Society, 15*(3), 3–26.

Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y. S., Gašević, D., & Mello, R. F. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence*, 2.

Chaparro-Peláez, J., Iglesias-Pradas, S., Rodríguez-Sedano, F. J., & Acquila-Natale, E. (2020). Extraction, processing and visualization of peer assessment data in Moodle. *Applied Sciences, 10*(1), 163.*

Chaudy, Y., & Connolly, T. (2019). Specification and evaluation of an assessment engine for educational games: Integrating learning analytics and providing an assessment authoring tool. *Entertainment Computing*, 30.

Cheng, J., & Lei, J. (2021). A description of students' commenting behaviours in an online blogging activity. *E-Learning and Digital Media, 18*(2), 209–225.*

Chiu, H. Y., Kang, Y. N., Wang, W. L., Chen, C. C., Hsu, W., Tseng, M. F., & Wei, P. L. (2019). The role of active engagement of peer observation in the acquisition of surgical skills in virtual reality tasks for novices. *Journal of Surgical Education, 76*(6), 1655–1662.*

Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. In *Proceedings of the LAK'12* (pp. 134–138).

Choi, H., Dowell, N., Brooks, C., & Teasley, S. (2019). Social comparison in MOOCs: Perceived SES, opinion, and message formality. In *Proceedings of the LAK'19* (pp. 160–169).*

Cope, B., & Kalantzis, M. (2016). Big data comes to school: Implications for learning, assessment, and research. *aera Open*, 2(2).

Deeva, G., Bogdanova, D., Serral, E., Snoeck, M., & De Weerdt, J. (2021). A review of automated feedback systems for learners: Classification framework, challenges and opportunities. *Computers and Education*, 162.

---

[1] *indicates papers included in the scoping review.

Divjak, B., & Maretić, M. (2015). Learning analytics for e-assessment: The state of the art and one case study. In *Proceedings of the CECIIS'15.**

Djelil, F., Brisson, L., Charbey, R., Bothorel, C., Gilliot, J. M., & Ruffieux, P. (2021). Analysing peer assessment interactions and their temporal dynamics using a graphlet-based method. In *Proceedings of the EC-TEL'21* (pp. 82–95).*

Ellis, C. (2013). Broadening the scope and increasing the usefulness of learning analytics: The case for assessment analytics. *British Journal of Educational Technology, 44*(4), 662–664.

Er, E., Dimitriadis, Y., & Gašević, D. (2021a). Collaborative peer feedback and learning analytics: Theory-oriented design for supporting class-wide interventions. *Assessment and Evaluation in Higher Education, 46*(2), 169–190.*

Er, E., Villa-Torrano, C., Dimitriadis, Y., Gašević, D., Bote-Lorenzo, M. L., Asensio-Pérez, J. I., Gómez-Sánchez, E. & Martínez Monés, A. (2021b). Theory-based learning analytics to explore student engagement patterns in a peer review activity. In *Proceedings of the LAK'21* (pp. 196–206).*

Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., Rienties, B., Ullmann, T., & Vuorikari, R. (2016). *Research evidence on the use of learning analytics: Implications for education policy*. Joint Research Centre.*

Gunnarsson, B. L., & Alterman, R. (2014). Peer promotions as a method to identify quality content. *Journal of Learning Analytics, 1*(2), 126–150.*

Han, Z. M., Huang, C. Q., Yu, J. H., & Tsai, C. C. (2021). Identifying patterns of epistemic emotions with respect to interactions in massive online open courses using deep learning and social network analysis. *Computers in Human Behavior*, *122.*

Huang, B., Hwang, G. J., Hew, K. F., & Warning, P. (2019). Effects of gamification on students' online interactive patterns and peer-feedback. *Distance Education, 40*(3), 350–379.*

Hunt, P., Leijen, Ä., & van der Schaaf, M. (2021). Automated feedback is nice and human presence makes it better: Teachers' perceptions of feedback by means of an e-portfolio enhanced with learning analytics. *Education Sciences*, *11*(6).*

Khosravi, H., Gyamfi, G., Hanna, B. E., & Lodge, J. (2020). Fostering and supporting empirical research on evaluative judgement via a crowdsourced adaptive learning system. In *Proceedings of the LAK'20* (pp. 83–88).*

Knight, S. (2020). Augmenting assessment with learning analytics. In *Re-imagining University Assessment in a Digital World* (pp. 129–145).*

Koh, E., Shibani, A., Tan, J. P. L., & Hong, H. (2016). A pedagogical framework for learning analytics in collaborative inquiry tasks: An example from a teamwork competency awareness program. In *Proceedings of the LAK'16* (pp. 74–83).*

Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R. S., & Hatala, M. (2015). Penetrating the black box of time-on-task estimation. In *Proceedings of the LAK'15* (pp. 184–193).

Lárusson, J. A., & White, B. (2012). Monitoring student progress through their written "point of originality." In *Proceedings of the LAK'12* (pp. 212–221).*

Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science, 5*(69), 1–9.

Lin, C. J. (2019). An online peer assessment approach to supporting mind-mapping flipped learning activities for college English writing courses. *Journal of Computers in Education, 6*(3), 385–415.

Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279–290.

Misiejuk, K., & Wasson, B. (2021). Backward evaluation in peer assessment: A scoping review. *Computers and Education*, *175.*

Misiejuk, K., Wasson, B., & Egelandsdal, K. (2021). Using learning analytics to understand student perceptions of peer feedback. *Computers in Human Behavior*, *117.**

Mørch, A. I., Engeness, I., Cheng, V. C., Cheung, W. K., & Wong, K. C. (2017). EssayCritic: Writing to learn with a knowledge-based design critiquing system. *Journal of Educational Technology and Society, 20*(2), 213–223.*

Nalli, G., Amendola, D., Perali, A., & Mostarda, L. (2021). Comparative analysis of clustering algorithms and Moodle plugin for creation of student heterogeneous groups in online university courses. *Applied Sciences, 11*(13).*

Nyland, R. (2018). A review of tools and techniques for data-enabled formative assessment. *Journal of Educational Technology Systems, 46*(4), 505–526.

Pardo, A. (2018). A feedback model for data-rich learning experiences. *Assessment and Evaluation in Higher Education, 43*(3), 428–438.

Ryan, T., Gašević, D., & Henderson, M. (2019). Identifying the impact of feedback over time and at scale: Opportunities for learning analytics. In *The impact of feedback in higher education* (pp. 207–223).

Saqr, M. (2017). Assessment analytics: The missing step. *International Journal of Health Sciences, 11*(1).

Sedrakyan, G., Snoeck, M., & De Weerdt, J. (2014). Process mining analysis of conceptual modeling behavior of novices–empirical study using JMermaid modeling and experimental logging environment. *Computers in Human Behavior, 41*, 486–503.*

Shibani, A., Knight, S., & Buckingham Shum, S. (2019). Contextualizable learning analytics design: A generic model and writing analytics evaluations. In *Proceedings of the LAK'19* (pp. 210–219).*

Vogelsang, T., & Ruppertz, L. (2015). On the validity of peer grading and a cloud teaching assistant system. In *Proceedings of the LAK'15* (pp. 41–50).*

Vozniuk, A., Holzer, A., & Gillet, D. (2014). Peer assessment based on ratings in a social media course. In *Proceedings of the LAK'14 (*pp. 133–137).*

Wahid, U., Chatti, M. A., & Schroeder, U. (2016). Improving peer assessment by using learning analytics. In *DeLFI Workshops* (pp. 52–54).

Support Student Integration of Multiple Peer Feedback on Research Writing in Thesis Circles

3

Ya Ping Hsiao and Kamakshi Rajagopal

3.1 Introduction

3.1.1 Student Peer Review and Feedback in Thesis Circles

The didactic principles of collaborative learning, peer learning, and the process and social interaction of writing are becoming increasingly important in Dutch Higher Education (HE), with the uptake of undergraduates' theses at the exit level (Elbow, 1998; Rajagopal et al., 2021; Romme & Nijhuis, 2002). Following these, peer review, defined as "an instructional writing activity in which students read and provide commentary on one another's writing, and the purpose of this activity is to help students improve their writing and gain a sense of audience" (Breuch, 2004, p. 1), has been an important learner-centered activity in the context of thesis circles, a form of group supervision, in which a number of students are supervised under one or two academic supervisors in the process of writing their graduation thesis (Rajagopal et al., 2021; Romme & Nijhuis, 2002). In thesis circles, students often receive feedback from multiple peers to compensate for little and targeted supervisor feedback (Romme & Nijhuis, 2002). Starting from student independent work and critical thinking, students de facto act as non-formal co-supervisors of their peers and co-regulate each other's learning (Romme & Nijhuis, 2002).

Y. P. Hsiao (✉)
Teacher Development, Tilburg University, Room 406, Academia Building, Warandelaan 2, 5037 AB Tilburg, The Netherlands
e-mail: y.p.hsiao@tilburguniversity.edu

K. Rajagopal
Itec, Imec Research Group and Educational Development Unit, KU, Leuven, Belgium

Reviewing each other's work helps students make sense of the quality criteria of academic writing and this understanding in turn helps them reflect on their own writing and increases the potential to improve their writing products (Cho & MacArthur, 2011; Huisman et al., 2018; Nicol et al., 2014; Noroozi et al., 2023). One challenge for students is the integration of multiple information sources by considering the contextual constraints and personal stand, which is an emerging theme of critical thinking in higher education (Elder & Paul, 2009; Facione, 2011).

3.1.2   Multiple Peer Feedback and the Need for Student Support

As suggested in the large-scale assessment literature, involving students in giving each other peer feedback is a cost-effective solution to compensate for supervisor feedback (Broadbent et al., 2018). However, the quality of peer feedback varies. Compared to teacher feedback based on profound didactic and content expertise (Gielen et al., 2010), peer feedback is not always treated seriously because students are uncertain of feedback quality from their equals (Latifi et al., 2021; Taghizadeh et al., 2022). In addition, students do not feel obliged to use peer feedback because there is no consequence on their grades if they do not use feedback in their revision (Zhao, 2010). To deal with this, involving multiple peers to give feedback seems to be a solution because applying a four-eyes principle is likely to ensure feedback quality. Students also suggest having "more reviews as then you had a better chance of getting one of good quality" (Nicol et al., 2014, p. 109).

Research of peer feedback and epistemological understanding suggests that students need training and support on how to deal with feedback made from reviewers with multiple perspectives and with different research interests and foci (Falchikov, 2013; Kuhn, 2020). This support can concern the quality of peer feedback, and on how students engage in deep processing of feedback (Ajjawi et al., 2021; Berndt et al., 2018) as well as how they integrate multiple feedback into a coherent set of suggestions for improving their writing.

3.1.3   Students Need Support on Assessing Feedback Quality

Regarding feedback quality, literature states that students attribute high-quality feedback to be attentive to their own work and to show emotions with detailed suggestions that are useful for them to make improvements on the subsequent tasks (Dawson et al., 2019). Training activities for students to give peer feedback is therefore often based on these quality criteria (Hsiao et al., 2015; Nicol & McCallum, 2021), but how students should judge the quality of received feedback has received less attention. As pointed out by recent research on feedback literacy, students need to develop "the understandings, capacities and dispositions needed to make sense of information and use it to enhance work or learning strategies" (Carless & Boud, 2018, p. 1316). Without an appropriate level of feedback literacy, it is difficult for students to judge the quality of peer feedback and determine

which feedback is useful for their own task improvement, especially when students do not have sufficient criterion knowledge (i.e., how well quality work should look like) of quality feedback and integration strategies of multiple feedback.

3.1.4 Students Need Support on Integrating Multiple Feedback

As for student deep engagement with received feedback, recent attention has focused more on supporting students to transform external feedback (from teachers or peers) to their internal feedback, which is defined as "the new knowledge that students generate when they compare their current knowledge and competence against some reference information" (Nicol, 2021, p. 2). This theme aims to draw attention to the ultimate goal of feedback practices: to enhance student learning. The notion of generating new knowledge requires students to engage in higher order thinking skills, such as analysis, evaluation and synthesis. According to Nicol's model, various types of external reference information can stimulate students to generate internal feedback (Nicol, 2021). The most effective one is comparing their own work with others. This kind of comparative judgment against concrete external reference information (others' work) is analogical/holistic, reasoning from what is known about one exemplar or case to infer new information about another exemplar or case (Gentner et al., 2001). Analogical comparisons are different from analytical comparisons based on rubric consisting of criteria and standards, which students perceive as abstract and difficult (Nicol, 2021; Sadler, 2009). Although comparative judgement seems to be easier for students to generate internal feedback (Nicol, 2021), its validity that justifies the rationales of these judgments, still needs more research in peer feedback studies (Nicol & McCallum, 2021). In addition, when doing comparative judgement, it can be difficult for students to "identify the shared principles and rational structures" (Nicol, 2021, p. 6) which require higher order thinking skills (analysis and synthesis) to generate new knowledge (creation) and to improve their own work. Therefore, students need guidance to generate high quality internal feedback (e.g., using prompt questions to process and uptake feedback) from external multiple peer feedback. Also, learning activities should bridge the gap between student internal feedback and how to use new knowledge in the revision, to improve their own work.

3.1.5 Integrating Multiple Peer Feedback: Developing Instructional Design for a Complex Student Activity

Before integrating multiple peer feedback, students need to make evaluative judgements of feedback quality based on multiple assessment criteria of feedback content and form. They also need to organize multiple interpretations of their own work into a coherent action plan, based on task and personal learning goals. This integration consists of multiple comparative analyses and multiple relation constructions among different components of student work and multiple assessment

criteria (Kuhn, 2020). These processes, without support, can overload students, especially for those who are not yet developed to deal with multiple perspectives (Kuhn, 2020).

Although several didactic strategies in peer feedback studies are proposed to guide the student process of feedback (Banihashem et al., 2022; Latifi & Noroozi, 2021), they mainly focus on a single feedback source, either from the teacher or one peer at a time (Falchikov, 2013; Nicol & McCallum, 2021; Winstone et al., 2017a, 2017b). In addition, students' uptake of peer feedback and their efficiency of using peer feedback to improve her or his own work still needs more research. Some authors have advocated to embed these feedback processing in a broad context of course instructional design (Berndt et al., 2018; Dawson et al., 2019; Mercader et al., 2020). Taken all together, this chapter aims to build such an instructional design to support student integration of multiple peer feedback in a thesis circle context, drawing on academic knowledge in feedback literacy research and epistemological understanding.

3.2   Methodology

We follow the paradigm of Educational Design Research (McKenney & Reeves, 2014) to develop our instructional design for the integration of multiple feedback. In particular, we used a design conjecture mapping approach to identify conjectures (i.e., "unproven propositions that are thought to be true" [McKenney & Reeves, 2014, p. 32]) and theoretical principles (e.g., students need support and structure before doing peer review and feedback) for the specific instructional activities of multiple peer feedback on written work. We mapped out "how they are predicted to work together to produce desired outcomes" (Sandoval, 2014, p. 19).

A conjecture map is made to illustrate the salient design elements and how these elements function together to achieve the desired outcomes. Before identifying design characteristics (i.e., dimensions, elements and principles), we carried out a problem analysis by examining the complexities of undergraduate thesis writing, and looked at the student cognitive developmental stage to describe the challenges faced by undergraduate students when dealing with multiple peer feedback in a specific context of thesis circles. Through this analysis, we identified important needs for specific structure, scaffolding and learning activities. Based on the literature study, we formulated design questions and identified design conjectures to understand which features we need to integrate and which outcomes we aim to achieve.

Based on this conjecture map, we described an integrated instructional design that supports students to deal with multiple peer feedback, including sense-making and uptake of feedback. Our design then becomes synthesis of the theories and studies from feedback literacy, integration of multiple texts in reading comprehension, and cognitive processing and biases in decision making processes. We describe the theoretical and empirical research base underlying each stage of this design.

3.2.1   Complexities and Challenges of Multiple Peer Feedback Practices

A graduation thesis is perceived by students as the most challenging academic work in their bachelor's program because it requires a greater degree of independent learning than previous assessments in the program curriculum (Huang, 2010; Todd et al., 2004). An undergraduate graduation requires students to use critical thinking, research, and writing skills for a specific problem statement or research question. It requires students to take responsibility and work independently in making decisions about the choice of thesis subject and supervisor, setting goals and making personalized planning, monitoring own progress and evaluating quality (Todd et al., 2004). The supervisor plays the central role in guiding and supporting this independent learning process, in a way that balances student autonomy and guidance (de Kleijn et al., 2012; Todd et al., 2004). Unfortunately, it is not easy to find an appropriate balance, because most senior undergraduates still rely on authority (i.e., supervisors, tutors, more competent peers) to deal with uncertainty arising from decision making and carrying out the tasks (Baxter Magolda, 2001). Independent learning becomes even more challenging in thesis circles, because students are supposed to co-supervise their peers (Romme & Nijhuis, 2002) while they are each other's equals and everyone works on a different topic (within a shared theme) and while they work on their own topic and thesis.

From the perspective of epistemological development, independent inquiry requires students to reach the stage of contextual relativism or become evaluativists (i.e., both terms are used interchangeably in the following texts) that they know some solutions are better than others, depending on context (Hofer & Pintrich, 1997; Kuhn, 2020). Students need to go beyond the lower stages of dualism (seeing solutions are correct or wrong) and multiplicity (seeing each solution takes a different perspective). Instructing students to actively engage in critical reflection, perspective taking, and sense-making is likely to develop them to the stage of contextual relativism (Baxter Magolda, 2001; King & Kitchener, 2002; Moore, 2002).

In terms of writing a bachelor's thesis, students are supposed to achieve contextual relativism (Moore, 2002): to judge an argument by its reasoning and supporting evidence, and consistency of how the argument is made within a certain context (King & Kitchener, 2002), to determine the most reasonable or probable argument based on the quality of justifications, and to draw adequate conclusions "representing the most complete, plausible, or compelling understanding of an issue on the basis of the available evidence" (King & Kitchener, 2002, p. 42). Making appropriate decisions for a thesis context requires students to deal with uncertainty (i.e., knowledge is subjective when facts are unknown (Kurfiss, 1990) and multiplicity (i.e., knowledge is conjectural, uncertain and open to interpretations) (Moore, 2002).

Unfortunately, the majority of undergraduate students are at the multiplicity stage: they accept that there are different degrees of sureness and they can be sure enough if they take a personal stance on an issue (King & Kitchener, 2002). We

observe that students at this stage still look for well-defined criteria and standards to evaluate facts and knowledge. They find it difficult to judge something without a clear set of references. These difficulties not only lead to more uncertainties when working on different sections of students' own theses, but also result in challenges for peer feedback uptake when students have to integrate comments from multiple reviewers. Whereas dealing uncertainties and multiplicity is particularly important in thesis circles when teacher feedback is replaced by peer feedback, research shows students tend to rely on sources from authority rather than their epistemic value (Baxter Magolda, 2001).

Moreover, independent inquiry and student epistemological understanding (i.e., epistemic beliefs and cognition) ideally should be developed over time and embedded in the program curriculum. Nonetheless, students do not always receive guidance or support on dealing with uncertainties and multiplicity during decision making (Moore & Felten, 2018; Todd et al., 2004). This implication for instructional design is that we should provide students with just-in-time scaffolds on their thesis writing to ensure their transition from multiplicity to contextual relativism. In particular, we find it important to make students aware of their biased perception towards feedback givers (i.e., preferring teacher over peer feedback), as part of developing student feedback literacy (Carless & Boud, 2018).

3.2.2 Design Hypothesis

In our endeavor to support student integration of multiple peer feedback in thesis circles, we work with the following overarching design hypothesis: Asking students to do analogical and analytical comparisons with epistemic reflection helps them integrate multiple peer feedback and transit to contextual relativism. We work within the context of thesis circles.

We use the three building blocks of conjecture mapping to make design choices on the embodiment, mediating processes, and outcomes (Sandoval, 2014) (see Fig. 3.1). The design elements, principles, and their inter-relationships in embodiment and mediating processes are translated from (i) the integrated framework of multiple texts (Barzilai et al., 2018; List & Alexander, 2019), including learner epistemological beliefs, learners' strategic processing, and argument construction, and (ii) feedback literacy research (e.g., Carless & Boud, 2018; Dawson et al., 2019; Nicol, 2021; Nicol & McCallum, 2021).

3.3 Instructional Design

3.3.1 Embodiment

A basic instructional design requires the structure of the learning environment (set design, artifacts and tools), resources (set design, materials), sequence of tasks (epistemic/cognitive design), and social arrangements (social design, working in
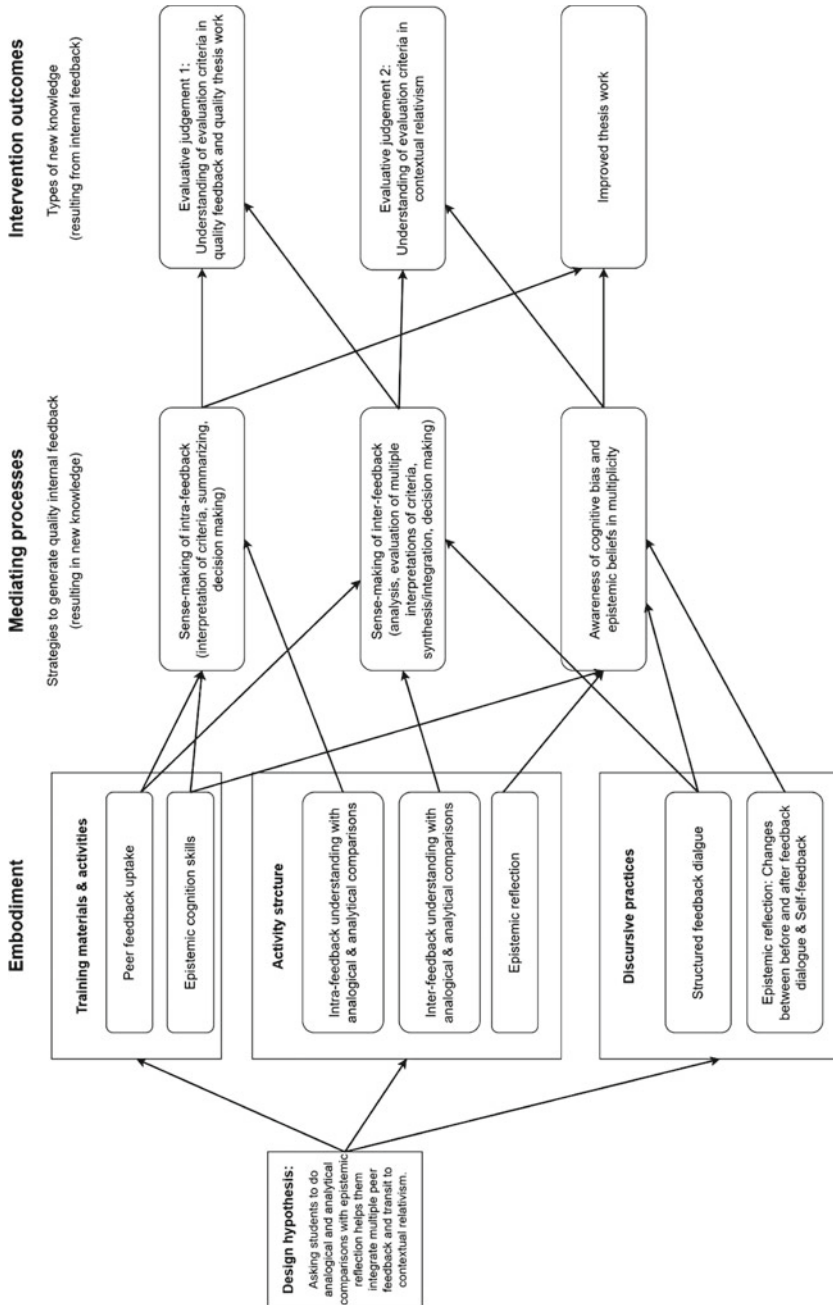
Fig. 3.1 Conjecture map for supporting student integration of multiple peer feedback

small groups, roles of receivers and peer reviewers, and their role tasks), such as the Activity Centred Analysis and Design (ACAD) framework (Yeoman & Carvalho, 2019). To develop a focused design on uptake of multiple feedback, we identify the three stages of feedback processing, preparation, execution and production, based on the literature on cultivating feedback literacy and integrated framework of multiple texts. Figure 3.2 gives an overview of these fundamental design elements of our instructional design for feedback uptake at the three stages, developed based on our conjecture map and the ACAD framework.

At the preparation stage, students should be provided with trainings on feedback literacy and structure to give feedback (Ajjawi et al., 2021). The published training materials of feedback literacy can be directly used together with our instructional design, such as instructional videos of the three processes of feedback (feed-up, feed-back, feed-forward) (Hattie & Timperley, 2007) and how to formulate constructive peer feedback. For example, supervisors can use or adapt materials from the Developing Engagement with Feedback Toolkit (DEFT) (Winstone & Nash, 2017). In addition, students who are feedback receivers can use a cover sheet (Bloxham & Campbell, 2010) to specify their personalized learning goals (i.e., specific aspects on which they are looking for feedback), accompanying their submitted thesis work.

As for the structure to give feedback, a peer feedback report for reviewers (see Table 3.1) can be used to summarize in-text comments and classify them based on assessment criteria of thesis content quality (e.g., what makes good introduction, literature review). The form of using a peer feedback report guides reviewers to relate written comments to the criteria and standards and it is more likely to induce process-related feedback (affirmations, argumentations), and to feed-forward suggestions (Dirkx et al., 2021).

3.3.1.1 Training Materials and Activities (at the Preparation Stage)

The training in our instructional design focuses on feedback uptake and epistemic cognition skills. The materials for feedback uptake include evaluative criteria of quality feedback (see the next paragraph), exemplars with good and poor feedback, and strategies for students to self-aware of their epistemic beliefs (Table 3.2).

Based on literature review, we select four evaluative criteria of quality feedback (Brookhart, 2008; Dawson et al., 2019; O'Donovan et al., 2021): purposefulness (i.e., task and writer's personalized learning goals are considered in the feedback), validity (i.e., qualitative comments are based on assessment criteria of thesis content quality), specificity (i.e., explanations why thesis work does not meet the content criteria), and constructiveness (i.e., starting with appraisals and then critiques, followed by providing suggestions how to improve the work). Purposefulness and validity are particularly important to develop students to evaluativist stage, because students need to determine whose feedback is more appropriate for their goals (purposefulness) and more helpful for them to improve their work to meet thesis assessment criteria (validity). Also, specificity and constructiveness are indispensable to effectively deliver the explanations of purposefulness and validity (Gielen & De Wever, 2015).
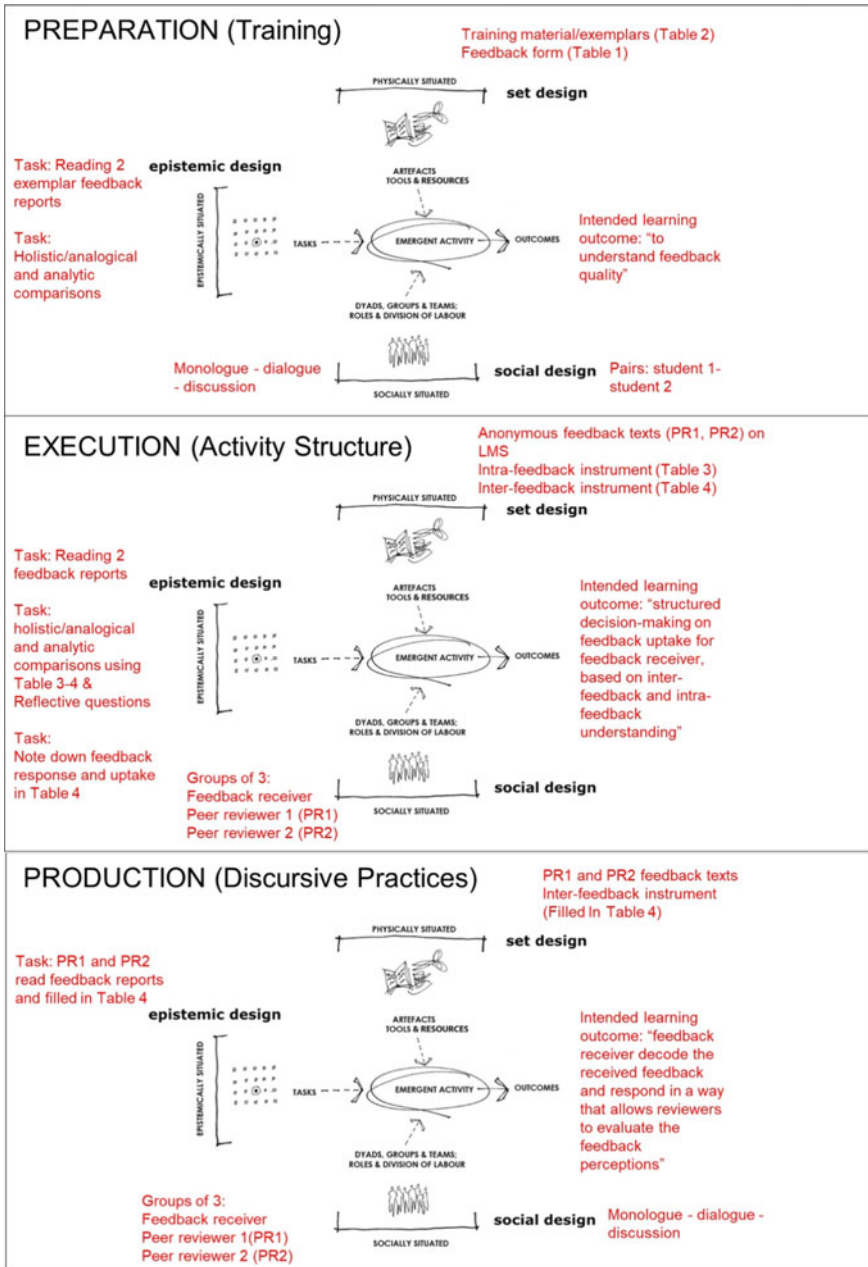
Fig. 3.2 Instructional design for feedback uptake, based on Fig. 3.1 and Yeoman (2019, p. 69)

Table 3.1 Reviewer's peer feedback report

Assessment criteria of thesis content quality	Standards	Higher level feedback based on in-text comments
Criterion 1	• Exceeds expectations • Expectations • Needs improvement	
Criterion 2	• Exceeds expectations • Expectations • Needs improvement	
Criterion 3	• Exceeds expectations • Expectations • Needs improvement	

Table 3.2 Dimensions of epistemic beliefs and instructional strategies, modified from Bråten (2011)

Dimension of beliefs	Definition: whether knowledge…	Instructional strategies that guide students
Certainty	Is absolute or evolving	To ask students to pay attention to the changes of knowledge
Simplicity	Consists of an accumulation of isolated facts or highly interrelated concepts	To have a holistic view of multiple elements and to examine inter-relatedness among different elements
Source	Comes from external authority or can be actively constructed by the person through social interactions with others	To discuss interpretations with others and to examine whether own subjectivity influences interpretations
Justifications	Is based on claims through observation and authority or based on scientific inquiry, evaluation and integration of different sources	To examine whether justification through reasoning (i.e., critical thinking), prior domain knowledge, scientific inquiry, and cross-checking of sources and multiple perspectives (e.g., considering counterarguments)

The normative models for peer feedback training are often based on analytical comparisons (Evans, 2013; Jonassen, 2011), such as using a rubric with criteria and standards to evaluate a simple piece of student work and determine its quality levels. Unfortunately, analytical comparisons based on established criteria and standards are often abstract and difficult. In the case of feedback quality, evaluation criteria may be new to students, resulting from insufficient feedback literacy in the program curriculum. Therefore, using an exemplar to show how to apply criteria and standards is regarded as a more effective training method because students are supported by both analogical and analytical reasoning. An effective exemplar should be "authentic and user-friendly" (Carless & Chan, 2017, p. 930),

similar or the same to student current assignment (e.g., feedback on thesis work) (Hendry et al., 2011), and explicit about how assessment criteria are applied to the feedback content (i.e., to show teacher tacit knowledge in evaluative judgments and quality expectations of the thesis work) (Lipnevich et al., 2014) and feedback form/technical aspect (i.e., constructiveness). Therefore, using past student work with peer feedback reports seems to be the best choice for exemplars.

As for epistemic cognition skills, literature shows evaluativists use more cognitive and metacognitive strategies, compared to people at lower levels of epistemological development (Greene & Yu, 2016). Therefore, reflection questions are used to make students aware of different dimensions of their epistemic beliefs (see Table 3.2) and to guide them to make different types of justifications.

Training activities provide students with practices to deal with multiple peer feedback and should simulate actual feedback processes in the Activity structure and Discursive practices of the conjecture map. In addition, supervisors and students should discuss exemplars so that they co-construct meanings of quality feedback and form reasoned justifications why it is good based on its interpretation of feedback criteria. Co-construction is essential to avoid the pitfalls that students regard exemplars as model answers and this in turn restricts student endeavor to make quality feedback (Carless & Chan, 2017). But before students can co-construct meaning, they need to first engage in deeper thinking processes rather than immediately participating in interactive dialogues with others. Following these rationales, we propose the following training design based on Carless and Chan's dialogic model (2017) and a step-wise monologue-dialogue-discussion (Manning & Jobbitt, 2019). Our training design consists of both analogical/holistic and analytical comparisons and emphasizes the importance of sequencing attentive and active listening before interactive dialogues (which is fundamental for feedback uptake).

At the beginning of the training, students are informed of the purpose of using exemplars for feedback uptake training. Each student reads two exemplars of feedback reports (A and B) based on a thesis work and carries out holistic/analogical (as a whole, which feedback report is better) and analytic comparisons (which one is better per criterion). During the discussion, students work in pairs and in three rounds. During the first round (monologue), Student 1 talks about her/his analyses in three minutes and Student 2 listens and takes notes. During the second round (monologue), Student 2 talks about her/his analyses in three minutes and Student 1 listens and takes notes. This monologue step forces students to focus on important findings at a higher level and listening to each other first can stimulate confrontations and avoid minimal contributions. During the third round (dialogue), both students compare their analyses and collectively determine which exemplar is better, on which they need to provide justifications to explain why. Supervisors use the strategies in Table 3.2 to probe students' epistemic beliefs. After these, the supervisor carries out the whole class discussions on each pair's findings. Through the training sessions, students understand the feedback quality processes they need to apply to their own work in further learning activities.

3.3.1.2 Activity Structure (at the Execution Stage)

During the training activities, students do not relate multiple peer feedback to their own work and feedback yet. The Activity Structure aims to engage feedback receivers in understanding and evaluating individual (intra-feedback processing) and multiple (inter-feedback processing) peer feedback through analogical/holistic and analytical comparisons.

The design principles of Activity Structure are:

- Align feedback uptake activities with the training materials and activities.
- Analogical and holistic comparisons take place before analytic comparisons.
- Guide decision making based on explanations and justifications.
- Reflect why their decisions change.
- Use organizational tools to make sense of and integrate multiple peer feedback.

As discussed in the introduction, feedback uptake is possibly influenced by receivers' perception of reviewers' level in thesis writing. Therefore, receivers carry out *anonymous* comparisons, by using any Learning Management System (LMS) that supports peer review procedures (e.g., Canvas). In the following texts, two peer reviewers are abbreviated as PR1 and PR2.

Intra-Feedback Understanding with Analogical and Analytical Comparisons. Understanding each reviewer's feedback is the first step to deal with feedback. Feedback receivers are usually asked to read each peer feedback report and relate it to the in-text comments added to their own thesis work. Unfortunately, reading alone is *not* sufficient (Kuhn, 2020) and as Winstone and Nash stated, "Many students don't even take any notice of their feedback!" (2017, p. 17). When being receivers, students need to be equipped and motivated to engage in and use feedback (Winstone et al., 2017a, 2017b). As informed by research in comparative judgements, comparing feedback quality is a purposeful activity that motivates students to read feedback carefully (otherwise they cannot compare) (Lesterhuis et al., 2017).

By holistic/analogical comparisons, receivers first identify the general impression that integrates several comments made by each reviewer by answering three questions: Is the reviewer positive, negative, constructive/neutral about your work? Which feedback report is better? Why do you make these choices?

By analytical comparisons, receivers compare the quality of each peer feedback report based on the criteria of purposefulness, validity, specificity, and constructiveness (see Table 3.3). They also need to justify their choices.

Inter-Feedback Understanding with Anonymous Analogical and Analytical Comparisons. Receivers at this stage need to identify the relationships between two reviewers' feedback and select points for feedback dialogue in discursive practices. Again, receivers carry out two types of comparisons, but this time they focus on the *content* of peer reviewers' feedback. By holistic/analogical comparisons, receivers now identify a pattern between two reviewers: Are two feedback reports complementary or conflicting each other (see Table 3.4)?

Table 3.3 Intra-feedback understanding

Holistic/analogical			
	PR1	PR2	Why do you make these choices?
Is the reviewer positive, negative, constructive/neutral about your work?	• Positive • Negative • Constructive/neutral	• Positive • Negative • Constructive/neutral	
Which feedback report is better?	• PR1	• PR2	
Analytical			
Evaluative criteria of feedback quality	Which one is better on each criterion?	Why do you make these choices?	
Purposefulness	• PR1	• PR2	
Validity	• PR1	• PR2	
Specificity	• PR1	• PR2	
Constructiveness	• PR1	• PR2	

By analytic comparisons, receivers go through two rounds of comparisons. First, they identify a relation pattern between two reviewers on each *content* criterion and justify why it is complementary or conflicting. Secondly, they compare two reviewers' feedback reports and indicate whether s(he) makes a *tentative* decision by indicating whether (s)he agrees or disagrees with analytic feedback on each criterion and justify why. In addition, they select points for feedback dialogue. Finally, they re-rank feedback quality made during intra-feedback understanding by answering this question: Which feedback report is better now? Why?

3.3.1.3 Discursive Practices: Student Feedback Dialogue and Self-feedback (at the Production Stage)

The importance of feedback dialogues has been advocated by multiple researchers in feedback literacy (e.g., Ajjawi & Boud, 2018; Carless & Chan, 2017; Winstone et al., 2017a, 2017b). As pointed out by Winstone et al. (2017a, 2017b), feedback receivers must decode the received feedback and respond in a way that allows reviewers to evaluate the feedback perceptions. In addition, receivers should play a proactive role in peer feedback dialogue (Zhu & To, 2021). In our Activity Structure, receivers have been decoding feedback content and evaluating feedback quality (see Tables 3.3 and 3.4), without knowing who reviewers are.

Before the feedback dialogue, PR1 and PR2 read each other's feedback report and receiver's completed Table 3.4, because the reviewers need to evaluate how the feedback is perceived and interpreted. As a Discursive Practice, the receiver attends to this evaluation and needs to actively find out "what to do differently, and how" (Winstone & Nash, 2017, p. 17). The feedback dialogue should be structured to facilitate different role tasks and be aligned with the training activities. We propose to adapt Manning and Jobbit's model (2019) to dialogue-monologues-discussion

Table 3.4 Inter-feedback comparisons

Holistic/analogical comparisons					
Are two feedback reports complementary or conflicting each other? • Complementary • Conflicting					

Analytical comparisons—identifying a relation pattern

Assessment criteria of thesis *content* quality	Relation patterns between PR1 and PR2 per criterion	Why is it complementary (similarities) or conflicting (differences or contradictory)?
Criterion 1	• Complementary • Conflicting	
Criterion 2	• Complementary • Conflicting	
Criterion 3	• Complementary • Conflicting	

Making a tentative decision of each reviewer's feedback on each criterion

Assessment criteria of thesis *content* quality	PR1	Receiver's justifications on why agree/disagree	PR2	Receiver's justifications on why agree/disagree	Discussion points for feedback dialogue
Criterion 1	• Agree • Disagree		• Agree • Disagree		
Criterion 2	• Agree • Disagree		• Agree • Disagree		
Criterion 3	• Agree • Disagree		• Agree • Disagree		
Which feedback report is better now?			Why do you make this choice?		
• PR1 • PR2					

(see Fig. 3.2). First, PR1 and PR2 have a dialogue to discuss whether they agree or disagree with the relation patterns in Table 3.4. For the complementary patterns, PR1 and PR2 elaborate on what the receiver can do. For the conflicting patterns, PR1 and PR2 need to find out why these differences occur in their feedback. The receiver listens, takes notes, and reacts to PR1 and PR2's dialogue results. Then PR1 and PR2 take turns to react to the receiver's disagreements (in Table 3.4) in a monologue while the receiver listens and takes notes. Finally, the receiver goes through the discussion points in Table 3.4 to have a group discussion with both PR1 and PR2. Then the receiver answers three reflective questions: (1) At the beginning of the feedback dialogue session, are you surprised when you know who the reviewers are? If so, why are you surprised? (2) After this feedback dialogue, which peer feedback report do you find better? PR1 or PR2? (3) What would you change your own feedback to PR1 and PR2 now and why? The detailed steps in this feedback dialogue are shown in Appendix.

Table 3.5 Feedback receiver's self-feedback report

Re-evaluate relation patterns between PR1 and PR2		
Assessment criteria of thesis content quality	Relation patterns between PR1 and PR2	If a pattern is changed, justify why*
Criterion 1	• Complementary • Conflicting	
Criterion 2	• Complementary • Conflicting	
Criterion 3	• Complementary • Conflicting	

Making a final decision of each reviewer's feedback on each criterion				
Assessment criteria of thesis content quality	PR1	If your decision is changed, justify why**	PR2	If your decision is changed, justify why**
Criterion 1	• Agree • Disagree		• Agree • Disagree	
Criterion 2	• Agree • Disagree		• Agree • Disagree	
Criterion 3	• Agree • Disagree		• Agree • Disagree	

Action plan		
Assessment criteria of thesis *content* quality	Action points to improve own thesis work quality	Action points to improve own feedback quality
Criterion 1		
Criterion 2		
Criterion 3		

*Examples are: My interpretation of their feedback was not entirely correct. Their elaborations during the feedback dialogue became clear

**Examples are: My interpretation of this criterion was not correct. The reviewer's elaborations during the feedback dialogue convinced me that (s)he is right about XX of my work. My tentative decision was influenced by the strict tone in this reviewer's feedback report. But during the feedback dialogue, I think (s)he is right about XX of my work, I did not XX

At the end of the feedback dialogue, the receiver makes a self-feedback report by re-evaluating the relation patterns, making a final decision of each reviewer's feedback on each criterion, and making an action plan (see Table 3.5).

3.3.2 Mediating Processes

The mediating processes are the hypothesized interactions triggered by Activity Structure and are directly contributed to the outcomes (Sandoval, 2014). Stimulating students to construct personal understanding from external feedback information is a prerequisite for putting it into action. As described in Activity

structure, students are prompted to use effective cognitive strategies to understand each individual's and multiple peers' work. Macrostructure strategies are effective to enhance both intra- and inter-feedback understanding, such as identifying main ideas and organizational tools (Castells et al., 2021).

3.3.2.1 Sense-Making of Intra-Feedback

When doing analogical/holistic and analytic comparisons, receivers (with or without awareness) carry out comprehension monitoring (i.e., students' self-evaluations of their understanding), epistemic monitoring (i.e., students' monitoring of feedback not violating their prior knowledge, epistemic standards for trustworthiness), and the monitoring of cognitive product formation (i.e., students' monitoring of their task goals and their achievement of expected cognitive outcomes) (List & Alexander, 2019). These strategies are important for students to make sense of the criteria of both feedback quality and thesis content and the relationship between these two sets of criteria. For example, a comment about research questions can be "The specific focus of the study only becomes clear at the end". Receivers examine to what extent this comment is relevant to the criterion of research questions (validity, comprehension monitoring) and check their prior knowledge about research questions (epistemic monitoring): Is this comment elaborated with explanations? Is the focus characteristics of research questions only or does it relate more to the introduction section?

3.3.2.2 Sense-Making of Inter-Feedback

Several comparisons and reflective questions guide receivers to make sense of inter-feedback by constructing a mental representation of each peer reviewer's feedback (i.e., holistic judgement), comparing and contrasting different interpretations of multiple criteria from multiple reviewers (i.e., complementary or conflicting), synthesizing complementary comments or reconciling conflicting comments (i.e., Table 3.4). The integration of multiple peer feedback is likely to take place, when receivers identify relation patterns among two reviewers, combine and organize information into a coherent whole, connect multiple inter-feedback links (e.g., whether two reviewers agree with each other holistically or analytically), and make decisions on which reviewer's feedback to agree with.

3.3.2.3 Awareness of Epistemic Beliefs and Cognitive Bias

As discussed in the introduction, undergraduates need support on improving their epistemic beliefs to further develop from multiplicity to contextual relativism so that they can deal with the high complexity of their own thesis work and multiple peer feedback. Epistemic beliefs refer to students' feelings and ideas about the nature and source of knowledge (Hofer & Pintrich, 1997) which are important in the peer feedback activities (Banihashem et al., 2023; Noroozi, 2018, 2022). Table 3.2 lists four dimensions of epistemic beliefs that are likely to influence student understanding and making judgment of others' work and instructional strategies to make students examine their beliefs (Bråten et al., 2011).

As for cognitive bias, human mental processing relies on analogical reasoning. When encountering a new situation, we look for prior knowledge in our schema and try to locate similar knowledge or experience to help us make decisions. Unfortunately, prior knowledge is not always a reliable source because memories can fade and past experience was situated in a different context. Therefore, the Activity Structure explicitly asks students to compare peer feedback reports to their prior experiences (e.g., training activities, earlier comparison results).

Our conjecture map ends at the activity that students complete a self-feedback report (Table 3.5). We do not expand on how students use the feedback on their actual improvement of their work.

3.4 Outcomes

There are three learning outcomes of supporting students in the integration of multiple peer feedback. First, both analogical/holistic and analytical comparisons are likely to improve student levels of evaluative judgements based on a better understanding of criterion knowledge of quality feedback and quality thesis work. Second, different types of comparisons and questions engage students in all of the four dimensions of epistemic beliefs (in Table 3.2) and these in turn contribute to student development towards evaluativist (contextual relativism). Third, asking students to fill out sense-making tables (Tables 3.3 and 3.4) and to generate a self-feedback report (Table 3.5) is likely to result in improved work (Nicol et al., 2014; Wu et al., 2019).

3.5 Conclusion

Research on peer feedback has been exploding in numbers and diversity. However, the specific focus of each research school makes it difficult for teachers to interconnect all of these aspects in their instructional design (Nieminen et al., 2022). With this in mind, based on integration of research findings, we hope that a concrete instructional design with activity descriptions can support teachers in designing peer review activities in thesis circles. In the future study, we will implement each step in Activity Structure to corroborate the occurrence of Mediating Processes and Outcomes.

For peer feedback to be effective, students need a proper training and multiple practices to process and integrate multiple peer feedback so that integrated multiple peer feedback is likely to replace supervisor feedback effectively. It is inevitable that supervisors need to invest certain transition costs on training and multiple practices in the beginning. Fortunately, thesis circles often involve a group of supervisors to design and organize activities together. Through collaboration with others, in a long term, each supervisor's transition costs will be paid off by implementing the proposed activities of our design.

Although this chapter focuses on feedback receivers, we are aware that feedback effectiveness cannot only count on the receivers' uptake. Feedback is always interactive and reviewers' feedback influences how feedback uptake takes place (Latifi et al., 2023). Still, when students are supported with these activities, materials (i.e., Tables 3.2, 3.3, 3.4 and 3.5) and reflective questions, they are more likely to change their own feedback giving behavior.

Finally, although this chapter focuses on multiple reviewers' feedback in thesis writing, the support in this design can be applicable for students to deal with real-world discussions that often involve multiple voices and opinions. Integrating epistemological development to instruction design is important for students to gradually develop from multiplicity to contextual relativism and this should receive more attention in undergraduate curriculum design.

Appendix: Steps in Student Feedback Dialogue
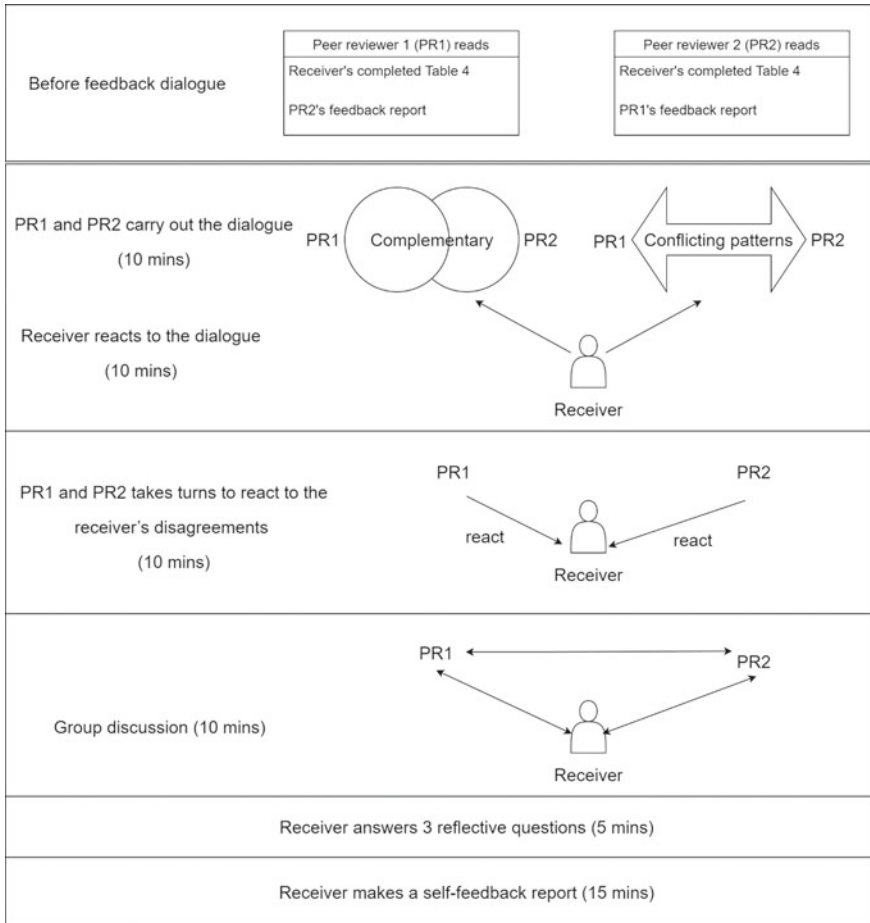
References

Ajjawi, R., & Boud, D. (2018). Examining the nature and effects of feedback dialogue. *Assessment and Evaluation in Higher Education, 43*(7), 1106–1119. https://doi.org/10.1080/02602938.2018.1434128

Ajjawi, R., Kent, F., Broadbent, J., Tai, J. H. -M., Bearman, M., & Boud, D. (2021). Feedback that works: A realist review of feedback interventions for written tasks. *Studies in Higher Education*, 1–14. https://doi.org/10.1080/03075079.2021.1894115

Banihashem, S. K., Noroozi, O., Biemans, H. J. A., & Tassone, V. C. (2023). The intersection of epistemic beliefs and gender in argumentation performance. *Innovations in Education and Teaching International*. https://doi.org/10.1080/14703297.2023.2198995

Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. A. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review, 100489*. https://doi.org/10.1016/j.edurev.2022.100489

Barzilai, S., Zohar, A. R., & Mor-Hagani, S. (2018). Promoting integration of multiple texts: A review of instructional approaches and practices. *Educational Psychology Review, 30*(3), 973–999. https://doi.org/10.1007/s10648-018-9436-8

Baxter Magolda, M. B. (2001). *Making their own way: Narratives for transforming Higher Education to promote self-development*. Stylus Publishing. https://books.google.nl/books?id=KPvxDwAAQBAJ

Berndt, M., Strijbos, J.-W., & Fischer, F. (2018). Effects of written peer-feedback content and sender's competence on perceptions, performance, and mindful cognitive processing. *European Journal of Psychology of Education, 33*(1), 31–49. https://doi.org/10.1007/s10212-017-0343-z

Bloxham, S., & Campbell, L. (2010). Generating dialogue in assessment feedback: Exploring the use of interactive cover sheets. *Assessment and Evaluation in Higher Education, 35*(3), 291–300. https://doi.org/10.1080/02602931003650045

Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J.-F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist, 46*(1), 48–70. https://doi.org/10.1080/00461520.2011.538647

Breuch, L. -A. K. (2004). *Virtual peer review: Teaching and learning about writing in online environments.* State University of New York Press. http://site.ebrary.com/id/10594775

Brookhart, S. M. (2008). How to give effective feedback to your students. *Association for Supervision and Curriculum Development*

Broadbent, J., Panadero, E., & Boud, D. (2018). Implementing summative assessment with a formative flavour: A case study in a large class. *Assessment and Evaluation in Higher Education, 43*(2), 307–322. https://doi.org/10.1080/02602938.2017.1343455

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment and Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Carless, D., & Chan, K. K. H. (2017). Managing dialogic use of exemplars. *Assessment and Evaluation in Higher Education, 42*(6), 930–941. https://doi.org/10.1080/02602938.2016.1211246

Castells, N., Minguela, M., Solé, I., Miras, M., Nadal, E., & Rijlaarsdam, G. (2021). Improving questioning–answering strategies in learning from multiple complementary texts: An intervention study. *Reading Research Quarterly*, *n/a*(n/a). https://doi.org/10.1002/rrq.451

Cho, K., & MacArthur, C. (2011). Learning by reviewing. *Journal of Educational Psychology, 103*(1), 73–84. https://doi.org/10.1037/a0021950

Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D., & Molloy, E. (2019). What makes for effective feedback: Staff and student perspectives. *Assessment and Evaluation in Higher Education, 44*(1), 25–36. https://doi.org/10.1080/02602938.2018.1467877

de Kleijn, R., Meijer, P., Brekelmans, M., & Pilot, A. (2012). Curricular goals and personal goals in master's thesis projects: Dutch student-supervisor dyads. *International Journal of Higher Education, 1*, 1–11. https://doi.org/10.5430/ijhe.v2n1p1

Dirkx, K., Joosten-ten Brinke, D., Arts, J., & van Diggelen, M. (2021). In-text and rubric-referenced feedback: Differences in focus, level, and function. *Active Learning in Higher Education, 22*(3), 189–201. https://doi.org/10.1177/1469787419855208

Elbow, P. (1998). *Writing without teachers* (2nd edn.). Oxford University Press.

Elder, L., & Paul, R. (2009). Close reading, substantive writing and critical thinking: Foundational skills essential to the educated mind. *Gifted Education International, 25*(3), 286–295. https://doi.org/10.1177/026142940902500310

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research*, *83*(1), 70–120. https://doi.org/10.3102/0034654312474350

Facione, P. (2011). Critical thinking: What it is and why it counts. Retrieved Mar. 9th, 2020, from https://www.insightassessment.com/article/critical-thinking-what-it-is-and-why-it-counts-pdf

Falchikov, N. (2013). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. Taylor & Francis.

Gentner, D., Holyoak, K. J., & Kokinov, B. N. (2001). *The analogical mind: Perspectives from cognitive science*. MIT Press.

Gielen, M., & De Wever, B. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior*, *52*, 315–325. https://doi.org/10.1016/j.chb.2015.06.019

Gielen, S., Tops, L., Dochy, F., Onghena, P., & Smeets, S. (2010). A comparative study of peer and teacher feedback and of various peer feedback forms in a secondary school writing curriculum. *British Educational Research Journal, 36*(1), 143–162. https://doi.org/10.1080/014119209028 94070

Greene, J. A., & Yu, S. B. (2016). Educating critical thinkers: The role of epistemic cognition. *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 45–53. https://doi.org/10.1177/237 2732215622223

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hendry, G. D., Bromberger, N., & Armstrong, S. (2011). Constructive guidance and feedback for learning: the usefulness of exemplars, marking sheets and different types of feedback in a first year law subject. *Assessment & Evaluation in Higher Education*, *36*(1), 1–11. https://doi.org/ 10.1080/02602930903128904

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67*(1), 88–140. https://doi.org/10.3102/00346543067001088

Hsiao, Y. P., Brouns, F., van Bruggen, J., & Sloep, P. B. (2015). Effects of training peer tutors in content knowledge versus tutoring skills on giving feedback to help tutees' complex tasks. *Educational Studies*, *41*(5), 499–512. https://doi.org/10.1080/03055698.2015.1062079

Huang, L.-S. (2010). Seeing eye to eye? The academic writing needs of graduate and undergraduate students from students' and instructors' perspectives. *Language Teaching Research, 14*(4), 517–539. https://doi.org/10.1177/1362168810375372

Huisman, B., Saab, N., van Driel, J., & van den Broek, P. (2018). Peer feedback on academic writing: Undergraduate students' peer feedback role, peer feedback perceptions and essay performance. *Assessment and Evaluation in Higher Education*, 1–14. https://doi.org/10.1080/026 02938.2018.1424318

Jonassen, D. H. (2011). Learning to solve problems: A handbook for designing problem-solving learning environments. Routledge.

King, P., & Kitchener, K. (2002). The reflective judgment model: Twenty years of research on epistemic cognition. In B. K. Hofer, & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 37–61). Taylor & Francis Group.

Kuhn, D. (2020). Why is reconciling divergent views a challenge? *Current Directions in Psychological Science, 29*(1), 27–32. https://doi.org/10.1177/0963721419885996

Kurfiss, J. G. (1990). Critical thinking: Theory, research, practice, and possibilities. *Teaching Sociology, 18*, 581.

Latifi, S., Noroozi, O., & Talaee, E. (2021). Peer feedback or peer feedforward? Enhancing students' argumentative peer learning processes and outcomes. *British Journal of Educational Technology, 52*(2), 768–784. https://doi.org/10.1111/bjet.13054

Latifi, S., & Noroozi, O. (2021). Supporting argumentative essay writing through an online supported peer-review script. *Innovations in Education and Teaching International, 58*(5), 501–511. https://doi.org/10.1080/14703297.2021.1961097

Latifi, S., Noroozi, O., & Talaee, E. (2023). Worked example or scripting? Fostering students' online argumentative peer feedback, essay writing and learning. *Interactive Learning Environments, 31*(2), 655–669. https://doi.org/10.1080/10494820.2020.1799032

Lesterhuis, M., Verhavert, S., Coertjens, L., Donche, V., & De Maeyer, S. (2017). Comparative judgement as a promising alternative to score competences. In E. Cano, & G. Ion (Eds.), *Innovative practices for higher education assessment and measurement* (pp. 119–138). IGI Global. https://doi.org/10.4018/978-1-5225-0531-0.ch007

Lipnevich, A. A., McCallen, L. N., Miles, K. P., & Smith, J. K. (2014). Mind the gap! Students' use of exemplars and detailed rubrics as formative assessment. *Instructional Science*, *42*(4), 539–559.

List, A., & Alexander, P. A. (2019). Toward an integrated framework of multiple text use. *Educational Psychologist, 54*(1), 20–39. https://doi.org/10.1080/00461520.2018.1505514

Manning, S. J., & Jobbitt, T. (2019). Engaged and interactive peer review: Introducing peer review circles. *RELC Journal, 50*(3), 475–482. https://doi.org/10.1177/0033688218791832

McKenney, S., & Reeves, T. C. (2014). Educational design research. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 131–140). Springer. https://doi.org/10.1007/978-1-4614-3185-5_11

Mercader, C., Ion, G., & Díaz-Vicario, A. (2020). Factors influencing students' peer feedback uptake: Instructional design matters. *Assessment and Evaluation in Higher Education, 45*(8), 1169–1180. https://doi.org/10.1080/02602938.2020.1726283

Moore, J. L., & Felten, P. (2018). Academic development in support of mentored undergraduate research and inquiry. *International Journal for Academic Development, 23*(1), 1–5. https://doi.org/10.1080/1360144X.2018.1415020

Moore, W. S. (2002). Understanding learning in a postmodern world: Reconsidering the Perry scheme of intellectual and ethical development. In B. K. Hofer, & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 17–36). Pintrich, Taylor & Francis Group.

Nicol, D. (2021). The power of internal feedback: Exploiting natural comparison processes. *Assessment and Evaluation in Higher Education, 46*(5), 756–778. https://doi.org/10.1080/02602938.2020.1823314

Nicol, D., & McCallum, S. (2021). Making internal feedback explicit: Exploiting the multiple comparisons that occur during peer review. *Assessment and Evaluation in Higher Education*, 1–19. https://doi.org/10.1080/02602938.2021.1924620

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment and Evaluation in Higher Education, 39*(1), 102–122. https://doi.org/10.1080/02602938.2013.795518

Nieminen, J. H., Bearman, M., & Tai, J. (2022). How is theory used in assessment and feedback research? A critical review. *Assessment and Evaluation in Higher Education*, 1–18. https://doi.org/10.1080/02602938.2022.2047154

Noroozi, O. (2018). Considering students' epistemic beliefs to facilitate their argumentative discourse and attitudinal change with a digital dialogue game. *Innovations in Education and Teaching International, 55*(3), 357–365. https://doi.org/10.1080/14703297.2016.1208112

Noroozi, O. (2022). The role of students' epistemic beliefs for their argumentation performance in higher education. *Innovations in Education and Teaching International*, 1–12. https://doi.org/10.1080/14703297.2022.2092188

Noroozi, O., Banihashem, S. K., Biemans, H. J. A., Smits, M., Vervoort, M. T. W., & Verbaan, C. (2023). Design, implementation, and evaluation of an online supported peer feedback module to enhance students' argumentative essay quality. *Education and Information Technologies*, 1–28. https://doi.org/10.1007/s10639-023-11683-y

O'Donovan, B. M., den Outer, B., Price, M., & Lloyd, A. (2021). What makes good feedback good? *Studies in Higher Education*, *46*(2), 318–329. https://doi.org/10.1080/03075079.2019.1630812

Rajagopal, K., Vrieling-Teunter, E., Hsiao, Y. P., Van Seggelen-Damen, I., & Verjans, S. (2021). Guiding thesis circles in higher education: Towards a typology. *Professional Development in Education*, 1–18. https://doi.org/10.1080/19415257.2021.1973072

Romme, G., & Nijhuis, J. (2002). *Collaborative learning in thesis rings*.

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment and Evaluation in Higher Education, 34*(2), 159–179. https://doi.org/10.1080/02602930801956059

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the Learning Sciences, 23*(1), 18–36. https://doi.org/10.1080/10508406.2013.778204

Taghizadeh Kerman, N., Noroozi, O., Banihashem, S. K., Karami, M. & Biemans, Harm. H. J. A. (2022). Online peer feedback patterns of success and failure in argumentative essay writing. *Interactive Learning Environments*, 1–10. https://doi.org/10.1080/10494820.2022.2093914

Todd, M., Bannister, P., & Clegg, S. (2004). Independent inquiry and the undergraduate dissertation: Perceptions and experiences of final-year social science students. *Assessment and Evaluation in Higher Education, 29*(3), 335–355. https://doi.org/10.1080/0260293042000188285

Winstone, N. E., & Nash, R. A. (2017). The "Developing Engagement with Feedback Toolkit (DEFT)": Integrating assessment literacy into course design. In S. Elkington & C. Evans (Eds.), *Transforming assessment in higher education: A case study series* (pp. 48–52). Higher Education Academy.

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017a). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist, 52*(1), 17–37. https://doi.org/10.1080/00461520.2016.1207538

Winstone, N. E., Nash, R. A., Rowntree, J., & Parker, M. (2017b). 'It'd be useful, but I wouldn't use it': Barriers to university students' feedback seeking and recipience. *Studies in Higher Education, 42*(11), 2026–2041. https://doi.org/10.1080/03075079.2015.1130032

Wu, W. -H., Kao, H. -Y., Wu, S. -H., & Wei, C. -W. (2019). Development and evaluation of affective domain using student's feedback in entrepreneurial Massive Open Online Courses. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01109

Yeoman, P., & Carvalho, L. (2019). Moving between material and conceptual structure: Developing a card-based method to support design for learning. *Design Studies*, *64*, 64–89. https://doi.org/10.1016/j.destud.2019.05.003

Zhao, H. (2010). Investigating learners' use and understanding of peer and teacher feedback on writing: A comparative study in a Chinese English writing classroom. *Assessing Writing*, *15*(1), 3–17. https://doi.org/10.1016/j.asw.2010.01.002

Zhu, Q., & To, J. (2021). Proactive receiver roles in peer feedback dialogue: Facilitating receivers' self-regulation and co-regulating providers' learning. *Assessment and Evaluation in Higher Education*, 1–13. https://doi.org/10.1080/02602938.2021.2017403

Part II

Methodological Contributions on Peer Learning

Peer Assessment Using Criteria or Comparative Judgement? A Replication Study on the Learning Effect of Two Peer Assessment Methods

4

Tine van Daal, Mike Snajder, Kris Nijs, and Hanna Van Dyck

4.1 Introduction

Learning complex skills as for instance writing or problem-solving in the domain of physics is not an easy endeavor for many students. To optimize students' learning, showing examples can be helpful (To et al., 2021; Sadler, 1989). Viewing examples should enable students to gain a better understanding of what constitutes quality (Orsmond et al., 2002; Rust et al., 2003; Sadler, 1989, 2009) and can support students' self-regulation (To et al., 2021). However, merely presenting examples is not sufficient. Students should also engage with these examples to reach a deeper understanding of quality (Carless & Chan, 2017; Handley & Williams, 2011; Sadler, 1989; Tai et al., 2018). This raises the question how students should ideally interact with examples to optimize their learning. A promising way to do so is setting-up a peer assessment where students assess each other's work (Carless & Boud, 2018; Tai et al., 2018; To et al., 2021). Several peer assessment methods can be adopted to support students in judging their peers' work. Using a predefined list of criteria to assess pieces of work one by one is the most commonly used method (Carless & Chan, 2017; Rust et al., 2003) because it results in reliable judgements and makes quality criteria explicit for students (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013). However, research shows that learning gains are higher if people compare two examples rather than looking at only one example (e.g., Alfieri et al., 2013). This suggests that comparative

T. van Daal (✉) · M. Snajder · K. Nijs · H. Van Dyck
Department of Training and Education Sciences, University of Antwerp, Sint-Jacobstraat 2, 2000 Antwerp, Belgium
e-mail: tine.vandaal@uantwerpen.be

judgement, in which students compare two pieces of work and choose the better one, might also be a valuable peer assessment method.

Only a limited number of studies explicitly compared the effectiveness of using a criteria list or comparative judgement in the context of peer assessment. Jones and Wheadon (2015) examined the reliability and validity of the outcomes of both approaches to peer assessment but did not dig into its learning effect. The latter was done by Bouwer et al. (2018) and only recently also by Stuulen et al. (2022). In the study of Bouwer and colleagues (2018), forty second-year students enrolled in the course International Trade English 2A in the Bachelor of a Business Management program assessed essays of their fellow students using either a criteria list or comparative judgement. Findings show that assessment method influences the quantity and quality of the feedback: students in the comparative condition give more feedback in general and look more often at higher-order aspects and less at lower-order aspects when giving negative feedback than students in the criteria condition. Furthermore, peer assessment method also impacts students' writing performance. Students in the comparative condition outperform their peers in the criteria condition (Bouwer et al., 2018). This suggests that the comparative approach might be valuable in supporting students. However, Stuulen and colleagues (2022) find no effect of assessment condition on high school students' writing performance in Dutch and the opposite effect regarding the quality of feedback: students in the criteria condition give more higher-order feedback than students in the comparative condition. This raises the question to what extent the learning effects that Bouwer et al. (2018) find in the context of writing English essays in higher education can be generalized to other contexts and other subjects. Investigating the generalizability of findings to other contexts and subjects requires conceptual replication (Hendrick, 1990; Schmidt, 2009).

Therefore, this study sets out to conceptually replicate the study by Bouwer et al. (2018) in other contexts and for other subjects. More specifically, this study investigated the effect of both peer assessment methods on a) the quality of students' peer feedback in the context of writing in French (secondary education) and scientific reporting of statistical results (university education) and b) on students' performance. The latter was also examined in the context of problem-solving in physics (secondary education).

4.2   Theoretical Framework

This theoretical framework first explains what is referred to with quality of peer feedback in the context of writing. This is followed by a discussion of both peer assessment methods, and their expected learning effects.

4.2.1   Quality of Peer Feedback

During peer assessment, students are often asked to provide feedback. It is expected that this encourages students to process information in a deep way (Lundstrom & Baker, 2009; Topping, 2009). This stimulates students to critically assess the works of their peers, to formulate strengths and weaknesses with which that student could improve his work (Nicol & Macfarlane-Dick, 2006). Hence, it is important that the feedback that students give is of high quality (Patchan et al., 2016). Bouwer et al. (2018) conceptualize quality of feedback as the content and quantity of feedback. They define quantity as the number of unique aspects per essay that a student refers to in their feedback. For the content of feedback, a distinction is made between higher-order and lower-order feedback. In the context of writing, higher-order aspects are related to, for example, content, structure, and style of the essay. Lower-order aspects refer to, for example, spelling, grammar, length, and layout (Bouwer et al., 2018; Cumming et al., 2002; Lesterhuis et al., 2018). Feedback that focuses on higher-order aspects is preferred as it contributes more to improving the quality of a text than feedback on lower-order aspects (Bouwer et al., 2018; Patchan et al., 2016).

4.2.2   Peer Assessment Using Criteria

Assessing pieces of work one by one using a list of criteria requires students to break down the quality of a piece of work into several separate aspects (Weigle, 2002). These criteria make it transparent how a piece of work will be assessed and what the expectations are (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013; Sadler, 1989). The student evaluates each criterion one by one. The final grade for a piece of work is obtained by summing these criterion scores (Norton, 2004; Sadler, 2009).

It is expected that by scoring each other's work based on criteria, students learn how high-quality pieces of work differ from works of lower quality. This increases students' knowledge of text quality and makes criteria and standards concrete (Bloxham & Boyd, 2007; Handley & Williams, 2011; Orsmond et al., 2002; Rust et al., 2003). Furthermore, understanding quality criteria helps students in monitoring and evaluating their own progress and performance (Tai et al., 2018). This helps them in self-regulating their own learning and makes them less dependent on the teacher (Bloxham & Boyd, 2007). It is important that self-regulation and self-evaluation skills are developed as they have been shown to be a strong predictor of better writing performance (Boud, 2000; Zimmerman & Risemberg, 1997).

Although studies show that students can reliably assess the work of fellow students using a criteria list (Topping, 1998, 2009), there are also some criticisms regarding this method. Some studies indicate that there is no certainty that the use of criteria results in reliable and valid outcomes (Sadler, 1989, 2009; Weigle, 2002). Assessors are not always consistent in their judgements, and they often

disagree (Schoonen et al., 1997). Some assessors are stricter than others (Weigle, 2002). Furthermore, when evaluating text quality, assessors differ in their interpretation of the criteria (Eckes, 2008). It is also difficult to define all criteria in concrete terms (Chapelle et al., 2008). As a result, this approach prevents students from reaching a full understanding of the entire quality of a piece of work. Students may have the tendency to only consider the predefined criteria while other aspects may also be relevant for assessing quality (Bouwer et al., 2015). Finally, this approach does not allow students to develop skills to determine for themselves which criteria are relevant for a given task. It is important that these skills are developed in students so that they are ready for life outside school where no predetermined criteria are available. Finally, when students perceive the criteria as demands from teachers, this may be associated with only superficial learning and achievement (Bell et al., 2013; Torrance, 2007).

4.2.3   Peer Assessment Using Comparative Judgement

Comparative judgement asks students to compare two pieces of work and indicate which is better in terms of the skill under assessment (Pollitt, 2012a, 2012b). All students make several comparative judgements. These judgements are statistically modelled to create a rank-order that orders the pieces of work from low to high quality (Pollitt, 2012a, 2012b). Comparative judgement requires students to make a holistic judgement which implies that a student evaluates the pieces of work as a whole and directly arrives at an overall judgement (Pollitt, 2012a, 2012b; Sadler, 2009). In addition, comparative judgement gives students the opportunity to reflect on how they conceptualize the quality of a piece of work (Sadler, 2009; Williamson & Huot, 1992).

Evidence from research into learning from comparison underpins that learning gains are higher when comparing examples than viewing examples separately. While comparing, students look for similarities and differences between two pieces which make different aspects of each piece of work salient (Alfieri et al., 2013; Gentner, 2010; Gentner & Markman, 1997). For example, in one comparison, the content of an essay may stand out, while in another comparison, spelling mistakes may be noticeable. In this way, students come to a better understanding of important characteristics that a good piece of work must satisfy (Alfieri et al., 2013; Gentner, 2010; Pachur & Olsson, 2012), which in turn enables them to deliver tasks of higher quality (Orsmond et al., 2002; Sadler, 1989).

That students gain a better understanding in quality criteria through comparison is also demonstrated in the context of peer assessment using comparative judgement (Bartholomew et al., 2019; Jones & Alcock, 2014; Seery et al., 2012). For example, the study by Seery et al. (2012) underpins that comparative judgement has a positive influence on the development of higher-order thinking of student teachers who comparatively assessed design projects of their peers. Similarly, the study by Bartholomew et al. (2019) shows that students in secondary education

gain a better understanding of the assignment's criteria by making comparative judgements on the work of their peers.

In addition, comparative judgement can be expected to have an impact on the quality of students' feedback although evidence regarding the direction of this effect is unclear. Students in the comparative condition of the study by Bouwer et al. (2018) provide more feedback in general than students in the criteria condition. This is also found by Stuulen et al. (2022) but only for positive feedback, while students in the criteria condition give more negative feedback. Also, the content of the feedback differs depending on peer assessment method. Results of Bouwer et al. (2018) indicate that students in the comparative condition give more negative feedback on higher-order aspects and less on lower-order aspects of their peers' text than students in the criteria condition. Positive feedback does not differ between conditions. The reverse is found in the study by Stuulen et al. (2022) as students in the criteria condition give more higher-order feedback than students in the comparative condition. No differences in lower-order feedback are found.

Peer assessment using comparative judgement can also improve students' performance (Bartholomew et al., 2019; Bouwer et al., 2018). In the study by Bartholomew et al. (2019), the performance of students who participated in a peer assessment via comparative judgement is improved compared to that of students who only discussed their work with peers. The study by Bouwer et al. (2018) also shows that the performance of students who comparatively judged essays is higher than that of students who assessed essays using a criteria list. However, Stuulen and colleagues (2022) find no difference in students' performance after participating in a peer assessment exercise using either a criteria list or comparative judgement.

4.3 This Study

The current study conceptually replicated the study of Bouwer et al. (2018) on the learning effect of two peer assessment methods (use of criteria and comparative judgement). In doing so, the extent to which the results of Bouwer et al. (2018) can be generalized to other contexts and subjects was examined (Hendrick, 1990; Schmidt, 2009). For this purpose, three small scale studies were set up in Flanders (Belgium). The first two studies were run in secondary education and focused on problem-solving in physics and writing in French. For the third study, data on scientific reporting of statistical results was collected in one pre-master program of a Flemish university.

In line with Bouwer et al. (2018), two research questions were answered. The first research question investigated the effect of the use of criteria and comparative judgement on the quality of the peer feedback that students provided. Based on the results of the original study (Bouwer et al., 2018), it was expected that students in the comparative condition would provide more feedback in general and focus more on higher-order aspects than students in the criteria condition. Furthermore, the latter students were expected to focus more on lower-order aspects.

The second research question examined the effect of both assessment methods on students' performance. In line with Bouwer et al. (2018), students' prior knowledge and self-efficacy were controlled for. Based on the results of the original study, it was expected that students in the comparative condition would perform better than students in the criteria condition.

4.4   Method

Three small-scale studies were set up to conceptually replicate the findings of Bouwer et al. (2018). This section describes the methodology underpinning each of these studies. First, an outline is given of the three samples. Then, the three phases of the research design are discussed including a description of the instruments employed in each sample. Finally, operationalization of the key variables and the analysis approach are detailed upon.

All code used to clean and prepare the data sets, run the analyses and report on the results can be found online. All data files, fitted models, tables and figures can be consulted at the Open Science Framework.

4.4.1   Samples

Sample A (physics) was collected in one secondary school in Flanders (Belgium). All pupils who were enrolled in the third grade (aged 14 or 15 years) of the study track "Sciences" or "Sports sciences" were asked to voluntarily participate in this study. After being informed about the study, 81 pupils gave their written consent for participation (response rate: 94%). However, three pupils were not able to complete all assignments due to medical reasons. Excluding these pupils left data of 78 participants available for analysis. Most pupils were enrolled in the "Sciences" track (68%). The sample consisted for 59% of boys ($n = 46$).

The sample on writing in French (sample B) was collected in the fourth grade of the same secondary school. The participants were 42 pupils within the "Human sciences" ($n = 22$) or "Latin" study track ($n = 20$). All participants gave their written consent for participation in the study (response rate: 100%). The group of participants was composed of 30 girls and 12 boys, all aged 15 or 16 years.

Sample C (scientific reporting) was collected in a statistics course of a pre-master program[1] at a Flemish University (Belgium). Of the 27 students who completed the consent form (response rate: 26%), 26 students participated in one or more phases of the study. Most students were female ($n = 18$) with an average age of 37.8 years ($SD = 8.81$).

---

[1] Successful completion of a pre-master program allows students with a professional bachelor's degree to enroll in a master program.

For the samples collected in secondary education, ethical advice was asked and granted. No ethical advice was required for sample C. Nonetheless, the same ethical guidelines were implemented in collecting the data.

4.4.2   Design and Instruments

The design of all studies replicated the set-up used in the study of Bouwer et al. (2018): a pre-test to capture students' prior knowledge and self-efficacy, an intervention with students randomly allocated to either the criteria or comparative condition and a post-test to measure students' performance. Since data collection took place during covid, data was mainly captured online. Next, each phase is discussed briefly. For more information on the materials that were used, interested readers are referred to the Open Science Framework. Table 4.1 summarizes the essential information per intervention phase for each sample.

4.4.2.1 Pre-test

During the pre-test, students' prior knowledge was mapped using one or more open questions. In sample A, students were presented with a math problem on the topic of speed ("How long does it take to cover a distance of 5.25 km at an average speed of 13.8 m/s?"). Answers were scored using eight criteria that were agreed upon by four domain experts. Students could score either 0 or 1 for each criterion. These criteria tapped into students' procedural (e.g., "Only symbolic language used") and conceptual knowledge regarding physics (e.g., "Correct identification of the physics concepts"). Internal consistency ($\alpha = 0.61$) and inter-rater reliability ($ICC = 0.90$) were checked. Students' prior knowledge in sample B was measured by asking them to write down as many features of a good, emotive text in French they could think of. Two raters independently coded the number of features provided ($ICC = 0.91$). Students in sample C were given a test consisting of five open questions to measure their prior knowledge. Two questions tapped into students' factual knowledge regarding t-tests, while the three other questions required students to interpret the output of a t-test. Rules to score the responses were developed and discussed. Responses were partly double coded by two researchers ($ICC_{Q1} = 0.91$, $ICC_{Q2} = 1$, $ICC_{Q3} = 1$, $ICC_{Q4} = 0.94$, $ICC_{Q5} = 1$).

A survey was administered to measure students' self-efficacy. For sample A, an adapted version of Usher and Pajares' (2009) survey on four sources of self-efficacy for mathematics mapped students' vicarious experience, mastery experience, social persuasion, and psychological state (24 items rated on a six point-scale). In sample B, an adapted version of the Bruning et al. (2013) survey on self-efficacy for writing was administered. The instrument consisted of 15 items that captured students' self-efficacy for ideation, conventions and self-regulation of writing using a slider ranging from 0 to 100. To map students' self-efficacy in sample C, 11 items were developed that measured students' self-efficacy regarding the content to be reported, interpreting statistical results, scientific writing style

Table 4.1 Overview of the three phases of the experimental design for each sample. Aspects printed in bolded italics refer to differences in set-up across the samples

	Sample A (Problem-solving in physics)	Sample B (Writing in French)	Sample C (Scientific reporting)
Pre-test	In class/online	Online	Online
Prior knowledge	One (physics) problem on speed	One open question tapping students' knowledge of aspects of good emotive writing in French	Five open questions tapping students' knowledge on t-tests
Self-efficacy	Adapted survey of Usher and Pajares (2009)	Adapted survey of Bruning et al. (2013)	Survey developed by authors
Intervention	Online—randomized No time limits 8 solutions to problem on speed	Online—randomized No time limits 5 short, French emotive texts	Online—randomized No time limits 6 reports on t-tests
Criteria condition[*]	$n_{students}$: 36; $n_{judgements}$: 8 Online (Qualtrics) ***Feedback not allowed***	$n_{students}$: 21; $n_{judgements}$: 5 Online (Qualtrics) Feedback allowed	$n_{students}$: 13; $n_{judgements}$: 6 Online (Qualtrics) Feedback allowed
Comparative condition	$n_{students}$: 36; $n_{comparisons}$: 10 Online (Comproved) Feedback allowed	$n_{students}$: 21; $n_{comparisons}$: 10 Online (Comproved) Feedback allowed	$n_{students}$: 11; $n_{comparisons}$: 3 Online (Comproved) Feedback allowed
Post-test	In class Solve 2 physic problems ('speed', 'force') 50 min	In class Write emotive text in French 50 min	***Online*** Write report on ANOVA ***No time limit***

Note Aspects in italic indicate differences in set-up between the three samples
[*]*Note* The criteria lists used in the criteria condition can be consulted online

and language use (slider ranging from 0 to 100). These dimensions were aligned with the dimensions of the criteria list that was used in the criteria condition (see Intervention).

4.4.2.2 Intervention

Respectively eight, five and six pieces of work of different quality were selected to be assessed during the intervention phase in samples A, B and C. These works were either constructed based on common mistakes of students (sample A), selected from the texts of previous year (sample B), or selected from an authentic (optional) assignment that students made during the statistics course (sample C). Examples were anonymized in all samples.

Because all peer assessments were set up online, students could be randomly assigned to the comparative or criteria condition (even within classes). Students in the criteria condition scored pieces of work using a predefined criteria list that was

implemented in Qualtrics. The criteria list in sample A was constructed by experts. The same eight criteria that were developed to score students' prior knowledge were rephrased into questions (e.g., "Does the pupil use only symbolic language?") that students had to answer by either 'yes' or 'no'. In the two other samples, the criteria list was adapted from the one used by Bouwer et al. (2018). To assess their peers' emotive texts in French (sample B), students had to judge the vocabulary, spelling, grammar, syntax, and content of a text by awarding maximum four points per criterion. To aid students' judgements, descriptions were provided per criterion that were indicative of a good, mediocre, or poor performance (e.g., descriptions for grammar: one or two grammatical errors—multiple grammatical errors—lots of grammatical errors). In sample C, structure and content, correct interpretation of results, scientific style, and language use had to be judged. Each aspect was further divided into sub criteria that were rated on a five-point scale (0: not at all good, 5: very good). Students rated respectively eight (sample A), five (sample B) or six pieces of work (sample C). Students in sample B and C were also allowed to give open feedback on the strengths and weaknesses of each piece of work. As it was felt that the criteria used to judge the physics problems didn't leave any room for additional open feedback, this was not implemented in the criteria condition of sample A. Consequently, the data of sample A couldn't be used to examine the effect of assessment method on the quality of the feedback (RQ1).

Students in the comparative condition chose the better of two pieces of work presented side-by-side using Comproved (https://comproved.com/en/). Students were instructed to "Choose the most correct or complete solution" (sample A), "Choose the better text" (sample B), or "Choose the report that is overall of better quality" (sample C). Also, they were allowed to give open feedback regarding the strengths and weaknesses of each piece of work (see Fig. 4.1 for screen shots of the implementation in Comproved in sample A). Students in the comparative condition were also provided with the assessment criteria, but the criteria list wasn't discussed with the students. Students made respectively ten (sample A), five (sample B) or three comparative judgements (sample C).

4.4.2.3 Post-test

In the final phase of the experiment, students' performance was captured using a writing task (samples B and C) or by letting students solve two math problems in the context of physics (sample A). Students in sample A and B had only 50 minutes to perform the task, while no time restrictions were given to the students in sample C.

The two math problems concerned the topic of speed ("How much time does it take a cyclist to cover a distance of 17.3 km at an average speed of 6.2 m/s?") and force ("Professor Jones has landed on an unknown planet. A mass of 500 g exerts a force of 17.6 N. What is the gravitational field strength on this planet?"). Students' responses were scored using the same criteria as in the pre-test ($\alpha = 0.56$, *ICC*: 0.82). In sample B, students had to write a short emotive text (between 120 and 150 words) in French that described a confidant from the family with whom they have a strong relationship. The texts were uploaded to Comproved and assessed by ten
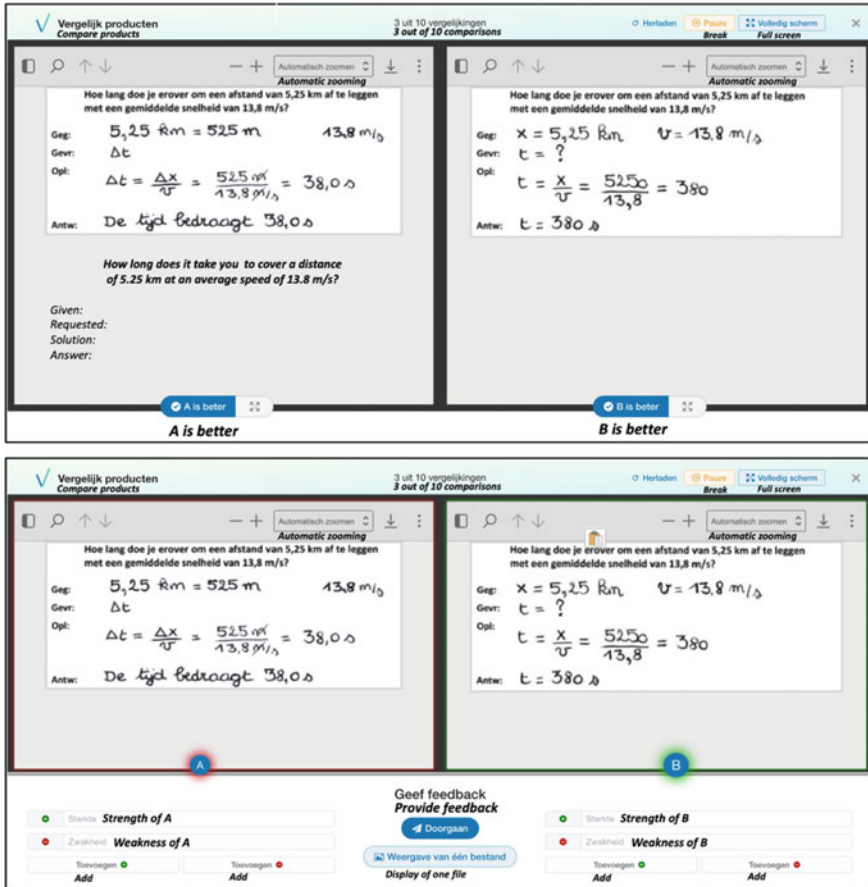
Fig. 4.1 Screen shots of the peer assessment exercise in the comparative condition of sample A (top: comparative judgement, bottom: feedback). Translations are added as bold text

experts. Each expert made 32 comparative judgements which resulted in a reliable rank-order of the texts ($SSR = 0.80$; see Verhavert et al., 2019 for more information on the SSR). Students in sample C were given a research question and the output of a t-test and asked to write a report using that information. The resulting reports were comparatively judged by six experts. Each expert made about 60 comparisons which resulted in rank-order of acceptable reliability ($SSR = 0.61$).

4.4.3 Variables

4.4.3.1 Prior Knowledge, Self-efficacy, and Performance

Students' scores on the prior knowledge tests were summed (in samples A and C). Scores could range between 0 and 8 (sample A and C) and from 0 onwards (sample B). Tables 4.2 and 4.3 provide an overview of the minimum and maximum scores, average and standard deviation of prior knowledge. In sample A, seven students scored the maximum on the prior knowledge test (score of 8). This was accounted for when examining the effect of both approaches on students' performance (RQ2). Prior knowledge was standardized before analysis to facilitate comparison across samples.

Exploratory factor analysis (EFA) with oblique rotation was used to examine the factor structure of the self-efficacy instruments. For sample A, three scales were retained tapping students' mastery experience (4 items, $\alpha = 0.84$), social persuasion (6 items, $\alpha = 0.86$) and psychological state (6 items, $\alpha = 0.84$). EFA on the self-efficacy items of sample B and sample C indicated that only one factor could be retained with respectively 14 items ($\alpha = 0.95$) and 8 items ($\alpha = 0.98$). Full results of EFA can be found online. Items were summed and divided by the number of items to create the variables on self-efficacy. The self-efficacy measures in sample A could range from 1 to 6, while self-efficacy scores were bounded between 0 and 100 in sample B and C (see Tables 4.2 and 4.3). Self-efficacy measures were standardized before analysis.

Table 4.2 Range (Min, Max), mean (*M*) and standard deviation (*SD*) of prior knowledge, sources of self-efficacy and performance for sample A

	Sample A (*n* = 78)			
	Min	Max	*M*	*SD*
Prior knowledge	0.00	8.00	5.42	1.78
Self-efficacy (mastery)	2.25	6.00	4.56	0.73
Self-efficacy (persuasion)	1.00	5.83	2.78	1.11
Self-efficacy (state)	2.00	6.00	5.46	0.72
Performance	7.00	16.00	13.42	1.96

Table 4.3 Range (Min, Max), mean (*M*) and standard deviation (*SD*) of prior knowledge, self-efficacy and performance for sample B and sample C

	Sample B (*n* = 42)				Sample C (*n* = 22)			
	Min	Max	*M*	*SD*	Min	Max	*M*	*SD*
Prior knowledge	0.00	7.00	2.62	1.90	0.20	7.75	4.88	1.83
Self-efficacy	14.64	90.64	53.50	18.39	10.50	76.75	50.39	19.07
Performance	−3.34	3.75	−0.02	1.85	−5.20	3.24	−0.19	2.00

Students' performance refers to their total score on the physics problems (sample A) or the quality of their text (samples B and C). Total scores for problem-solving were calculated by adding up students' scores on the criteria of both post-test problems (range 0 to 16). The quality of the texts was estimated using the comparative judgements of all experts. This resulted in a score per text (expressed in logits) which indicates the probability that this text would be judged as better than an average text. Thus, texts with a positive score are relatively better than an average text, while the opposite is true for texts with a negative score. Tables 4.2 and 4.3 show descriptive statistics of the variables operationalizing students' performance. All performance scores were standardized before analysis.

4.4.3.2 Feedback

To operationalize the quantity and content of the feedback, students open comments were qualitatively coded. First, students' open comments were divided into feedback statements that referred to a single aspect which resulted in 380 statements for sample B and 386 statements for sample C. The variable representing the amount of feedback was created by summing the number of unique aspects mentioned per piece of work. Similarly, a variable representing the amount of positive and of negative feedback was created for sample B (positive: 170, negative: 210) and sample C (positive: 181, negative: 187). Table 4.4 provides descriptive statistics for the variables that represent the total amount of feedback and the amount of positive and negative feedback. Overall, students provided feedback about at least one positive or negative aspect in 77.1% of the judgements in sample B and in all judgements in sample C. Figure 4.2 shows for both samples the relative share of the number of arguments provided (positive and negative) per judgement.

Then, content of the feedback was operationalized by assigning each feedback statement to one of the categories also included in the criteria list. For sample B, this could be either 'content', 'syntax', 'grammar', 'vocabulary', or 'spelling'. The two first categories were considered as statements referring to higher-order aspects of writing in French in the fourth grade of secondary education, the three remaining categories were labelled as lower-order aspects. A distinction was made between positive and negative statements. 14% of all feedback statements were independently coded by two raters resulting in *ICC*'s ranging from 0.73 to 0.96. The feedback statements in sample C were also deductively coded (based on the criteria list) resulting in the categories of 'content', 'interpretation', 'scientific

Table 4.4 Range (Min, Max), mean (*M*) and standard deviation (*SD*) of amount of (positive and negative) feedback statements for sample B and sample C

	Sample B (*n* = 42)				Sample C (*n* = 22)			
	Min	Max	*M*	*SD*	Min	Max	*M*	*SD*
Total amount of feedback	0.00	5.00	1.72	1.47	1.00	7.00	3.15	1.21
Amount of positive feedback	0.00	5.00	0.80	1.03	0.00	5.00	1.62	0.88
Amount of negative feedback	0.00	4.00	0.93	0.99	0.00	4.00	1.53	0.95
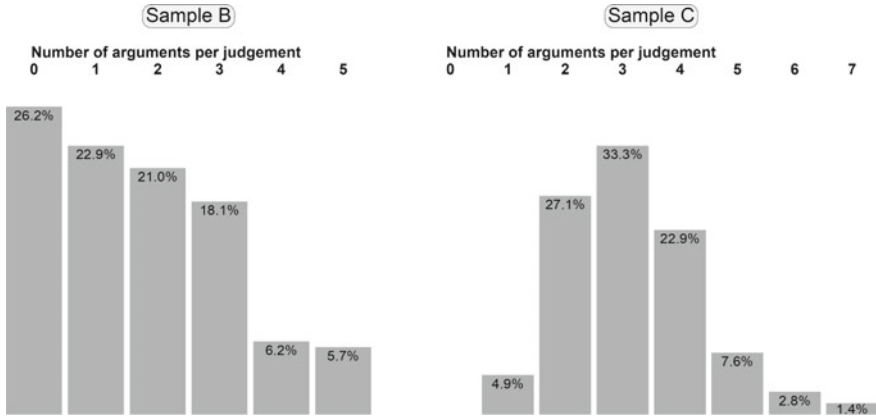
Fig. 4.2 Relative frequencies of number of arguments provided per judgement for sample B and sample C

style', and 'language use'. Three categories were inductively added that referred to making a holistic judgement, the length of the report, or other aspects. The categories of 'content', 'interpretation' and 'scientific writing' were considered as referring to higher-order aspects of scientific writing. As in sample B, positive and negative statements were distinguished. Inter-rater reliability was calculated using Cohen's kappa and ranged from acceptable to good ($0.6 < \kappa < 1$). Sometimes, a student referred more than once to the same category. Therefore, all variables representing the content of the feedback were recoded to dummy variables taking a value of 0 (aspect not mentioned) or 1 (aspect mentioned). Table 4.5 presents descriptive statistics for the dummy variables on the content of the positive and negative feedback.

Table 4.5 Absolute frequency (*N*) and relative frequency (%) of the dummy variables representing the content of the positive and negative feedback for sample B and sample C

Sample B	Positive		Negative		Sample C	Positive		Negative	
	N	*%*	*N*	*%*		*N*	*%*	*N*	*%*
Content	59	28.10	29	13.81	Content	87	60.42	82	56.94
Syntax	39	18.57	50	23.81	Interpretation	0	0.00	30	20.83
Grammar	25	11.90	45	21.43	Scientific style	36	25.00	39	27.08
Vocabulary	31	14.76	33	15.71	*Language use*	39	27.08	43	29.86
Spelling	10	4.76	32	15.24	*Holistic judgement*	19	13.19	4	2.78
					Length	21	14.58	15	10.42
					Other	6	4.17	8	5.56

Note Aspects in *italic* refer to lower-order aspects

4.4.4 Analyses

Before answering both research questions, randomization of students was checked. Randomization failed only in two instances. Students in the comparative condition of sample A ($M = 0.13$, $SD = 1.02$) scored 0.27 SD higher ($t(75.78) = -1.21$, $p = 0.22$, $d = -0.27$) on psychological state (self-efficacy) than students in the criteria condition ($M = -0.13$, $SD = 0.97$). In sample B, students' self-efficacy was 0.22 SD higher ($t(39.79) = 0.70$, $p = 0.49$, $d = 0.22$) in the criteria ($M = 0.11$, $SD = 1.04$) than the comparative condition ($M = -0.11$, $SD = 0.97$). Results of all randomization checks can be consulted online.

The effect of condition on the quality of the feedback provided (RQ1) was tested using generalized cross-classified linear mixed-effect models fitted with the R-package lme4 (version 1.1-25; Bates et al., 2015). These models account for hierarchy in the data by examining the effect of condition on the amount/content of feedback for an average student (fixed effects) while also taking differences in amount/content of feedback between students and between products (random effects) into account (Fielding & Goldstein, 2006). Dependent variables were not normally distributed as they were either counts (amount of feedback) or binary variables (content of feedback). Therefore, generalized mixed-effect models assuming a Poisson-distribution with log-link (amount of feedback) or a binomial distribution with logit-link (content of feedback) were used. Two effect sizes were calculated using the MuMIn-package (version 1.43.17; Bárton, 2022). The marginal R^2 represents differences in amount/content of the feedback attributable to the average effect of condition (fixed effects). Its interpretation is analogous to that of the R^2 in ordinary linear regression models. The conditional R^2 represents the differences in amount/content of feedback that can be explained by the whole model (fixed and random effects). Consequently, the difference between both R^2-statistics gives an indication of variation in amount/content of the feedback attributable to differences between students and between products (random effects). However, interpreting these effect sizes should be done cautiously given that their size depends on the location of the intercept[2] (see Johnson, 2014; Nakagawa & Schielzeth, 2013).

To examine the effect of both assessment methods on students' performance (RQ2), two analyses were performed. First, an independent sample Welch t-test was applied to examine differences in average performance across both conditions. All t-tests were performed assuming unequal variances (see Delacre et al., 2017). Cohen's d was also estimated to gain insight into the size of the effect. Then, a regression analysis was run with condition, prior knowledge and self-efficacy as independent variables and students' performance as dependent variable. For sample A, these analyses were performed using the full data set and a data set excluding information of students with maximum scores on prior knowledge (see 4.4.3 Variables).

---

[2] In this case, the size of the random effects is estimated for students in the criteria condition.

4.5   Results

Results are discussed per research question. Findings on the effect of peer assessment method on the quality of the feedback (RQ1) are presented using visualizations. Tables with full results for RQ1 can be consulted in the appendix of this chapter.

4.5.1   Quality of Feedback

Results regarding sample B show that condition only impacts the number of positive feedback statements that an average student provides. An average student in the comparative condition mentions 0.8 arguments per judgement compared to 0.4 arguments for a student in the criteria condition (see Fig. 4.3). The marginal R^2-statistic points to a moderate effect (marginal $R^2 = 0.09$). The results also indicate that students differ in the total amount of feedback ($SD = 0.93$), in the amount of positive feedback ($SD = 0.99$) and in the amount of negative feedback ($SD = 0.69$) they give.

In study C, opposite results are found as an average student in the criteria condition provides more feedback per judgement (3.5 arguments) than a student in the comparative condition (2.8 arguments; see Fig. 4.3). An average student in the criteria condition also mentions more negative aspects per judgement (1.8 arguments) than a student in the comparative condition (1.2 arguments). The amount of positive feedback does not differ between conditions (1.7 positive arguments per judgement). Also, students do not vary in the amount of feedback
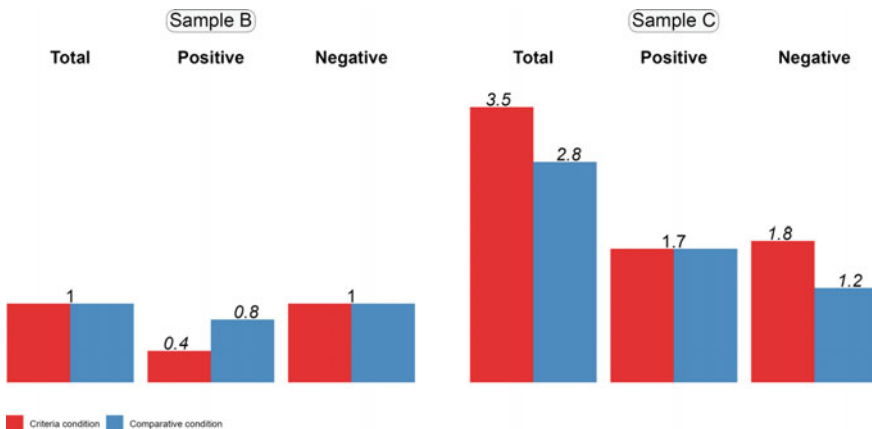
Fig. 4.3   Estimated average number of arguments per judgement by condition
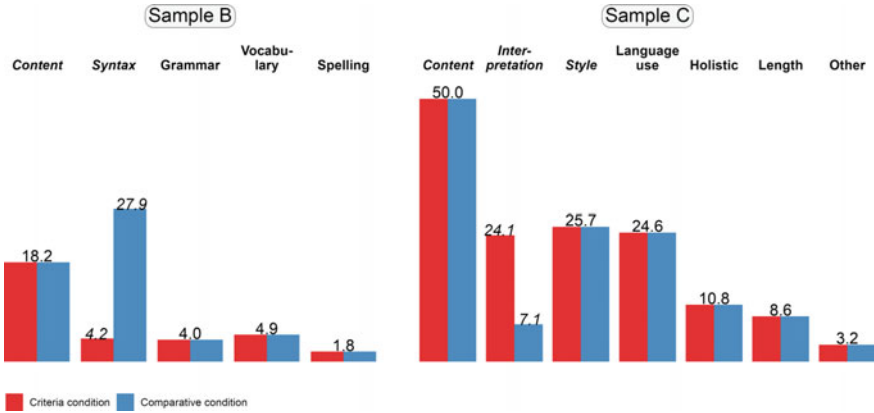
Fig. 4.4 Estimated average probability of giving positive feedback for each quality aspect by condition

they provide which is reflected in the small conditional R^2-statistics (< 0.1). Consequently, differences can be mainly attributed to peer assessment condition. Marginal R^2-statistics point to a small effect for the total amount of feedback (marginal $R^2 = 0.05$) and to a moderate effect for amount of negative feedback (marginal $R^2 = 0.07$). Full results can be consulted in Table 4.10 of the appendix.

Further analysis of the content of the positive feedback shows that the probability of mentioning most aspects is, on average, the same across both conditions. Figure 4.4 visualizes the estimated average probability of mentioning each quality aspect when giving positive feedback for both samples.

Only two differences are found that can be attributed to condition (see Fig. 4.4). First, the average student in the criteria condition of sample B has a lower probability of mentioning the higher-order aspect 'Syntax' than an average student in the comparative condition (4.2% versus 27.9%). This points to a moderate effect (marginal $R^2 = 0.13$). Second, the probability of mentioning the higher-order aspect 'Interpretation' is higher in the criteria than in the comparative condition of sample C (24.1% versus 7.1%). Again, the marginal R^2-statistic indicates a moderate effect (marginal $R^2 = 0.07$). In addition to the effect of condition, it also appears that especially in sample B students differ in the probability of mentioning the higher-order aspect 'Content' ($SD = 1.52$) and the lower-order aspects 'Grammar' ($SD = 2.16$) and 'Vocabulary' ($SD = 2.36$). In sample C, differences between students are only found regarding the higher-order aspect 'Content' ($SD = 1.17$). The complete results of the analyses can be found in Table 4.11 of the appendix.

In-depth analysis of the content of the negative feedback does not find any average effect of condition. Figure 4.5 shows that the estimated average probability of mentioning a quality aspect is the same across both conditions in sample B and
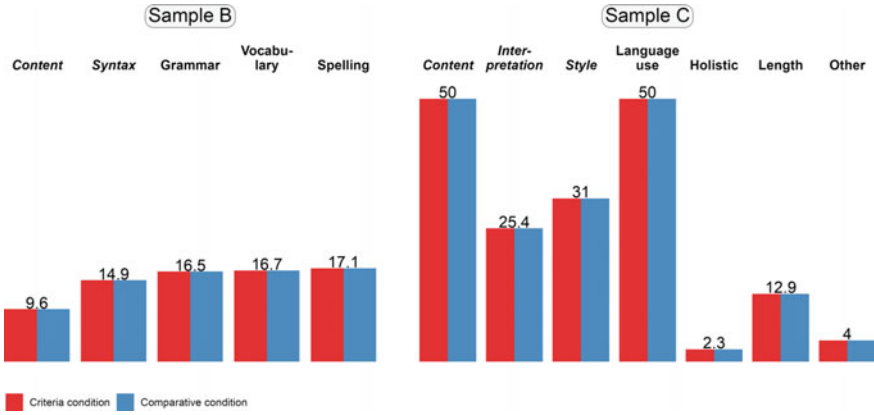
Fig. 4.5 Estimated average probability of giving negative feedback for each quality aspect by condition

sample C. Moreover, all effect sizes are negligible or small (marginal $R^2 \leq 0.03$). Only differences between students or products in the probability of mentioning certain aspects are found. In sample B, the probability of mentioning the higher-order aspect 'Syntax' ($SD = 1.33$) and the lower-order aspect 'Grammar' ($SD = 1.24$) varied across students. Differences between products are found in sample C regarding the higher-order aspect 'Content'. Hence, the probability of mentioning this aspect was higher for some products than for others ($SD = 1.24$). All results on the content of negative feedback can be consulted in Table 4.12 f the appendix.

4.5.2 Effect on Performance

The results in Table 4.6 indicate that students' performance after the intervention doesn't differ across both conditions. The effect sizes (Cohen's d) vary between -0.15 (sample A) and 0.16 (sample B). Hence, whether a student judged their peers' works using criteria or made comparative judgements does not impact their performance differently.

This lack of effect remains after controlling for prior knowledge and self-efficacy (see Tables 4.7 and 4.8). The difference between the criteria and the comparative condition ranges between -0.06 (sample B) and 0.13 (sample C). However, the 95% confidence intervals indicate that this effect cannot be generalized to the population of students in any sample (see Tables 4.7 and 4.8). Thus, assessment method has no differential effect on students' performance in any of the samples.

Table 4.6 Mean (*M*), standard deviation (*SD*) per condition and results of independent sample Welch t-tests

	Criteria condition		Comparative condition		Results of t-test[*]			
	M	*SD*	*M*	*SD*	*t*	*df*	*p*	*d*
Sample A	−0.06	1.15	0.06	0.83	−0.52	69.22	0.61	−0.12
Sample A (filtered)	−0.10	1.17	0.05	0.86	−0.62	62.69	0.53	−0.15
Sample B	0.08	0.90	−0.08	1.11	0.51	38.34	0.61	0.16
Sample C	0.03	0.90	−0.04	1.16	0.14	16.76	0.89	0.06

[*]t-value = *t*, degrees of freedom = *df*, p-value = *p*, Cohen's d = *d*

Table 4.7 Estimates (Est.) and 95% confidence intervals (95% CI) of the regression models that examine the impact of condition on performance and control for prior knowledge and self-efficacy (SE) for sample A

	Sample A (full)		Sample A (filtered)	
	Est.	95% CI	Est.	95% CI
Intercept[*]	−0.03	−0.33|0.27	−0.04	−0.37|0.29
Condition	0.06	−0.37|0.49	0.06	−0.41|0.52
Prior knowledge (Z)	0.11	−0.12|0.34	0.09	−0.17|0.35
SE mastery (Z)	0.15	−0.09|0.40	0.13	−0.12|0.39
SE state (Z)	**0.31**	0.08|0.55	**0.34**	0.10|0.58
SE persuasion (Z)	−0.10	−0.32|0.12	−0.05	−0.30|0.20

Note **Results in bold** can be generalized to the population
Note Prior knowledge and self-efficacy are standardized
[*]Criteria condition is the reference category

Table 4.8 Estimates (Est.) and 95% confidence intervals (95% CI) of the regression models that examine the impact of condition on performance and control for prior knowledge and self-efficacy for samples B and C

	Sample B		Sample C	
	Est.	95% CI	Est.	95% CI
Intercept[*]	0.03	−0.38|0.44	−0.06	−0.53|0.42
Condition	−0.06	−0.63|0.52	0.13	−0.58|0.84
Prior knowledge (Z)	−0.03	−0.34|0.29	**0.74**	0.36|1.13
Self-efficacy (Z)	**0.48**	0.16|0.79	−0.27	−0.66|0.11

Note **Results in bold** can be generalized to the population
Note Prior knowledge and self-efficacy are standardized
[*]Criteria condition is the reference category

4.6   Discussion

This study conceptually replicated the study of Bouwer et al. (2018) on the learning effects of peer assessment using either predefined criteria or comparative judgement. Three small scale studies were set up, two in secondary education (problem-solving in physics, writing in French) and one in university education (scientific reporting of statistical results). After mapping students' prior knowledge and self-efficacy, students were randomly allocated to a peer assessment condition. Students in the criteria condition assessed the work of their peers using a predefined criteria list, while students assigned to the comparative condition made comparative judgements. Students in the study on writing in French and on scientific reporting were also allowed to provide open feedback on the strengths and weaknesses of the works they assessed. Analyses were analogue to the ones performed by Bouwer et al. (2018) and focused on the effect of both assessment methods on the quantity and content of the peer feedback and on students' performance. Overall, the results of the conceptual replications showed that the effects found by Bouwer and colleagues (2018) can only be replicated to a limited extent. Table 4.9 compares the results of the studies reported in this chapter to those of the Bouwer et al.-study (2018). Because the study of Stuulen et al. (2022) had a similar set-up, findings of this study are also added to the table.

Regarding the quantity of the feedback, results of the conceptual replications are partly in line with those of Bouwer et al. (2018). The original study found a

Table 4.9 Comparison of the results regarding the impact of comparative judgement (CJ) and the use of criteria (CRIT) on the quantity and content of feedback and on students' performance found in the three studies reported in this chapter, the study by Bouwer et al. (2018) and the study by Stuulen et al. (2022)

	This chapter			Bouwer et al. (2018)	Stuulen et al. (2022)
	Sample A	Sample B	Sample C		
Quantity of FB°					
Total amount		CJ = CRIT	CJ < **CRIT**		
Amount +		**CJ** > CRIT	CJ = CRIT	CJ = CRIT	**CJ** > CRIT
Amount −		CJ = CRIT	CJ < **CRIT**	CJ = CRIT	CJ < **CRIT**
Content of + FB					
Higher-order		**CJ** > CRIT	CJ < **CRIT**	**CJ** > CRIT	
Lower-order		CJ = CRIT	CJ = CRIT		
Content of—FB					
Higher-order		CJ = CRIT	CJ = CRIT		CJ < **CRIT**
Lower-order		CJ = CRIT	CJ = CRIT		CJ = CRIT
Performance					
	CJ = CRIT	CJ = CRIT	CJ = CRIT	**CJ** > CRIT	CJ = CRIT

*Educational level = Edu. Level; secondary education (SE) or higher education (HE)
°Feedback = FB

positive effect of comparative judgement on the total amount of feedback given. The replication studies also found differences in quantity of feedback across peer assessment conditions. However, the direction of the effect differed between samples. In one sample (secondary education, writing in French), students in the comparative condition gave more positive feedback (in line with Bouwer et al., 2018), while in the other sample (higher education, scientific reporting) the opposite was found as students in the criteria condition gave more negative feedback. These results are in line with those of Stuulen and colleagues (2022) who also found students in the comparative condition to give more positive feedback but less negative feedback than students in the criteria condition. One explanation for the inconclusive findings might be the difference in the number of works students assessed. In the studies by Bouwer et al. (2018), Stuulen et al. (2022) and two of the samples in this chapter, students in the comparative condition judged more pieces of work than their peers in the criteria condition. This gave these students more feedback opportunities than students in the criteria condition. Moreover, the number of judgements could also vary within the comparative condition. Therefore, future research should control for the number of judgements made across and within condition. Another explanation for the results relates to students' initial task experience. Students in the sample on scientific reporting hardly had any experience with the task at hand before the intervention, while students in the study of Bouwer et al. (2018) and in the sample on writing in French all had prior experience with writing in English or French. Hence, students in the sample on scientific reporting might have been less able to fall back on their own understanding of quality than students in the other samples. This might have benefitted students in the criteria condition since they could draw on the predefined criteria to formulate feedback (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013; Sadler, 2009). Future research can investigate to what extent the interaction between students' prior task experience and assessment method influences the amount of feedback students give.

Looking at the results on the content of the feedback, a complex picture emerges. Whereas Bouwer and colleagues (2018) found differences between both assessment methods in the type of negative feedback provided, results of the replication studies presented in this chapter found only two differences across both conditions in the content of the positive feedback that students gave. Students who comparative judged French texts gave more positive feedback on the syntax of the texts, while students in the criteria condition of the sample on scientific reporting provided more positive feedback on the aspect of interpretation. These effects refer in both cases to higher-order aspects of writing which is partly in line with the results of Bouwer et al. (2018) and of Stuulen et al. (2022). Reasons for the differences found are unclear. One important aspect that might have been at play and hasn't been considered in any of the studies is which pieces of work the students assessed. Research into comparative judgement underpins that the aspects assessors look at depends on the pair composition (Lesterhuis, 2018). This also applies when using criteria because students are assumed to compare examples to their own internal standards or previous work (Nicol, 2020). Hence,

future research should be set up that looks at the impact of confronting students with specific (pairs of) examples.

The positive effect of comparative judgement on students' writing performance found by Bouwer et al. (2018) wasn't replicated in any of the samples. This lack of effect is in line with the findings of Stuulen et al. (2022) who also didn't find differences in performance due to peer assessment method. However, the design of the replication studies in this chapter (and of the Bouwer et al.-study) did not allow drawing any conclusions regarding improvement in students' performance due to assessment condition. Furthermore, according to Nicol (2020) and To et al. (2021), it might be beneficial if students already have some experience with a task before engaging in peer assessment. In that case, they already developed a sense of quality and generated internal feedback on their own work (strengths, gaps). This allows them to be more focused on information that is relevant for them during the peer assessment which can enhance the learning effect of the peer assessment exercise. Together this calls for future studies that capture students' performance before the intervention and allow them to revise the same task after they have participated in a peer assessment exercise. Furthermore, future studies should also dig into students' learning processes while engaging in peer assessment. Students' feedback statements only capture those quality aspects that students were aware of and that they reported. These statements do not reveal how students came to noticing these aspects nor how they cognitively process the examples. Therefore, studies should combine feedback statements with online measures such as eye-tracking and log data to fully map students' learning processes. Eye-tracking data and log data provide objective measures of cognitive processes that student engage in (e.g., attention allocation). Replaying students' eye movements can also be used to capture retrospective cued recall data of students' learning processes.

This study set out to conceptual replicate the findings of the study by Bouwer et al. (2018). Overall, it can be concluded that results are only replicated to a limited extent. According to literature on replication research, conceptual replication studies that fail in replicating results add little insights to the scientific knowledge base (Hendrick, 1990). However, replicating the Bouwer et al.-study (2018) three times in a different context sheds light on individual characteristics (e.g., difference in initial task experience) and characteristics of the peer assessment design (e.g., difference in number of judgements) that might explain the variety in the results found. In this respect, a systematic replication study would be interesting. Then, different foundational aspects of a study (e.g., number of judgements made, characteristics of respondents in sample) are systematically varied, while the hypothesis behind the study (learning effect of peer assessment method) is retained (Hendrick, 1990). This would provide systematic insight into which type of peer assessment method (use of criteria or comparative judgement) and peer assessment design (e.g., number of judgements, type of exemplars) is most beneficial for which student (e.g., high prior task experience).

Some additional limitations of this study should be mentioned. First, results are based on small samples making the results of this study more uncertain. Further replication research that uses bigger samples is needed. Second, all samples were

collected amidst the covid-pandemic. Consequently, some procedures were less controlled than common in experimental designs. This is especially important for the peer assessment exercises which were run online in all samples. Although this mirrors actual classroom conditions, it makes it unclear to what extent students engaged in the peer assessment exercise as intended which can have biased the results. Finally, students' effort and time investment were not considered which might have confounded the results of this study. Despite these limitations, this study provided insights into the effect of using criteria and comparative judgement in the context of peer assessment. Furthermore, it also highlights the need for (conceptual) replication studies within the educational sciences as this can shed light on the replicability of effects and can provide an avenue for further research and theory development.

Appendix

See Tables 4.10, 4.11 and 4.12.

Table 4.10 Results of the generalized mixed-effect models on the effect of condition on the total amount of feedback, the amount of positive feedback and the amount of negative feedback: estimates (Est.) and bootstrapped 95% confidence intervals (95% CI) of the fixed (average impact of condition expressed in logs) and random effects (differences between students and products expressed in standard deviations), including estimates of effect sizes

Sample	Type feedback	Fixed effects*				Random effects				Effect sizes[a]	
		Intercept		Condition		Students		Products		Mar. R^2	Con. R^2
		Est.	95% CI	Est.	95% CI	SD	95% CI	SD	95% CI		
B	Total	0.10	−0.28|0.60	0.33	−0.33|0.93	**0.93**	0.64|1.17	0.00	0.00|0.14	0.02	0.69
	Positive	**−1.01**	−1.39|−0.33	**0.80**	0.02|1.44	**0.99**	0.64|1.23	0.13	0.00|0.29	0.09	0.63
	Negative	−0.27	−0.61|0.23	0.06	−0.50|0.55	**0.69**	0.43|0.87	0.30	0.00|0.50	0.00	0.46
C	Total	**1.25**	1.13|1.37	**−0.23**	−0.42|−0.04	0.07	0.00|0.18	0.00	0.00|0.13	0.05	0.06
	Positive	**0.52**	0.35|0.69	−0.08	−0.33|0.18	0.06	0.00|0.23	0.03	0.00|0.19	0.00	0.01
	Negative	**0.59**	0.39|0.80	**−0.40**	−0.65|−0.11	0.00	0.00|0.24	0.14	0.00|0.28	0.07	0.10

Note **Results in bold** can be generalized to the population

*Criteria condition is the reference category

[a]Mar. R^2 = marginal R^2; Con. R^2 = conditional R^2

Table 4.11 Results of the generalized mixed-effect models on the effect of condition on the content of the positive feedback: estimates (Est.) and bootstrapped 95% confidence intervals (95% CI) of the fixed (average impact of condition expressed in logits) and random effects (differences between students and products expressed in standard deviations), including estimates of effect sizes

Sample	Content feedback	Fixed effects*				Random effects				Effect sizes[a]	
		Intercept		Condition		Students		Products		Mar. R^2	Con. R^2
		Est.	95% CI	Est.	95% CI	SD	95% CI	SD	95% CI		
B	Content	**−1.50**	−2.08l−0.58	0.66	−0.53l1.70	**1.52**	0.68l2.24	0.00	0.00l0.49	0.02	0.34
	Syntax	**−3.12**	−4.65l−1.76	**2.17**	0.87l3.53	1.10	0.00l1.80	0.87	0.00l1.61	0.13	0.36
	Gram.°	**−3.18**	−3.77l−1.53	0.97	−0.66l2.26	**2.16**	0.87l3.15	0.00	0.00l0.62	0.02	0.43
	Vocabu.°	**−2.96**	−3.58l−1.20	0.87	−0.83l2.27	**2.36**	1.08l3.36	0.42	0.00l0.95	0.02	0.51
	Spelling	**−3.98**	−22.57l−2.64	1.38	−0.07l19.95	0.95	0.00l1.90	0.00	0.00l0.81	0.03	0.08
CC	Content	0.68	−0.16l1.31	−0.41	−1.34l0.73	**1.17**	0.30l1.85	0.00	0.00l0.59	0.01	0.25
	Interpr.°	**−1.15**	−1.92l−0.41	**−1.42**	−3.13l−0.22	0.92	0.00l1.62	0.00	0.00l0.64	0.07	0.18
	Style	**−1.06**	−1.86l−0.28	−0.25	−1.21l0.76	0.73	0.00l1.38	0.49	0.00l0.91	0.00	0.14
	Langu.°	**−1.12**	−1.86l−0.42	0.18	−0.70l1.16	0.77	0.00l1.28	0.00	0.00l0.52	0.00	0.11
	Holistic	**−2.11**	−3.29l−1.02	0.07	−1.22l1.29	1.02	0.00l1.85	0.66	0.00l1.32	0.00	0.17
	Length	**−2.37**	−3.64l−1.41	1.00	−0.03l2.30	0.60	0.00l1.30	0.50	0.00l1.12	0.03	0.11
	Other	**−3.34**	−22.54l−2.03	0.20	−18.20l19.06	0.76	0.00l1.94	0.72	0.00l1.90	0.00	0.06

Note **Results in bold** can be generalized to the population

Note Aspects in *italic* refer to lower-order aspects

*Criteria condition is the reference category

[a]Mar. R^2 = marginal R^2; Con. R^2 = conditional R^2

°Grammatica = Gram.; Vocabulary = Vocabu.; Interpretation = Interpr.; Language use = Langu.

Table 4.12 Results of the generalized mixed-effect models on the effect of condition on the content of the negative feedback: estimates (Est.) and boot-strapped 95% confidence intervals (95% CI) of the fixed (average impact of condition expressed in logits) and random effects (differences between students and products expressed in standard deviations), including estimates of effect sizes

Sample	Content feedback	Fixed effects*				Random effects				Effect sizes[a]	
		Intercept		Condition		Students		Products		Mar. R^2	Con. R^2
		Est.	95% CI	Est.	95% CI	SD	95% CI	SD	95% CI		
B	Content	**−2.24**	−3.61|−1.20	0.27	−0.54|1.21	0.00	0.00|0.98	1.00	0.00|1.91	0.00	0.11
	Syntax	**−1.74**	−2.41|−0.65	0.60	−0.52|1.46	**1.33**	0.46|1.87	0.74	0.00|1.34	0.01	0.32
	Gram.°	**−1.62**	−2.25|−0.64	0.22	−0.76|1.10	**1.16**	0.28|1.72	0.65	0.00|1.18	0.00	0.25
	Vocabu.°	**−1.61**	−2.17|−0.87	−0.35	−1.27|0.53	0.91	0.00|1.39	0.33	0.00|0.71	0.00	0.13
	Spelling	**−1.58**	−2.19|−0.65	−0.74	−1.67|0.28	1.12	0.00|1.68	0.50	0.00|0.95	0.02	0.20
CC	Content	0.75	−0.52|1.95	−0.86	−1.84|0.32	1.03	0.00|1.65	**1.24**	0.14|2.10	0.03	0.41
	Interpr.°	**−1.08**	−1.70|−0.50	−0.65	−1.61|0.29	0.53	0.00|1.16	0.00	0.00|0.61	0.02	0.06
	Style	**−0.80**	−1.54|−0.16	−0.58	−1.45|0.45	0.80	0.00|1.37	0.38	0.00|0.82	0.01	0.15
	Langu.°	−0.55	−1.28|0.08	−0.80	−1.68|0.17	0.73	0.00|1.22	0.20	0.00|0.59	0.03	0.14
	Holistic	**−3.75**	−23.57|−2.19	0.02	−19.88|19.92	1.27	0.00|2.60	0.00	0.00|1.36	0.00	0.06
	Length	**−1.91**	−2.92|−0.95	−1.03	−2.61|0.32	0.96	0.00|1.70	0.53	0.00|1.17	0.03	0.14
	Other	**−3.17**	−4.60|−1.89	0.35	−1.61|2.36	1.13	0.00|2.21	0.00	0.00|0.98	0.00	0.08

Note **Results in bold** can be generalized to the population

Note Aspects in *italic* refer to lower-order aspects

*Criteria condition is the reference category

[a]Mar. R^2 = marginal R^2; Con. R^2 = conditional R^2

°Grammatica = Gram.; Vocabulary = Vocabu.; Interpretation

References

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist, 48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712

Bartholomew, S. R., Strimel, G. J., & Yoshikawa, E. (2019). Using adaptive comparative judgment for student formative feedback and learning during a middle school design project. *International Journal of Technology and Design Education, 29*(2), 363–385. https://doi.org/10.1007/s10798-018-9442-7

Bárton, K. (2022). *Package 'MuMIn'.* https://cran.r-project.org/web/packages/MuMIn/MuMIn.pdf

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *67*(1), 48. https://doi.org/10.18637/jss.v067.i01

Bell, A., Mladenovic, R., & Price, M. (2013). Students' perceptions of the usefulness of marking guides, grade descriptors and annotated exemplars. *Assessment and Evaluation in Higher Education, 38*(7), 769–788. https://doi.org/10.1080/02602938.2012.714738

Bloxham, S., & Boyd, P. (2007). *Developing Effective Assessment in Higher Education.* Open University Press.

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151–167. https://doi.org/10.1080/713695728

Bouwer, R., Béguin, A., Sanders, T., & Van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing, 32*, 83–100. https://doi.org/10.1177/0265532214542994

Bouwer, R., Lesterhuis, M., Bonne, P., & De Maeyer, S. (2018). Applying criteria to examples or learning by comparison: Effects on students' evaluative judgment and performance in writing. *Frontiers in Education, 3*, 86. https://doi.org/10.3389/feduc.2018.00086

Bruning, R., Dempsey, M., Kauffman, D. F., McKim, C., & Zumbrunn, S. (2013). Examining dimensions of self-efficacy for writing. *Journal of Educational Psychology, 105*(1), 25–38. https://doi.org/10.1037/a0029692

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment and Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Carless, D., & Chan, K. K. H. (2017). Managing dialogic use of exemplars. *Assessment and Evaluation in Higher Education, 42*(6), 930–941. https://doi.org/10.1080/02602938.2016.1211246

Chapelle, C. A., Enright, M. K., & Jamieson, J. (2008). *Building a validity argument for the test of English as a foreign language.* Routledge.

Cumming, A., Kantor, R., & Powers, D. (2002). Decision making while rating ESL/EFL writing tasks: A descriptive framework. *The Modern Language Journal, 86*(1), 67–96. https://doi.org/10.1111/1540-4781.00137

Delacre, M., Lakens, D., & Leys, C. (2017). Why Psychologists should by default use Welch's t-test instead of student's t-test. *International Review of Social Psychology, 30*(1), 92–101. https://doi.org/10.5334/irsp.82

Eckes, T. (2008). Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing, 25*(2), 155–185. https://doi.org/10.1177/0265532207086780

Fielding, A., & Goldstein, H. (2006). *Cross-classified and multiple membership structures in multilevel models: An introduction and review.* https://dera.ioe.ac.uk/6469/1/RR791.pdf

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science, 34*(5), 752–775. https://doi.org/10.1111/j.1551-6709.2010.01114.x

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*(1), 45–56. https://doi.org/10.1037/0003-066X.52.1.45

Handley, K., & Williams, L. (2011). From copying to learning: Using exemplars to engage students with assessment criteria and feedback. *Assessment and Evaluation in Higher Education, 36*, 95–108. https://doi.org/10.1080/02602930903201669

Hendrick, C. (1990). Replications, strict replications and conceptual replications: Are they important? *Journal of Social Behavior and Personality, 5*(4), 41–49.

Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth's R^2 *GLMM* to random slopes models. *Methods in Ecology and Evolution, 5*(9), 944–946. https://doi.org/10.1111/2041-210X.12225

Jones, I., & Alcock, L. (2014). Peer assessment without assessment criteria. *Studies Higher Education, 39*, 1774–1787. https://doi.org/10.1080/03075079.2013.821974

Jones, I., & Wheadon, C. (2015). Peer assessment using comparative and absolute judgment. *Studies in Educational Evaluation, 47*, 93–101. https://doi.org/10.1016/j.stueduc.2015.09.004

Jonnson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*, 130–144. https://doi.org/10.1016/j.edurev.2007.05.002

Lesterhuis, M., van Daal, T., Van Gasse, R., Coertjens, L., Donche, V., & De Maeyer, S. (2018). When teachers compare argumentative texts: decisions informed by multiple complex aspects of text quality. *L1-Educational Studies in Language and Literature. 18*, 1–22. https://doi.org/10.17239/L1ESLL-2018.18.01.02

Lesterhuis, M. (2018). The validity of comparative judgement to assess text quality: an assessor's perspective (Unpublished doctoral dissertation). University of Antwerp, Antwerp.

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing, 18*(1), 30–43. https://doi.org/10.1016/j.jslw.2008.06.002

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*, 199–218. https://doi.org/10.1080/03075070600572090

Norton, L. (2004). Using assessment criteria as learning criteria: A case study in psychology. *Assessment & Evaluation in Higher Education, 29*(6), 687–702. https://doi.org/10.1080/0260293042000227236

Orsmond, P., Merry, S., & Reiling, K. (2002). The use of exemplars and formative feedback when using student derived making criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education, 27*(4), 309–323. https://doi.org/10.1080/0260293022000001337

Pachur, T., & Olsson, H. (2012). Type of learning task impacts performance and strategy selection in decision making. *Cognitive Psychology, 65*(2), 1–34. https://doi.org/10.1016/j.cogpsych.2012.03.003

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review, 9*, 129–144. https://doi.org/10.1016/j.edurev.2013.01.002

Patchan, M. M., Schunn, C. D., & Correnti, R. J. (2016). The nature of feedback: How peer feedback features affect students' implementation rate and quality of revisions. *Journal of Educational Psychology, 108*(8), 1098–1120. https://doi.org/10.1037/edu0000103

Pollitt, A. (2012a). Comparative judgement for assessment. *International Journal of Technology and Design Education, 22*(2), 157–170. https://doi.org/10.1007/s10798-011-9189-x

Pollitt, A. (2012b). The method of adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice, 19*(3), 281–300. https://doi.org/10.1080/0969594X.2012.665354

Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education., 28*, 147–164. https://doi.org/10.1080/02602930301671

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–144.

Sadler, D. R. (2009). Transforming holistic assessment and grading into a vehicle for complex learning. In G. Joughin (Ed.), *Assessment, learning and judgement in higher education* (pp. 45–63). Springer.

Schoonen, R., Vergeer, M., & Eiting, M. (1997). The assessment of writing ability: Expert readers versus lay readers. *Language Testing, 14*(2), 157–184. https://doi.org/10.1177/026553229701 400203

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology, 13*(2), 90–100.

Seery, N., Canty, D., & Phelan, P. (2012). The validity and value of peer assessment using adaptive comparative judgement in design driven practical education. *International Journal of Technology and Design Education, 22*(2), 205–226. https://doi.org/10.1007/s10798-011-9194-0

Stuulen, J., Bouwer, R., & van den Bergh, H. (2022). Peerfeedback geven en gebruiken voor revisie; Een vergelijking tussen twee peerfeedbackmethoden. *Levende Talen Tijdschrift, 23*(1), 20–35. https://lt-tijdschriften.nl/ojs/index.php/ltt/article/view/2247

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3

To, J., Panadero, E. & Carless, D. (2021). A systematic review of the educational uses and effects of exemplars, Assessment & Evaluation in Higher Education, 47(8):1167–1182. https://doi.org/10.1080/02602938.2021.2011134

Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*, 249–276. https://doi.org/10.3102/00346543068003249

Topping, K. J. (2009). Peer assessment. *Theory into Practice, 48*(1), 20–27. https://doi.org/10.1080/00405840802577569

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education, 14*, 281–294. https://doi.org/10.1080/09695940701591867

Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology, 34*(1), 89–101. https://doi.org/10.1016/j.cedpsych.2008.09.002

Verhavert, S., Bouwer, R., Donche, V., & De Maeyer, S. (2019). A meta-analysis on the reliability of comparative judgement. *Assessment in Education: Principles, Policy & Practice, 26*(5), 541–562. https://doi.org/10.1080/0969594X.2019.1602027

Williamson, M. M., & Huot, B. A. (1992). *Validating holistic scoring for writing assessment: Theoretical and empirical foundations*. Hampton Press.

Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.

Zimmerman, B. J., & Risemberg, R. (1997). Self-regulatory dimensions of academic learning and motivation. In G. D. Phye (Ed.), *Handbook of academic learning: Construction of knowledge* (pp. 105–125). Academic.

Using Stochastic Actor-Oriented Models to Explain Collaboration Intentionality as a Prerequisite for Peer Feedback and Learning in Networks

5

Jasperina Brouwer and Carlos A. de Matos Fernandes

5.1   Introduction

Rooted in social constructivism (Vygotsky, 1978), within the student-centered learning environments students actively co-construct their knowledge in interaction with their peers, which is crucial within learning practices for deep learning (Baeten et al., 2010; O'Donnell, 2006). Next to peer interaction, higher education students discuss the study material, undertake hands-on assignments and provide each other peer feedback. Although peer feedback is often related to assessment, it can also be considered a learning practice within student-centered learning environments (Boud et al., 2001). In the current chapter, we follow Dingyloudi and Strijbos (2018) who go beyond the assessment framework of feedback and task-specific feedback and consider peer feedback more broadly as a process of interpersonal communication contributing to students' learning and performance. Peer feedback is a way in which students share their knowledge, advice, information, and learning experiences. Importantly, peer feedback takes place within the social context of the small group learning environment (i.e., learning communities, see Brouwer et al., 2018, 2022) and is based on the sociocultural perspective

---

We have no conflicts of interest to disclose.

J. Brouwer (✉)
Department of Educational Sciences, University of Groningen, Groningen, The Netherlands
e-mail: jasperina.brouwer@rug.nl

C. A. de Matos Fernandes
Department of Sociology/Interuniversity Center for Social Science Theory and Methodology (ICS), University of Groningen, Groningen, The Netherlands
e-mail: c.a.de.matos.fernandes@rug.nl

implying that learning is a social rather than merely a cognitive phenomenon (Vygotsky, 1978). Thus, we define feedback broadly in terms of academic help and advice-seeking in peer networks.

Peer feedback happens among students who are similar in status and educational level (Finn & Garner, 2011) and provide each other with information related to their performance, also informally outside the classroom. The advantage of these informal forms of peer feedback is that it is a safe and convenient way to increase their ability to advance in higher education. Peers are considered as equals and when provided in a non-evaluative way, it is less likely that peer feedback decreases their self-esteem. Moreover, the feedback is often more immediate and timely than feedback provided by the course instructors or teachers (Laydshewsky, 2013). The fact that non-evaluative and informal peer feedback takes place outside the classroom means that students actively need to seek feedback from their peers. Aleven et al. (2003) identified different steps for approaching a peer when he or she needs feedback to get a better understanding of the study material. First, they need to be aware that they need academic support and feedback. Second, they need to know who is an advanced peer who can provide adequate feedback (Sangin et al., 2011). Third, they need to initiate contact and ask for feedback, academic help or advice. Fourth, the other is willing to provide timely and adequate feedback. Fifth, students collaborate, help each other, and provide each other with feedback.

An important means-to-an-end to facilitate feedback processes comprises network relations. That is, network relations are one of the most important sources of support, help, advice, or peer feedback when they are study partners in higher education (Brouwer et al., 2018, 2022; Stadtfeld et al., 2019). For learning, it is crucial that peers do not merely interact, but that students are willing to function as scaffolds by sharing their knowledge from different perspectives (Sangin et al., 2011). However, students seem to prefer to ask for academic support from their friends, who are, in turn, more or less similar to them in terms of background characteristics or attitudes (Brouwer et al., 2018). This is consistent with an important network selection mechanism (i.e., to initiate a network connection), which is the so-called homophily or similarity effect. Homophily, famously known as the social mechanism "birds of a feather flock together" (McPherson et al., 2001), represents the tendency to preferentially connect to similar others. Similarity can be based on individual features such as gender, ethnicity, or achievement (Lomi et al., 2011; McPherson et al., 2001; Stadtfeld et al., 2019), but also on attitudes (McPherson et al., 2001), such as the intention to collaborate and the willingness to provide feedback and support. Another strand of research posits that similarity in individual features is based on influence mechanisms (Snijders et al., 2010; Steglich et al., 2010), stressing that network relations are social conduits through which individuals influence each other to behave similarly. We explain the role of selection and influence mechanisms in what follows as well as in Fig. 5.1.

Peer collaboration intentionality is a selection mechanism that may play a role in feedback seeking. Collaboration intentionality (CI), which is students' willingness to collaborate, seems an important prerequisite for peer feedback. Research within the educational context shows that school principals' and teachers' network
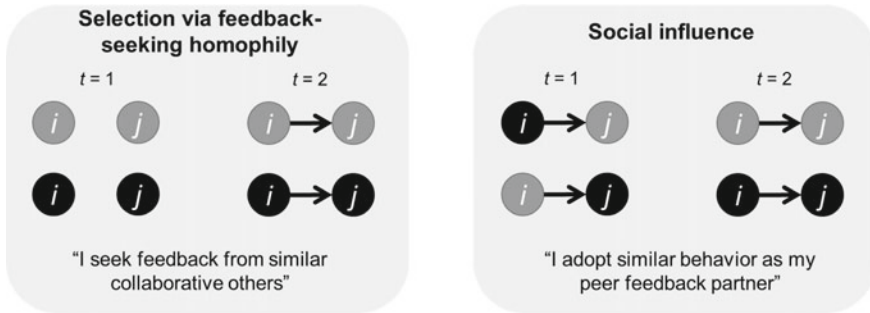
Fig. 5.1 A simple visualization between two individuals. Selection via homophily assumes that individual *i* preferentially nominate a similar other, *j*, for seeking feedback from (similarity is indicated via the color of the node), while influence assumes that a student adjusts his or her collaborative behavior (color of *i*) to behavior shown by peer feedback partners (*j*)

intentionality is associated with social capital formation. Network intentionality refers to the intention of someone to actively connect and interact with other network members (Coleman, 1990; Moolenaar et al., 2014). Van Waes et al. (2015) demonstrate that university teachers who are more *intentional*, actively seek advice and information from their colleagues about teaching. Someone has agency in actively initiating connections when this is of instrumental value, for example, for receiving ideas or feedback. Similarly, peer feedback can only take place within a collaborative learning approach and when students are willing to initiate feedback relationships with their peers (Er et al., 2021). In this respect, social exchange theory (Blau, 1964; Cook & Rise, 2003; Homans, 1961) helps us to understand why someone is willing to help a peer. The social exchange theory posits that someone is willing to do this when a valuable return is expected. Spitzmuller and Dyne (2013) distinguish reactive helping and proactive helping. The former means that others are supported because providing support is the social norm, whereas the latter is beneficial for the helpers contributing to their reputation and self-esteem. Students may also maintain a feedback relationship, for example, when a relationship is assumed to maintain mutually beneficial social exchange relationships (e.g., they obtain both higher grades).

Yet, to understand this complex link between peer feedback relationships and CI, we need to account for selection and influence mechanisms in feedback-seeking networks (Lomi et al., 2011; Snijders et al., 2010). *Selection* comprises whether students preferentially seek feedback from other fellow students because they have similar scores on CI. *Influence* means that students become more similar in CI over time when they provide each other feedback. Influence is an umbrella term for peer influence and social learning (e.g., Bandura, 1977; Steglich et al., 2010). Essentially, influence posits that network relations in place allow connected peers to influence one another in their collaboration, attitudes, opinions, and other behavioral topologies.

In this chapter, selection concerns that someone initiates to form a feedback relation, whereas influence is about the effect of feedback partners on one's CI. The feedback relation is either present (influence) or is under question whether it will be formed or not (selection). Influence and selection are social processes with opposite roles assigned to feedback-seeking relations and collaborative behavior, as indicated in Fig. 5.1. Influence affects CI. Selection, conversely, does not alter CI but only changes the network relation. The striking consequence of homophilous selection and social influence is that the outcome is the same: Connected peers tend to be similar on a certain individual feature (Fig. 5.1).

Not only peer feedback relations may be important for CI, but also gender and personality characteristics play a key role in collaboration and feedback processes (see Noroozi et al., 2020, 2022). Some research, for instance, shows that females tend to express more prosociality than males (Höglinger & Wehrli, 2017) and that this tendency for prosociality is stable over time (de Matos Fernandes et al., 2022). The Five-Factor Model (FFM) of personality consists of a taxonomy of five self-reported traits (McCrae & John, 1992): extraversion (being extravert rather than reserved), agreeableness (altruistic or oriented to cooperate rather than being selfish), openness to new experiences (rather than keeping conventions), conscientiousness (being self-organized rather than disorganized), and neuroticism (being anxious rather than calm). Previous work shows that FFM personality traits, particularly extraversion, agreeableness, and openness, positively affect seeking help or feedback from peers in higher education (Atik & Yalçin, 2011). Moreover, someone who has higher scores on agreeableness seems to be more intended to collaborate (Thielmann et al., 2020). In the current chapter, the main focus is on CI in peer feedback networks, while we control for gender and personality traits.

The interdependence of the social network data and of selection and influence urges researchers to employ a complementary statistical method, namely stochastic actor-orientated models (SAOMs) (Snijders, 2017; Steglich et al., 2010) to dissect underlying mechanisms that give rise to CI—or other individual attributes, such as gender or personality traits—similarity among peers. This approach is necessary because it remains otherwise unclear why students become similar in terms of CI within the feedback network over time. Influence and selection are competing mechanisms but SAOMs allow disentangling one from the other (and vice versa). We introduce this method in our chapter and provide an example using longitudinal feedback-seeking network data of 95 first-year students in higher education.

Although peer feedback takes place within peer networks, to our knowledge, it has been rarely investigated from a network perspective. One of the few examples is Dingyloudi and Strijbos (2018) who investigated peer feedback within learning communities. We want to go beyond Dingyloudi and Strijbos' work by applying the advanced SAOM method to disentangle selection from influence within peer feedback networks regarding CI. These peer networks are collected at two-time points and considered longitudinally in these models (Ripley et al., 2021). Analysis of longitudinally collected social network data informs us about the changes in the relationships and behavior simultaneously (i.e., the network dynamics) and by doing so, the underlying mechanisms of relationship formation within the learning

context. This is the so-called co-evolution modeling and allows us to investigate how social networks and attributes, such as characteristics, behavior, or attitudes change over time (Kalish, 2020; Snijders et al., 2010). In this chapter, we will address the following research question: To what extent does homophily of CI plays a role in selecting peers for feedback (selection), and to what extent do peer feedback relationships influence CI (i.e., social influence of CI)? We investigate the co-evolution of peer feedback-seeking network data (i.e., study-related advice or help-seeking) and CI, which is an individual attribute or in SOAM terms a "behavior" variable. We control for the impact of gender, personality traits, and whether feedback providers are friends. SAOM will be further explained in the next section.

The outline of our chapter is as follows. First, we introduce stochastic actor-oriented models and provide examples of the method. Second, we illustrate how this method can be applied to investigate CI within peer feedback networks. Overall, we introduce a new way to investigate peer feedback within longitudinal social network designs, which provides us a better understanding of how students select each other in terms of CI when seeking feedback and to what extent social influence from feedback seeking plays a role regarding CI? By doing so, we can address research questions about social network dynamics and get a better understanding of social mechanisms, such as social selection (e.g., homophily) and social influence. More specifically, do students ask for feedback from a peer who is similar in terms of the intentionality to collaborate, or do students become similar over time in terms of the intentionality to collaborate when they ask each other for feedback?

5.2   Introducing Stochastic Actor-Oriented Models

Stochastic actor-oriented models (SAOMs) represent an important methodological breakthrough in modeling the interdependence of networks and behavior. What do the following terms mean, such as 'stochastic', 'actor-oriented', and 'models'? SAOMs are *stochastic* given that they model changes in network and behavior via an individual decision-making model; SAOMs are *actor-oriented* given that students (i.e., actors) are the locus of modeling (oriented), instead of networks or groups of people. It is assumed that network and behavior changes are due to students' decisions; SAOMs are *models* because the simulation procedure ensures that we control for all possible interdependent network and behavior states between both waves (Kalish, 2020; Snijders, 2017; Snijders et al., 2010; Steglich et al., 2010). The term *behavior* is an umbrella term for individual attributes such as attitudes, opinions, grades, CI, smoking, drinking, bullying, and many more individual characteristics that change over time. *Networks* refer to friendship networks but they also comprise peer feedback-seeking networks, online social networks, acquaintance networks, positive or negative interactions in a network context, workplace networks, and many more other situations in which individuals are linked to one another in a network. Using SAOMs, we can test how behavior and the network co-evolve from one point in time to another. The role of feedback

is thus not only assessed theoretically, but it is also an inherent part of the SAOM approach. Namely, SAOMs operate in a feedback loop: behavior affects the network, whereas networks affect changes in behavior. A change in CI spills over to the feedback network, and a change in the network affects CI.

What kind of data do we need for SAOMs? SAOMs enable exploring interdependent longitudinal network and behavioral data (see Steglich et al., 2010), which permits researchers to link antecedents to the consequence of peer feedback-seeking network and CI change. To do so, the data requirements of SAOMs comprise complete (socio-centric) network and behavioral data (i.e., individual attributes) from at least two-time points to estimate co-evolution (Steglich et al., 2010; Veenstra & Steglich, 2012). Complete network data refer to whole networks with a specified boundary, e.g., within a school class, which may vary from 20 to 400 individuals (Niezink, 2018). The advantage of, for example, nominating students within one school class is that it informs us also about non-selection. Not selecting someone as a network partner is a requirement to understand selection (Steglich et al., 2010; Veenstra & Steglich, 2012). To know whether similarity in behavior (i.e., homophily) plays a role in selecting someone as a friend, we need to be informed about whether students who select each other are similar in terms of behavior and when students who do not select each other differ in terms of behavior.

How does the modeling take place within SAOM in the background? Changes in the network and behavior between waves are simulated via mini-steps. Mini-steps follow the actor-oriented paradigm that changes in the network or behavior are driven by individual choices (Ripley et al., 2021; Snijders, 2005; Snijders et al., 2010). In other words, each actor (i.e., individual or student) can make one change in his/her network connection or one change in the behavior variable (here CI) in each step. These steps are simulated based on longitudinal data and then estimated with a probability function based on changes in-between measured data waves. Thus, SAOMs build on the inherent assumption that students have a say over with whom they form network ties and in what way they change the initiative towards collaborative behavior (CI). Within the so-called mini-steps simulation procedure, an actor can decide in each step to form, dissolve, or maintain a network relation or report a higher or lower value on the behavior variable (see Fig. 5.2). A so-called mini-step thus captures a change in network relationships and a behavior change.

How many mini-steps—or, i.e., changes—students can take is modeled via the *rate function*, while which mini-step to take is determined by the *objective function*. The rate function provides a numerical value of how many changes a student can make in network relations or CI. Conversely, the objective function shows how attractive a network state or change in behavior for a student is, thereby controlling for various structural network (e.g., reciprocity, transitivity) parameters. 'Attractiveness' comprises, for example, whether it is attractive to change behavior to 6 instead of 4 (Fig. 5.2). Alternatively, in the network context: whether forming or maintaining no relation (for the blue actor in Fig. 5.2) is more attractive than
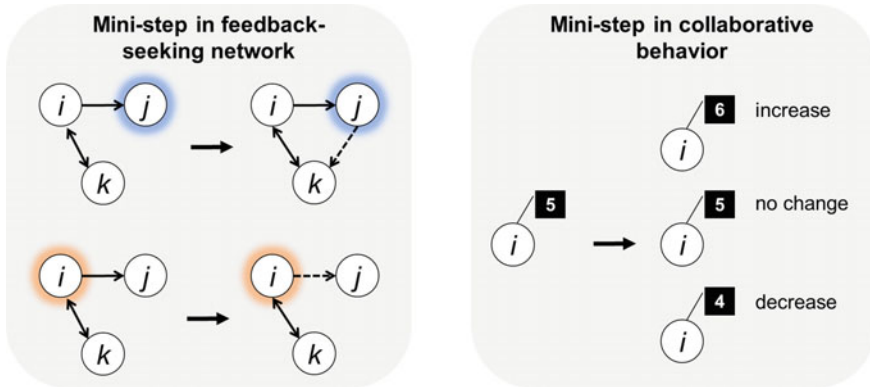
Fig. 5.2 Examples of so-called mini-steps in network selection and behavioral changes. On the left, one actor in blue (top) or orange (bottom) has the opportunity to change one network relationship (dashed arrow). A feedback-seeking relationship may be formed or remain absent for the blue actor, or a feedback-seeking tie may be dissolved or remain to be present for the orange actor. On the right, we see that collaboration scores (in this case 5) may go up, down, or an actor keeps the current score

the other network option (Snijders, 2001, 2005). In other words, the rate function explains the *frequency* of changes are made in the network (i.e., which actor makes a change in either the network relationship or the behavior). The rate function is a single number specifying the number of possible changes each one can make in behavior or the network. Conversely, the objective function determines *which* changes can be made based on the model specification. A model specification within the objective function is based on theory and the related hypotheses, mirroring model specification in more conventional regression analysis, such as logistic or linear regression.

The models are assessed in *R*—a free software system for statistical and graphical computing—using Simulation Investigation for Empirical Network Analysis (*RSiena*) (Ripley et al., 2021). *RSiena* estimates the coevolution of behavior and networks via stochastic actor-oriented models (Snijders et al., 2010). Next to the help function in *R*, potential effects, possibilities, and both in-depth and general information on *RSiena* are available in the free available online manual, written by Ripley et al. (2021). Example *R*-scripts and datasets and more information on the methodology are available on the *RSiena* homepage of Tom Snijders (one of the main developers of *Rsiena*), accessible via the following URLs: https://www.stats.ox.ac.uk/~snijders/siena/ or https://github.com/snlab-nl/rsiena. One can, for instance, find more information concerning the practical side of preparing the dataset, how to run the models in *R*, and other practicalities.

What are the steps a researcher should take when employing a SAOM using *RSiena* can be done via the following four steps (see also Kalish, 2020)?

1. The first step is to prepare the data accordingly to fit the *RSiena* framework. This requires that network data is dichotomized; that is, a feedback nomination is present (1) or not (0). The network data is fitted into an *n* by *n* matrix, where *n* stands for all the students in the network. A network data frame consists of 0's and 1's. A '1' represents a network relation with someone else, whereas a '0' is no network relation. Behavior, or individual characteristics, are included as a common dataset in which rows represent individuals and columns are the variables. Other individual-level data, such as gender, are included as an *RSiena* covariate

For example, we have feedback-seeking network data for waves 1 ($t = 1$) and 2 ($t = 2$). The *t* is a time point or wave. Longitudinal network and behavioral data (attributes) are separately imported in *RSiena* and in such a way that *RSiena* considers them the dependent variable when modeling selection (dependent variable = feedback network) or influence (dependent variable = collaborative intentionality).

2. The second step is to include effects in the network (selection) and behavioral change (influence) model. Luckily, *RSiena* provides modelers with a documentation file specifically applicable to the dataset at hand. Thus, based on the variables included in the previous step, *RSiena* provides a long list of potential effects to include in the selection and influence function. Some effects are commonly included; think of reciprocity, transitivity, and outdegree (Ripley et al., 2021). Other effects are included based on theoretical considerations. For example, one may include homophily effect regarding an attribute (e.g., CI).
3. The third step is estimating the SAOM using a simulation algorithm specified in *R*. We run the SAOM in *RSiena*, which eventually leads to the results in which selection and influence model-based findings are separated in the output.
4. The final step is to interpret the effects. An estimate can be either positive or negative. The interpretation is similar to the interpretation of a logit/log-odds estimate which can be re-calculated as an odds ratio via measuring the exponential function of the SAOM effect (i.e., e^x; with x as the SAOM effect). This means that the estimates can be considered as the likelihood that a connection will be formed or the behavior will be changed. In addition to the estimate of the rate functions (how many changes are made in the network or behavior), an effect in the objective function can be, for example, whether students similar in CI or gender preferentially are more likely to seek feedback from similar others. Such an effect is represented by a positive significant estimate. A negative parameter in the objective function may state, for example, whether reciprocity is unlikely over time in a feedback network or that men are less popular than women in the network. Results from the objective function are usually utilized to test hypotheses. The estimates are divided by the standard error to inspect significance. This is similar to significance testing as in logistic regression. The rate and objective function of both selection and influence operate simultaneously to model coevolution of network or behavior changes respectively (see

Ripley et al., 2021; Snijders, 2001, 2005). We will illustrate the interpretation of the effects in the next sections.

5.3 Illustration of Peer Feedback in Higher Education

We illustrate this method with a longitudinal study conducted in one bachelor's program in higher education among first-year students. We analyze data obtained from 95 first-year sociology students from a large university in the Netherlands. The complete data sample comprises 56 females (64%) and 32 males (36%) with a mean age of 19.5 years old (SD = 1.6). Students answered a 20–30 min computer-based questionnaire across two waves in an academic year (see Brouwer et al., 2018). The current dataset comprises variables on feedback-seeking relations, CI, gender, and personality traits. Wave 1 is often referred to as $t = 1$, and wave 2 is often noted as $t = 2$.

5.3.1 Variables

Peer feedback network. Students could nominate all members in their cohort, i.e., their academic year group, for feedback-seeking in terms of academic help or advice-seeking via a free-recall method. When a respondent started typing, the program automatically provided the respondent with potential names that correspond to the typed text. This eased the network nomination process. Students were allowed to indicate whom they asked for feedback when they do not understand the study material. In other words, students nominated others who they seek for feedback, help, support, or assistance in the academic environment. Students rated per fellow student on a 5-point Likert scale to what extent they agree that they would seek feedback from a certain fellow student (1 = strongly disagree to 5 = strongly agree). To analyze the peer feedback network using *RSiena*, it is necessary to dichotomize feedback nominations. Scores 4 and 5 result in a 1, while other scores resulted in a 0. There are 495 peer feedback nominations at $t = 1$ and 349 at $t = 2$. Using the Hamming statistics (Ripley et al., 2021), we infer 394 changes in feedback nominations between $t = 1$ and $t = 2$. A network generally changes slowly since too much instability and fluctuations pressure the reliability of the *RSiena* analysis (Ripley et al., 2021). The Jaccard index measures changes in tie presence between two waves. A Jaccard index value below 0.30 is deemed unfit for network analysis given too many unstable network relations (Snijders et al., 2010). In this feedback-seeking network, the Jaccard similarity index of 0.36 shows that there is sufficiently high enough stability in peer feedback nominations between both waves. The feedback network is visualized per wave in Fig. 5.3.

Collaboration intentionality (CI). It is difficult to reliably capture collaboration behavior, that is why we asked peers to indicate if they deem others in their year
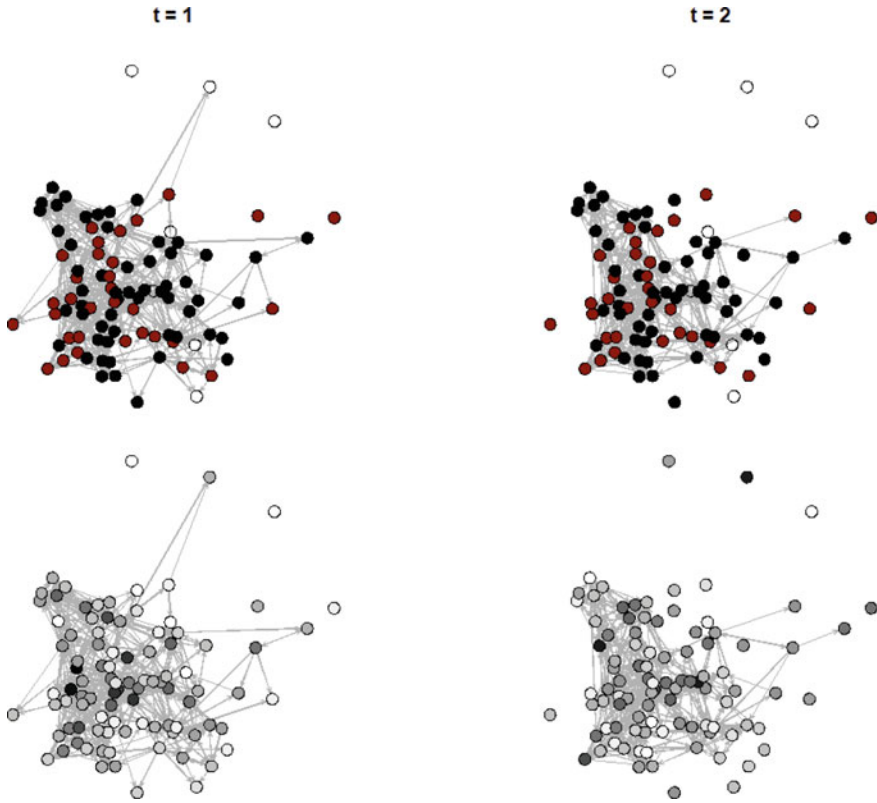
Fig. 5.3 The feedback network is visualized at $t = 1$ and $t = 2$. In the upper row, red nodes are males and black nodes are females (white is missing). The lower two networks show CI as the color of the nodes. The darker the node, the higher one CI score is. Black is score 16, while white is the lowest CI score possible (0)

group collaborative or not. Collaboration intentionality is measured by asking students to nominate others who they deem collaborative. If one is perceived as more collaborative, a student has a higher score. A more collaborative student is then more "popular" as a collaborator. The range of CI is 0–16. A score of 0 represents that a student is never mentioned as a collaborator and a score of 16 means that someone is 16 times nominated. The mean at $t = 1$ is 6.14 (SD = 3.43) and at $t = 2$ it is 5.43 (SD = 3.94). The high standard deviations indicate that there is some variation in CI among students. A combination of the feedback network and CI is presented in the lower row of Fig. 5.3. There is some change in CI scores over time. CI thus captures how collaborative one is via popularity. We assume that a more collaborative student is more popular (i.e., more often nominated as a collaborator).

__Gender__. Our sample comprises males (0) and females (1). Previous research using SAOMs showed that gender plays an important role in friendship network

selection (e.g., Brouwer et al., 2018). A visualization of gender and the feedback network at $t = 1$ and $t = 2$ is provided in the upper row in Fig. 5.3.

Five-Factor Model personality traits. The Five-Factor Model (FFM) measures five personality traits: agreeableness, extraversion, neuroticism, openness, and conscientiousness (McCrae & John, 1992). We relied on the Ten-Item Personality Inventory (Gosling et al., 2003) to assess the five latent traits. The following 10 items are distributed among the students: (1) 'I take time for a talk', (2) 'I try to avoid conflicts', (3) 'I work in a structured manner', (4) 'I am easily enthusiastic', (5) 'I am open to new experiences', (6) 'I ignore adversity quickly', (7) 'I see myself as someone who is generally trusting', (8) 'I can handle stress well', (9) 'I am interested in art', and 10) 'I am self-disciplined'. Students indicated if the statement applies to them on a 5-point Likert scale, ranging from 1 (very inappropriate) to 5 (very appropriate). Extraversion comprises the average of items 1 and 4 ($M = 3.84$, SD $= 0.70$), agreeableness items 2 and 7 ($M = 4.14$, SD $= 0.59$), conscientiousness items 3 and 10 ($M = 3.11$, SD $= 0.95$), neuroticism items 6 and 8 ($M = 3.14$, SD $= 0.78$), and openness to new experiences items 5 and 9 ($M = 3.56$, SD $= 0.77$).

5.3.2 Specifying Effects to Be Included in the SAOM

In *RSiena*, the researcher specifies—similar to more conventional regression analysis—the effects included based on theoretical considerations. We describe each included effect in detail and offer an example graphical interpretation of the included effect. We first describe SAOM effects included in the selection model and then discuss the influence model. Table 5.1 provides an explanation and an visualization of the included effects in the model.

5.3.3 *RSiena* Findings

This statistical method allows us to ask the following research question: To what extent does homophily of CI plays a role in selecting peers for feedback (selection), and to what extent do peer feedback relationships influence CI (i.e., social influence of CI)? However, stochastic actor-oriented models permit researchers to control for other factors that may affect the network-CI link: What is the role of gender and Five-Factor Model personality traits in feedback-seeking selection processes and how do these individual features influence individual changes in CI? The findings of the stochastic actor-oriented selection and influence model are presented in Table 5.2. A positive estimate represents that such a state is pursued ('it is more likely that..'), while a negative parameter indicates that such a state tends to be avoided by students if the opportunity comes to alter a feedback-seeking nomination or changes in CI ('it is less likely that…').

We first start with the selection model presented in Table 5.2 which investigates potential sources of why students seek certain students out for feedback and

Table 5.1 Effects included in the selection and influence SAOMs

Effect ("*RSiena* label")	Description	Simplified visualization
Selection SAOM		
1. Rate ("*rate*")	Rate indicates how many changes students make in their feedback nominations	
2. Outdegree ("*density*")	This effect models the tendency to form feedback relations	
3. Reciprocity ("*recip*")	The tendency towards reciprocal feedback relations	
4. Transitivity ("*transTrip*")	Modeling the tendency to have transitive relations	
5. Interaction reciprocity and transitivity ("*transRecTrip*")	This effect accounts for reciprocity in transitive structures	
6. Friendship ("*X*")	Effect of having a friendship relation (dashed line) on feedback relations	
7. Attribute popularity ("*altX*")	Whether attributes determine feedback popularity (receiving nominations)	
8. Attribute activity ("*egoX*")	Whether attributes determine feedback activity (sending out nominations)	
9. Attribute similarity ("*simX*")	The tendency for similar students to form feedback relations	
Influence SAOM		
10. Rate ("*rate*")	Rate indicates how many changes students make in their CI	
11. Linear shape ("*linear*")	This shape effect captures linear patterns in CI (positive or negative)	
12. Quadratic shape ("*quad*")	Accounting for non-linear distributions of CI	

(continued)

Table 5.1 (continued)

Effect ("*RSiena* label")	Description	Simplified visualization
13. Social influence ("*avSim*")	Adopting a CI similar to the average CI of feedback partners	
14. The effect from attribute ("*effFrom*")	Modeling the effect of an attribute (red) on changes in CI	

Note Attribute refers to an individual feature not related to the network, such as gender, CI, or Five-Factor Model personality traits. Instead of *simX*, we implement *sameX* for categorical variables (gender)

academic support. The dependent variable in the selection model is the feedback seeking network. The rate effect in the rate function shows that students had more than 12 opportunities to alter their feedback-seeking nominations. We are particularly interested in which feature affected feedback-seeking nominations, and we turn to the objective function in the selection model for answers. Students, on the whole, tend to have fewer nominations over time, per the negative outdegree parameter in Table 5.2. We furthermore find that students prefer reciprocated to non-reciprocated relations ('if you seek feedback from me, then I'm more likely to return the favor') and that students are more likely to be embedded in transitive structures ('if I seek feedback from student A and A seeks feedback from student B, then I'm more likely seek feedback from student B'), per the positive and significant reciprocity and transitivity effect in Table 5.2. Yet, the interaction term between reciprocity and transitivity indicates that a reciprocal feedback-seeking relationship is less likely when a student is embedded in a transitive triplet. There are thus multiple social sources for peers to form feedback relations with one another.

Feedback relations are an important source to receive help, support, and feedback from peers. To achieve this, feedback network relations may be utilized to seek others out who most readily can provide qualitative feedback to one another. Notably, we find that students preferentially seek feedback from other students with similar CI scores (estimate $= 0.80$, SE $= 0.36$, $p = 0.027$). As such, it is more likely that students seek feedback from students with similar collaboration tendencies.

Yet, CI is not the only defining feature for feedback-seeking selection; that is, gender, friendships, and personality significantly affect underlying features why some students are more likely to be nominated to seek feedback from than others, which in turn may explain why some are more able to provide feedback and receive support than others. Table 5.2 shows that females are less popular (estimate $= -0.57$, SE $= 0.16$, $p < 0.001$) for feedback-seeking nominations than their male counterparts. Even so, female-female and male-male feedback relations are more likely than cross-gender relations (estimate $= 0.57$, SE $= 0.15$, $p < 0.001$). Thus, similarity in gender is a prerequisite for seeking feedback from one another. Next,

Table 5.2 SAOM findings of *feedback-seeking* selection and influence of *feedback-seekers* on collaboration intentionality (CI), separated by rate and objective function*

Selection model (dep. var. = feedback-seeking nomination)			Influence model (dep. var. = CI)		
Parameter	Est. (SE)	p	Parameter	Est. (SE)	p
Rate function					
Rate effect	12.57 (1.49)	< 0.001	Rate effect	17.87 (4.29)	< 0.001
Objective function					
Feedback-seeking effects			Effects on CI change		
Outdegree (density)	–3.65 (0.19)	< 0.001	Linear shape	–0.19 (0.06)	0.002
Reciprocity	2.48 (0.25)	< 0.001	Quadratic shape	0.03 (0.01)	0.031
Transitivity	0.54 (0.06)	< 0.001	Influence of peers' CI scores on own CI	7.55 (2.50)	0.003
Reciprocity x transitivity	–0.43 (0.09)	< 0.001	Extraversion	–0.15 (0.08)	0.047
Friendship nominations	0.75 (0.16)	< 0.001			
CI similarity	0.80 (0.36)	0.027			
Gender (1 = female)					
Popularity	–0.57 (0.16)	< 0.001			
Similarity	0.57 (0.15)	< 0.001			
Five-Factor Model traits					
Openness popularity	0.31 (0.10)	0.001			
Openness similarity	0.63 (0.30)	0.033			

Note CI = collaboration intentionality; dep. var. = dependent variable; nom. = nomination; Est. = log-odds estimate; SE = standard error; ref. = reference category; Overall maximum convergence ratio = 0.21, which is below the critical value of good model convergence of 0.25 (Ripley et al., 2021)

*We only show marginally significant effects, meaning $p < 0.10$, to keep table as simple and interpretable as possible

having friendship relationships makes it more likely to seek feedback from one another (estimate = 0.75, SE = 0.16, $p < 0.001$). Relatedly, students higher in openness are perceived as more attractive to seeking feedback, and thus are more likely to receive feedback nominations, than students low in openness (estimate = 0.31, SE = 0.10, $p = 0.001$). Being open to new experiences and willing to try new things are considered attractive features for feedback popularity. Moreover,

students similar on openness are more likely to seek feedback from each other than students dissimilar in openness are (estimate = 0.63, SE = 0.30, p = 0.033). These findings suggest that students, who are more willing to embrace new things in higher education, and postulate more readily fresh ideas are also more inclined to select partners for feedback who display similar care for openness.

The influence model, conversely, allows studying whether it is more likely that students become more similar to their feedback partners in CI. Students had in total approximately 18 opportunities to change collaborative intentionality in-between the two waves. We find in the objective function that students tend to have lower scores on CI over time, per the negative linear shape effect (estimate = − 0.19, SE = 0.06, p = 0.002). This effect suggests that there is a linear downward trend in CI. The positive quadratic shape effect stresses that the negative trend is less step for students with higher values on CI (estimate = 0.03, SE = 0.01, p = 0.031).

More importantly, the influence model in Table 5.2 suggests that changes in CI are also driven by social influence (estimate = 7.55, SE = 2.50, p = 0.003). This shows that a student who is nominated to seek feedback from is more likely to adopt a similar value of CI as their peers. Yet, this effect may also exacerbate the problem for non-collaborative students. Namely, students with lower levels of CI tend to have feedback relationships with similar others, and if influence processes are dominant then they may influence each other to take an even lesser collaborative stance. Furthermore, we find that extraversion lowers changes in CI (estimate = − 0.15, SE = 0.08, p = 0.047), meaning that students high on extraversion report lower scores of CI over time.

5.4   Discussion and Outlook

Combining insights from selection and influence, we show that students who are similar in their intention to collaborate are more likely to request each other for feedback. Our network approach elucidates, furthermore, that students are more likely to seek feedback from friends, from students with the same gender, and students who are also open to new experiences. The same-gender effect and similarity in CI is consistent with the homophily principle in selecting peers for feedback (c.f., McPherson et al., 2001).

The novelty of this chapter and the advantage of using stochastic actor-oriented models (SAOMs) is that it allows to unravel social influence from the selection of peers—and vice versa—in feedback-seeking networks. Selection and influence mechanisms are dependent on each other. The major advantage of SAOMs is disentangling influence from selection in a statistically valid way. In our contribution, we show that SAOMs allow us to study the complex interdependence between behavior and network relations. Our methodology builds on an innate feedback loop from selection to influence and influence to selection. In our analysis, we provided a template to analyze and describe selection and influence effects using SAOMs.

Another advantage is that this chapter provides a short introduction to SAOMs but provides, by all means, *not* a full overview of what is possible with SAOMs. If interested, the following references show different applications of SAOMs, providing researchers with more features, possibilities, and information than described here: Snijders (2017), Kalish (2020), Snijders et al. (2010), Steglich et al. (2010), Henneberger et al. (2021), Ripley et al. (2021), Brouwer et al. (2020, 2022), or Veenstra et al. (2013). Here, we illustrated that behavior and networks are two fitting pieces in a puzzle when appropriate statistical methods are utilized. This chapter provides more understanding of the mechanisms underlying peer feedback—utilizing the power that feedback networks provide and SAOMs to monitor selection and influence processes—to advance in higher education.

References

Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research, 73*(3), 277–320. https://doi.org/10.3102/00346543073003277

Atik, G., & Yalçin, Y. (2011). Help-seeking attitudes of university students: The role of personality traits and demographic factors. *South African Journal of Psychology, 41*(3), 328–338. https://doi.org/10.1177/008124631104100307

Baeten, M., Kyndt, E., Struyven, K., & Dochy, F. (2010). Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. *Educational Research Review, 5*, 243–260. https://doi.org/10.1016/j.edurev.2010.06.001

Bandura, A. (1977). *Social learning theory*. Prentice Hall.

Blau, P. M. (1964). *Exchange and power in social life*. Wiley.

Boud, D., Cohen, R., & Sampson, J. (2001). Peer learning and assessment. In D. Boud, R. Cohen, & J. Sampson (Eds.), *Peer learning in higher education: Learning from and with each other* (pp. 67–81). Routledge.

Brouwer, J., Flache, A., Jansen, E. P. W. A., Hofman, W. H. A., & Steglich, C. E. G. (2018). Emergent achievement segregation in freshmen learning community networks. *Higher Education, 76*, 483–500. https://doi.org/10.1007/s10734-017-0221-2

Brouwer, J., Downey, C., & Bokhove, C. (2020). The development of communication networks of pre-service teachers on a school-led and university-led programme of initial teacher education in England. *International Journal of Educational Research, 100*. https://doi.org/10.1016/j.ijer.2020.101542

Brouwer, J., de Matos Fernandes, C .A., Steglich, C. E. G., Jansen, E. P. W. A., Hofman, W. H. A., Flache, A. (2022). The development of peer networks and academic performance in learning communities in higher education. *Learning and Instruction, 80*, 101603. https://doi.org/10.1016/j.learninstruc.2022.101603

Coleman, J. S. (1990). *Foundations of social theory*. Belknap Press of Harvard University Press.

Cook, K. S., & Rise, E. (2003). Social exchange theory. In J. Delamater (Ed.), *Handbook of social psychology* (pp. 53–76). Kluwer Academic/Plenum Publishers.

Dingyloudi, F., & Strijbos, J.-W. (2018). Just plain peers across social networks: Peer-feedback networks nested in personal and academic networks in higher education. *Learning, Culture and Social Interaction, 18*, 86–112.

de Matos Fernandes, C. A., Bakker, D. M., & Dijkstra, J. (2022). Assessing the test-retest reliability of the social value orientation slider measure. *Judgment and Decision Making, 17*(1), 31-49. https://doi.org/10.1017/S1930297500009013

Er, E., Dimitriadis, Y., & Gašević D. (2021). A collaborative learning approach to dialogic peer feedback: a theoretical framework. *Assessment & Evaluation in Higher Education, 46*(4), 586–600. https://doi.org/10.1080/02602938.2020.1786497

Finn, G. M., & Garner, J. (2011). Twelve tips for implementing a successful peer assessment. *Medical Teacher, 33*(6), 443–446. https://doi.org/10.3109/0142159X.2010.546909

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality, 37*(6), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Henneberger, A. K., Mushonga, D. R., & Preston, A. M. (2021). Peer influence and adolescent substance use: A systematic review of dynamic social network research. *Adolescent Research Review, 6*, 57–73. https://doi.org/10.1007/s40894-019-00130-0

Höglinger, M., & Wehrli, S. (2017). Measuring social preferences on amazon mechanical turk. In B. Jann & W. Przepiorka (Eds.), *Social dilemmas, institutions, and the evolution of cooperation* (pp. 527–546). de Gruyter.

Homans, G. C. (1961). *Social behavior and its elementary forms.* Harcourt, Brace and World.

Kalish, Y. (2020). Stochastic actor-oriented models for the co-evolution of networks and behavior: An introduction and tutorial. *Organizational Research Methods, 23*(3), 511–534. https://doi.org/10.1177/1094428118825300

Laydshewsky, R. K. (2013). The role of peers in feedback processes. In Boud, D., & Molloy, E. (2013). *Feedback in higher and professional education: Understanding it and doing it well* (pp. 174–189). Routledge. Retrieved March 12, 2022, from INSERT-MISSING-URL.

Lomi, A., Snijders, T. A. B., Steglich, C. E. G., & Torló, V. J. (2011). Why are some more peers than others? Evidence from a longitudinal study of social networks and individual academic performance. *Social Science Research, 40*(6), 1506–1520. https://doi.org/10.1016/j.ssresearch.2011.06.010

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*, 415–444.

Moolenaar, N., Daly, A. J., Daly, A. J., Cornelissen, F., Liou, Y.-H., Caillier, S., Riordan, R., Wilson, K., & Cohen, N. A. (2014). Linked to innovation: Shaping an innovative climate through network intentionality and educators' social network position. *Journal of Educational Change, 15*(2), 99–123.

Niezink, N. M. D. (2018). Modeling the dynamics of networks and continuous behavior. [Doctoral dissertation, University of Groningen].

Noroozi, O., Banihashem, S, K., Taghizadeh Kerman, N., Parvaneh Akhteh Khaneh, M., Babayi, M., Ashrafi, H., & Biemans, H. J. A. (2022). Gender differences in students' argumentative essay writing, peer review performance and uptake in online learning environments. *Interactive Learning Environments*, 1–16. https://doi.org/10.1080/10494820.2022.2034887

Noroozi, O., Hatami, J., Biemans, H. J. A., van Ginkel, S., & Bayat, A. (2020). Students' online argumentative peer feedback, essay writing, and content learning: Does gender matter? *Interactive Learning Environments, 28*(6), 698–712.

O'Donnell, A. M. (2006). The role of peers and group learning. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 781–802). Lawrence Erlbaum Associates Publishers.

Ripley, R. M., Snijders, T. A. B., Boda, Z., Vörös, A., & Preciado, P. (2021). *Manual for RSiena* (version December 2021). Oxford: University of Oxford, Department of Statistics, Nuffield College; Groningen: University of Groningen, Department of Sociology.

Sangin, M., Molinari, G., Nüssli, M. A., & Dillenbourg, P. (2011). Facilitating peer knowledge modeling: Effects of a knowledge awareness tool on collaborative learning outcomes and processes. *Computers in Human Behavior, 27*(3), 1059–1067. https://doi.org/10.1016/j.chb.2010.05.032

Simões, D., & Pinheiro, M. M. (2014). Collaborative strategies of teaching and learning in higher education. In Rutherford, S. (Ed.), *Collaborative learning: Theory, strategies, and educational benefits.* Nova Science Publishers, Inc.

Snijders, T. A. B. (2017). Stochastic actor-oriented models for network dynamics. *Annual Review of Statistics and Its Application, 4*(1), 343–363. https://doi.org/10.1146/annurev-statistics-060 116-054035

Snijders, T. A. B. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology, 31*, 361–395.

Snijders, T. A. B. (2005). Models for longitudinal data. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 215–247). Cambridge University Press.

Snijders, T. A. B., Van de Bunt, G. G., & Steglich, C. E. G. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks, 32*(1), 44–60. https://doi.org/10.1016/j. socnet.2009.02.004

Spitzmuller, M., & Van Dyne, L. (2013). Proactive and reactive helping: Contrasting the positive consequences of different forms of helping. *Journal of Organizational Behavior, 34*, 560–580. https://doi.org/10.1002/job.1848

Stadtfeld, C., Vörös, A., Elmer, T., Boda, Z., & Raabe, I. (2019). Integration in emerging social networks explains academic failure and success. *Proceedings of the National Academy of Sciences of the United States of America, 116*(3), 792–797. https://doi.org/10.1073/pnas.181138 8115

Steglich, C. E. G., Snijders, T. A. B., & Pearson, M. (2010). Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology, 40*(1), 329–393. https://doi.org/10. 1111/j.1467-9531.2010.01225.x

Thielmann, I., Spadaro, G., & Balliet, D. (2020). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychological Bulletin, 146*(1), 30–90. https://doi.org/10.1037/ bul0000217

van Waes, S., van den Bossche, P., Moolenaar, N. M., Stes, A., & van Petegem, P. (2015). Uncovering changes in university teachers' professional networks during an instructional development program. *Studies in Educational Evaluation, 46*, 11–28.

Veenstra, R., & Steglich, C. (2012). Actor-based model for network and behavior dynamics. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 598–618). Guilford.

Veenstra, D. R., Dijkstra, J. K., Steglich, C. E. G., & Van Zalk, M. H. W. (2013). Network-behavior dynamics. *Journal of Research on Adolescence, 23*(3), 399–412. https://doi.org/10.1111/jora. 12070

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental processes.* Harvard University Press.

Comparing Expert and Peer Assessment of Pedagogical Design in Integrated STEAM Education

6

Kyriaki A. Vakkou, Tasos Hovardas, Nikoletta Xenofontos, and Zacharias C. Zacharia

6.1 Introduction

Peer assessment aims to actively involve peers in employing their knowledge and skills to assess peer work (Cestone et al., 2008; Van Gennip et al., 2010). This may include providing peers with quantitative feedback, for instance, scores across assessment criteria, and/or qualitative feedback, with any justification of scores as well as recommendations for improving peer work (Hovardas et al., 2014; Tsivitanidou et al., 2011). The later would be decisive for letting peer assessees benefit from peer feedback. In education, a quite effective peer assessment format has been the formative/reciprocal one (Tsivitanidou et al., 2011), which engages students in both the roles of peer assessor and peer assessee. Usually this starts with all students undertaking the same set of learning activities to deliver a set of learning products to be assessed. Learning products are any physical or virtual artefacts created by students themselves as they go through a learning activity sequence (Hovardas, 2016; Hovardas et al., 2018). Having created the learning products to be assessed later on, the formative/reciprocal peer assessment procedure should

K. A. Vakkou (✉) · T. Hovardas · N. Xenofontos · Z. C. Zacharia
Research in Science and Technology Education Group, Department of Education, University of Cyprus, 75 Kallipoleos St., PO 20537, 1678 Nicosia, Cyprus
e-mail: vakkou.kyriaki@ucy.ac.cy

121

O. Noroozi and B. de Wever (eds.), *The Power of Peer Learning*,
Social Interaction in Learning and Development,
https://doi.org/10.1007/978-3-031-29411-2_6

be able to familiarize students with the main requirements and characteristics of the work needed to produce the objects of assessment and shape their background knowledge and skills to be able to act as peer assessors. To better support students in their peer assessor role, a training session is often preceding peer assessment (van Zundert et al., 2010; Xiao & Lucking, 2008). In the peer assessee role, peers screen peer feedback and use it constructively to rework and improve their learning products. The formative/reciprocal peer assessment arrangement lets students gain from multiple reflection processes, for example, when peer assessors compare their own learning products with those of their peers, and when peer assessees are about to rework their learning products taking into account peer feedback (Anker-Hansen & Andree, 2019; Hovardas et al., 2014).

Although peer assessment has been practiced quite often with pre-service teachers (Topping, 2021), there are too few studies engaging pre-service teachers in peer assessment for pedagogical design[1] (Fang et al., 2021; Lin, 2018; Ng, 2016; Tsai et al., 2002) and, with the exception of Tsai et al. (2002), who reported that peer assessment was not valid across all dimensions studied, no previous study reported either on the validity or the reliability of peer assessment for pedagogical design. What is more, peer assessment has not been yet implemented in pedagogical design for integrated Science, Technology, Engineering, Arts, and Mathematics (STEAM)[2] education (Margot & Kettler, 2019; Thibaut et al., 2018). Integrated STEAM education is understood as the inclusion of at least two STEAM subjects in designing learning activity sequences, whole lesson plans or even projects, with a concentration on real-world problems (Tasiopoulou et al., 2020). Peer assessment would be especially valuable in this case, where teacher collaboration for pedagogical design is indispensable (Margot & Kettler, 2019). STEAM integration seems to be quite demanding and challenging for primary and secondary teachers (Brown & Bogiages, 2019), despite the fact that STEAM education should already presuppose some interdisciplinarity. The silo approach, which compartmentalizes each STEM discipline within its own confines, is still prevailing in many curricula and in everyday school practice in most educational systems, presenting substantial barriers for promoting integrated STEAM education (Kelly & Knowles, 2016). To address these barriers, pre-service teachers need to be familiarized with good practice in pedagogical design in integrated STEAM education and to work with

---

[1] Pedagogical design begins with planning learning activities, which includes class arrangement (i.e., if activities will be performed by individual students, groups of students or the entire class), the description of learning products, and time needed for students to undertake each activity. Pedagogical design also involves the orchestration of separate activities into sequences of activities, lesson plans or projects. Pedagogical design should align with curriculum standards (e.g., learning objectives, assessment), while it depends on the pedagogical theories and instructional strategies to be chosen (see de Jong et al., 2021).

[2] STEM has been extended to also involve "Arts" (STEAM) and highlight the innovation and creativity of the concept; the "A" in STEAM is interpreted by some scholars as "All", which wishes to denote the inclusiveness of the approach (Iacovou, 2021). We will refer to "STEM" whenever we present findings of previous research, which also referred to "STEM".

their peers to design learning activity sequences, lesson plans and projects based on STEAM integration (Hovardas et al., 2020; Tasiopoulou et al., 2020). Using peer assessment for that purpose would allow pre-service teachers develop the competences and mindset needed to provide insightful feedback to peers as well as gain from such input. If peer assessment proves valid and reliable in the context of pedagogical design for integrated STEAM education, and if peer feedback can include constructive input, for instance, justifications for quantitative scores given by peers and suggestions for improving peer work, then it may be instrumental for pre-service teacher training.

Another considerable challenge for pedagogical design in integrated STEAM education, which could be tackled by peer assessment, is female engagement (Zacharia et al., 2020). Previous research has shown that female students in primary education do not differ from their male peers in their attitudes towards STEM (McGuire et al., 2020; Zhou et al., 2019). Moreover, girls in primary education tend to receive higher STEM grades than boys (O'Dea et al., 2018) and tasks related to ICT literacy (Siddiq & Scherer, 2019). It is quite interesting that career beliefs of female students in STEM do not correspond at all to their attitudes and ability in primary education (Sadler et al., 2012; Selimbegović et al., 2019). Indeed, girls do not expect to be as successful as boys in STEM-related careers, which results eventually in fewer girls than boys being interested in pursuing a STEM career at the beginning of high school. This mismatch between female attitudes and performance in STEM, on the one side, and female STEM career beliefs, on the other, is a distinguishing feature in the transition from primary to secondary education and marks female field-specific ability beliefs (Wang & Degol, 2017). What we confront here is a type of "bottleneck effect", where the overall decrease of students interested in following STEM careers is accompanied by a sharp decrease in the gender diversity of students who still remain interested. This bottleneck effect may be held responsible for any further reduction in female enrolment in STEM subjects and degrees in higher education (Zacharia et al., 2020). It would be crucial to examine if the implementation of peer assessment in pedagogical design for integrated STEAM education could offer input and insight for addressing female engagement. Specifically, qualitative feedback provided by peers can include justification of scores (quantitative part of feedback) and suggestions for improving pedagogical design in this direction. Female engagement will be one of the design dimensions on which we will focus in present study.

Our objective was to implement peer assessment for pedagogical design in integrated STEAM education and to compare expert and peer feedback, in this regard. To our knowledge, this is the first study to investigate if peer assessment can be employed for improving pedagogical design in integrated STEAM education. We engaged pre-service teachers registered in an undergraduate programme for primary education in a formative/reciprocal peer assessment arrangement, where they had the chance to act as both peer assessors and peer assessees. Participants delivered a short but comprehensive pedagogical scenario concentrating on educational robotics, where they had to refer to at least two STEAM subjects, describe

a real-world problem to be solved by primary students through thinking critically and creatively, include problem-solving activities for educational robotics, and engage girls as much as boys. Following a training session, participants acted as peer assessors providing quantitative feedback (scores) and qualitative feedback (justification of their scores; suggestions for improving pedagogical scenarios) to their peers. An expert also assessed each pedagogical scenario, and, based on these scores, we awarded badges to a number of participants for recognition of excellence in developing pedagogical scenarios. Moreover, we awarded assessment badges to pre-service teachers based on deviations of peer assessor scores from expert assessor scores for the same pedagogical scenario.

We first investigated if pre-service teachers were able to respond to expert assessment by improving their pedagogical design (Research question 1). This would provide a solid indication of understanding assessment criteria, grasping the dimensions of pedagogical design and working productively to improve pedagogical scenarios along these dimensions. Then, we examined if peer assessment was valid and reliable (Research question 2). If it was, then it could be exploited in pre-service teacher training for pedagogical design in integrated STEAM education. Our next objective was to compare between expert and peer feedback and outline the weaknesses of peer feedback, if any, for instance, where peer feedback was inferior to expert feedback (Research question 3). This would give us the opportunity to target such weaknesses in training sessions for peer assessment. Finally, we investigated the main determinants that led groups of peer assessees to choose a pedagogical scenario that they would then fully develop into a lesson plan. Here we aimed to explore if performance badges would feature out as significant determinants (Research question 4), which would imply that pre-service teacher training may benefit from exploiting performance badges and letting pre-service teachers use them in their social media and networks.

6.2   Methods

6.2.1   Participants

Participants were pre-service teachers (5 males and 20 females) who registered as undergraduate students in the compulsory course "Science Teaching Methods" offered in the fourth semester of the undergraduate programme for primary education at the Department of Education, University of Cyprus. The course content involved a strong component on integrated STEAM education. Participation in the study was part of an assignment given to students, which counted, upon completion of all related activities, towards 10% of their final mark in the course. Student performance in the assignment did not influence their final grade but, to receive the 10%, they had to submit all deliverables related to the assignment on time. Although all 29 students enrolled in the course agreed to take part in the study, only 25 managed to conclude all tasks and be included in the sample. All

participants were guaranteed anonymity. They were informed that their deliverables would be used within the frame of the current study and they provided their informed consent for using them as data sources. Participants were notified that they were free to withdraw at any time from the study, if they felt inclined to do so, without providing any further explanation and without their withdrawal having any impact on the allocation of the 10% of their grade. No participant had any prior experience in peer assessment.

6.2.2 Procedure

Overview
All participants followed an introductory session to the study and a training session on peer assessment, where the first and second authors acted as instructors (see Fig. 6.1 for a presentation of the whole procedure). Participants then developed pedagogical scenarios for integrated STEAM education concentrating on educational robotics. Each scenario was assessed twice by an expert and once by a peer (the second round of expert assessment was accompanied by peer assessment as well). The first round of expert assessment was planned to check if pre-service teachers would respond to expert feedback and improve their scenarios. This would also provide some additional guidance to pre-service teachers in terms of good practice in pedagogical design, concentrating on the first version of the scenarios they delivered. The second round of expert assessment was used to estimate the validity and reliability of peer assessment and investigate differences between expert and peer feedback. Based on expert scores for pedagogical scenarios in the second round, and overlap of peer scores with expert scores, two types of performance badges were granted to a selection of participants, namely, a scenario badge and an assessment badge. Participants then were randomly assigned to groups and they had to choose one scenario to fully develop into a lesson plan among the ones that group members had already delivered for assessment. The focus here was on whether performance badges were decisive for scenario selection.

Introductory session
In the introductory session, the aim and scope of the study was presented, specifications of participation were discussed and the participants granted their informed consent for the use of the data sources, which will be presented in the next section. Participants were informed that they would take part in a procedure of developing pedagogical scenarios for integrated STEAM education, which would involve two rounds of expert assessment and one round of peer assessment. Each participant
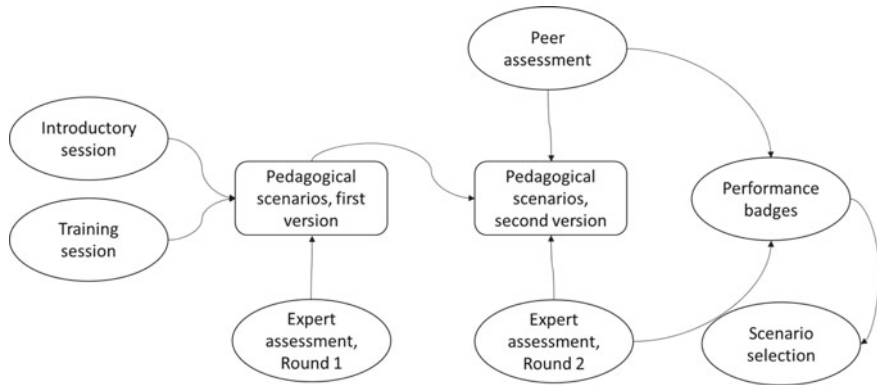
Fig. 6.1 Procedure: After an introduction session to the study and a training session for peer assessment, pre-service teachers developed a first version of pedagogical scenarios in integrated STEAM education with a focus on educational robotics. A first round of expert assessment followed and pre-service teachers reworked their scenarios. This second version of pedagogical scenarios were subjected to a second round of expert assessment and peer assessment. Based on expert and peer scores, a selection of pre-service teachers were awarded performance badges (scenario badge; assessment badge). Pre-service teachers then formed groups and selected one scenario among the ones already delivered by group members to fully develop into a lesson plan

would deliver one pedagogical scenario using the GINOBOT for designing learning activities for primary students (https://www.engino.com/w/index.php/products/innolabs-robotics/ginobot). The introductory session included a component of educational robotics focusing on the GINOBOT, the basic functionalities and capabilities of the robot, the KEIRO software for programming the GINOBOT (https://enginoeducation.com/downloads/), and prototype lesson plans concentrating on the GINOBOT. Pedagogical scenarios were meant to be comprehensive descriptions of pedagogical design that should meet four requirements: First, scenarios should address at least two STEAM subjects, which was used as an approach to operationalize integrated STEAM education. Each scenario should describe a real-world problem to be solved by primary students using the GINOBOT in problem-solving activities through thinking critically and creatively. Apart from these requirements, pedagogical scenarios should also seek to engage girls as much as boys to address the gender gap in STEAM education. The introductory session included examples of good practice in pedagogical design for all these dimensions.

Another objective of the introductory session was to familiarize participants with Open Badges, specifically, Open Badge Factory (https://openbadgefactory.com/en/), which is used by competent organizations to create, issue and manage Open Badges, and Open Badge Passport (https://openbadgepassport.com/), where badge owners can obtain and store a pdf certificate of their badge and share it with other users in their social media accounts. Open Badges can be issued for recognizing either intention (e.g., intention to enter a community of practice, intention

to communicate a message) or performance (e.g., knowledge, achievements, competences, skills, abilities). They have the form of a digital artefact with malleable visual identity and they carry relevant metadata. Open Badges can be employed in social media accounts to increase visibility of intention or performance of badge owners and shape interaction with other social media users accordingly. All participants created an account in Open Badge Passport and this infrastructure was employed for issuing participant badges for recognizing excelling performance in pedagogical design and peer assessment.

Training session
The training session on peer assessment focused on formative/reciprocal peer assessment. It started with all participants creating an account in HumHub (https://www.humhub.com/en), which was used by the expert assessor and peer assessors to rate pedagogical scenarios and submit expert and peer feedback to peer assessees. Participants rated two different ready-made scenarios provided by the instructors using four assessment criteria, which followed closely the good practice requirements given to participants for developing scenarios:

   Criterion 1: The scenario refers explicitly to the STEAM subjects involved;

   Criterion 2: The scenario describes a real-world problem to be solved through thinking critically and creatively;

   Criterion 3: The scenario includes problem-solving activities with the GINOBOT robot; Criterion 4: The scenario seeks to engage girls as much as boys.

   After rating the first ready-made scenario, participants discussed with instructors their scores and justifications for these scores. A comparison with expert scores followed and an elaboration upon deviations between expert and participant scores concluded that part of the training session. Then, participants rated the second ready-made scenario, and in this case, they were requested to provide justifications for their scores as well as suggestions for changes to improve the scenario. Another round of discussion followed, which involved all above aspects.

Delivery of scenarios, expert assessment, peer assessment, performance badges and student group formation
Each participant delivered one pedagogical scenario, which was first rated by an expert assessor (Senior Research Associate at the University of Cyprus holding a PhD in Science Education and having participated in five European research projects in STEAM education during the last decade). The expert assessor used the same assessment criteria which participants had used in the training session. The expert assessor also provided qualitative feedback to participants with justification of scores across criteria and changes proposed for improving pedagogical scenarios. Reworked scenarios were again assessed by the expert assessor in a second expert assessment round, as well as by participants themselves who acted as peer assessors. Each participant used the same four assessment criteria they had used in the training session to rate two peer pedagogical scenarios chosen randomly and provide qualitative feedback to peer assessees with justification of scores for each assessment criterion and suggestions for changes. The identity of

all assessors and asseesees was known to all participants. Excelling performance in pedagogical design (i.e., scenarios with the three highest total expert assessor scores, which belonged to 7 participants) as well as excelling performance in peer assessment (i.e., the three lower ranked deviations of total peer scores from the expert assessor for the same pedagogical scenario, which belonged to 10 participants) were recognized by being awarded specific badges (scenario badge and assessment badge, respectively). The identity of all pre-service teachers who received badges was known to all peers. Participants then were randomly assigned to groups and selected one pedagogical scenario from those that group members had already submitted for assessment, to further develop it into a lesson plan in integrated STEAM education.

6.2.3 Data Sources and Coding

Pedagogical scenarios, quantitative scores for each assessment criterion and qualitative feedback (justification of scores; suggestions for changes for improving pedagogical scenarios) provided by the expert assessor and peer assessors in HumHub were the data sources for the study. We coded expert and peer qualitative feedback for items justifying scores and changes proposed for improving pedagogical scenarios. An additional coding process focused on how different STEAM disciplines were integrated in pedagogical scenarios. The first and third author acted as independent coders for 10% of all data. Inter-rater reliability amounted to over 85% and the rest of the cases were resolved after a discussion between coders.

6.2.4 Statistical Analyses

We employed non-parametric statistics for all data we collected, since data distributions were non-normal. Specifically, we used Wilcoxon Signed Ranks Tests to ascertain whether size (word count) of pedagogical scenarios provided by participants, total expert assessor scores and expert scores for each assessment criterion differed significantly between the first and second round of expert assessment. These analyses would reflect if participants were able to respond to expert assessment and improve their pedagogical scenarios. To estimate the validity of peer assessment, we computed Spearman's *rho* correlation coefficients for total scores between expert and peer assessors as well as for scores given for each assessment criterion between expert and peer assessors. Another set of Spearman's *rho* correlation coefficients were computed for total scores and scores for each assessment criterion between the two different peer assessors who were assigned the same pedagogical scenario. This second correlational analysis concentrated on the reliability of peer assessment. Differences in the characteristics of expert and peer feedback, including size (word count) of feedback, scores for each assessment criterion, number of items justifying scores, and number of changes proposed to peer

assessees, were examined by means of Mann–Whitney Tests. Finally, we employed tree modeling to investigate if performance badges would be significant determinants for the selection of pedagogical scenarios by peer assessees for developing a lesson plan in integrated STEAM education. In this analysis, we used as independent variables the following parameters: Participants' gender, whether they had been granted a scenario badge and/or an assessment badge, total expert assessor score in the first and second round of expert assessment, the difference in total scores between the first and second round of expert assessment, total peer assessor scores, and the absolute value of the difference in total scores between peer assessors.

6.3 Results

6.3.1 Pre-service Teacher Responsiveness to Expert Assessment

Average word count of pre-service teachers' pedagogical scenarios increased from the first to the second round of expert assessment from 107.28 to 160.12 words (Wilcoxon Signed Ranks Test $Z = -3.58$, $p < 0.001$). This increase in the size of scenarios was accompanied by an analogous increase in the average value of the total score of the expert assessor from 5.88 (min = 4, max = 9; standard deviation = 1.33), in the first round, to 7.32 (min = 4, max = 10; standard deviation = 1.60) in the second round (Wilcoxon Signed Ranks Test $Z = -3.67$, $p < 0.001$). These results suggest that pre-service teachers, overall, responded to the suggestions of the expert assessor and were able to enrich the descriptions of their pedagogical scenarios and improve their scores. Examining each assessment criterion separately (Table 6.1), there was significant improvement of scenarios in three out of four criteria (Criterion 1: The scenario refers explicitly to the STEAM subjects involved, Wilcoxon Signed Ranks Test $Z = -2.71$, $p < 0.01$; Criterion 2: The scenario describes a real-world problem to be solved through thinking critically and creatively, Wilcoxon Signed Ranks Test $Z = -2.89$, $p < 0.01$; Criterion 4: The scenario seeks to engage girls as much as boys, Wilcoxon Signed Ranks Test $Z = -2.83$, $p < 0.01$). For problem-solving activities with the GINOBOT robot (Criterion 3), improvement was not significant. In this case, there was probably a ceiling effect, with the average expert score being already quite high in the first round of expert assessment. We need to highlight the rather low average scores for Criterion 4 ("The scenario seeks to engage girls as much as boys"). Despite the improvement that was recorded, most scenarios failed to effectively address female engagement after the first round of expert assessment.

6.3.2 Validity and Reliability of Peer Assessment

Spearman's *rho* correlation coefficients between total expert scores and total peer scores (global measure of the validity check for peer assessment) as well as

Table 6.1 Mean scores for pedagogical scenarios for each assessment criterion in the two rounds of expert assessment

	First round	Second round	Wilcoxon Signed Ranks Test Z
Criterion 1: The scenario refers explicitly to the STEAM subjects involved	1.68 (0.63)	2.04 (0.54)	-2.71^{**}
Criterion 2: The scenario describes a real-world problem to be solved through thinking critically and creatively	1.28 (0.61)	1.84 (0.80)	-2.89^{**}
Criterion 3: The scenario includes problem-solving activities with the GINOBOT robot	1.84 (0.80)	2.04 (0.68)	-1.51^{ns}
Criterion 4: The scenario seeks to engage girls as much as boys	1.08 (0.28)	1.40 (0.50)	-2.83^{**}

Note Each criterion was scored by the expert assessor along a three-point Likert-scale (1 = not addressed at all; 2 = partially addressed; 3 = fully addressed); standard deviations are given in parentheses; ns = non-significant; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$

between total scores provided by different peer assessors (global measure of the reliability check for peer assessment) revealed that, overall, peer assessment was valid (Spearman's *rho* correlation coefficient = 0.48, $p < 0.001$; $N = 50$) and reliable (Spearman's *rho* correlation coefficient = 0.70, $p < 0.001$; $N = 25$). Spearman's *rho* correlation coefficients for the validity and reliability check for each criterion separately are shown in Table 6.2. Peer assessment proved to be valid in three out of four assessment criteria (Criterion 2: The scenario describes a real-world problem to be solved through thinking critically and creatively, Spearman's *rho* correlation coefficient = 0.47, $p < 0.001$; $N = 50$; Criterion 3: The scenario includes problem-solving activities with the GINOBOT robot, Spearman's *rho* correlation coefficient = 0.42, $p < 0.01$; $N = 50$; Criterion 4: The scenario seeks to engage girls as much as boys, Spearman's *rho* correlation coefficient = 0.39, $p < 0.01$; $N = 50$). Reliability revealed somewhat worse results, with two out of the four assessment criteria having significant coefficients (Criterion 2: The scenario describes a real-world problem to be solved through thinking critically and creatively; Spearman's *rho* correlation coefficient = 0.61, $p < 0.01$, $N = 25$; Criterion 3: The scenario includes problem-solving activities with the GINOBOT robot; Spearman's *rho* correlation coefficient = 0.87, $p < 0.001$, $N = 25$). All the above findings indicate that peer assessment did not succeed in providing valid and reliable quantitative feedback across all assessment criteria, despite the training session that pre-service teachers had attended.

Table 6.2 Spearman's rho correlation coefficients for the validity and reliability check of peer assessment for each assessment criterion

	Validity check (between expert assessor and peer assessors; $N = 50$)	Reliability check (between first and second peer assessor, $N = 25$)
Criterion 1: The scenario refers explicitly to the STEAM subjects involved	0.14^{ns}	0.33^{ns}
Criterion 2: The scenario describes a real-world problem to be solved through thinking critically and creatively	0.47^{***}	0.61^{**}
Criterion 3: The scenario includes problem-solving activities with the GINOBOT robot	0.42^{**}	0.87^{***}
Criterion 4: The scenario seeks to engage girls as much as boys	0.39^{**}	0.45^{ns}

Note ns = non-significant; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$

6.3.3 Comparison Between Expert and Peer Feedback

Average scores for each assessment criterion in expert and peer feedback are presented in Table 6.3. All scores in peer feedback were higher than expert assessor scores and in three out of four criteria these differences were found to be significant (Criterion 1: The scenario refers explicitly to the STEAM subjects involved, Mann–Whitney $Z = -2.84$, $p < 0.01$; Criterion 2: The scenario describes a real-world problem to be solved through thinking critically and creatively, Mann–Whitney $Z = -2.90$, $p < 0.01$; Criterion 4: The scenario seeks to engage girls as much as boys, Mann–Whitney $Z = -4.79$, $p < 0.001$). The fact that there was no significant difference for Criterion 3 (The scenario includes problem-solving activities with the GINOBOT robot) should be linked to the ceiling effect that was underlined for this criterion in the section on "Pre-service teacher responsiveness to expert assessment" above (see also Table 6.1, in this regard). Overall, the consistently higher average scores of peers as compared to expert scores may indicate some type of positive bias towards peers.

Differences in average scores (quantitative feedback) combined with difference in feedback size (word count) can help us trace and interpret further differences in the qualitative elements of expert and peer feedback, i.e., items provided for justification of scores and changes proposed to peer assessees for improving their pedagogical scenarios. The size of expert feedback (average word count = 168 words; standard deviation = 27 words) was significantly larger compared to the size of peer feedback (average word count = 91 words; standard deviation = 24 words) (Mann–Whitney $Z = -6.73$, $p < 0.001$). At the same time, the average number of items justifying scores (Table 6.4) as well as the average number of

Table 6.3 Average scores for each assessment criterion in expert and peer feedback

	Average scores in expert feedback	Average scores in peer feedback	Mann–Whitney Z
Criterion 1: The scenario refers explicitly to the STEAM subjects involved	2.04 (0.54)	2.46 (0.65)	−2.84[**]
Criterion 2: The scenario describes a real-world problem to be solved through thinking critically and creatively	1.84 (0.80)	2.42 (0.74)	−2.90[**]
Criterion 3: The scenario includes problem-solving activities with the GINOBOT robot	2.04 (0.68)	2.46 (0.74)	−2.53 [ns]
Criterion 4: The scenario seeks to engage girls as much as boys	1.40 (0.50)	2.40 (0.79)	−4.79[***]

Each criterion was scored by the expert assessor and peer assessors along a three-point Likert-scale (1 = not addressed at all; 2 = partially addressed; 3 = fully addressed); standard deviations are given in parentheses; ns = non-significant; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

changes proposed to peer assessees (Table 6.5) were, for all assessment criteria, higher in expert feedback as compared to peer feedback. Although peer assessors were able to provide at least one item for justifying their quantitative scores for each assessment criterion, changes proposed to peer assessees were too few, with no change included in peer feedback for Criterion 3 ("The scenario includes problem-solving activities with the GINOBOT robot"). Taken together, the above findings imply that lower average scores across all assessment criteria in expert feedback were accompanied by more items to justify scores and more changes proposed to peer assessees, which led to a relatively increased word count of expert feedback.

Specifically, the average number of items justifying scores was significantly higher in expert feedback for Criterion 1 ("The scenario refers explicitly to the STEAM subjects involved") (Table 6.4; Mann–Whitney $Z = −3.34$, $p < 0.001$), while the average number of changes proposed to peer assesses was significantly higher in expert feedback for Criteria 3 ("The scenario includes problem-solving activities with the GINOBOT robot") (Table 6.5; Mann–Whitney $Z = −4.12$, $p < 0.001$) and 4 ("The scenario seeks to engage girls as much as boys") (Table 6.5; Mann–Whitney $Z = −3.27$, $p < 0.001$). Another interesting finding was that word count in peer feedback tended to increase when peer assessors proposed changes to peer assessees related to female engagement (Criterion 4) (Spearman' *rho* correlation coefficient = 0.37, p < 0.01). We computed a crosstabulation and ran

Table 6.4 Average number of items justifying scores in expert and peer feedback

	Average number of items justifying scores in expert feedback	Average number of items justifying scores in peer feedback	Mann–Whitney Z
Criterion 1: The scenario refers explicitly to the STEAM subjects involved	1.28 (0.46)	1.02 (0.14)	-3.34[***]
Criterion 2: The scenario describes a real-world problem to be solved through thinking critically and creatively	1.20 (0.50)	1.13 (0.33)	-0.47[ns]
Criterion 3: The scenario includes problem-solving activities with the GINOBOT robot	1.16 (0.37)	1.06 (0.24)	-1.33[ns]
Criterion 4: The scenario seeks to engage girls as much as boys	1.20 (0.50)	1.04 (0.20)	-1.76[ns]

Note Standard deviations are given in parentheses; ns = non-significant; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$

a relevant Chi-Square analysis to examine if participants' gender influenced the probability of proposing any changes to peer assessees for improving their scenarios in the criterion for female engagement (Criterion 4). We found that proposing changes for female engagement was neither associated with peer assessor gender nor with peer assessee gender.

6.3.4 Selection of Pedagogical Scenarios by Peer Assessees for Developing a Lesson Plan in Integrated STEAM Education

After receiving expert and peer feedback, peer assessees worked in groups to select one pedagogical scenario among those that group members had already delivered for assessment and process it further to develop a lesson plan in integrated STEAM education. There were three groups with three pre-service teachers and another four groups with four. We employed tree modeling to investigate the effect of several parameters on this selection, including pre-service teachers' gender, whether they had been granted a scenario badge and/or an assessment badge, total expert assessor score in the first and second round of expert assessment, the difference

Table 6.5 Average number of changes proposed in expert and peer feedback for improving pedagogical scenarios

	Average number of changes proposed in expert feedback	Average number of changes proposed in peer feedback	Mann–Whitney Z
Criterion 1: The scenario refers explicitly to the STEAM subjects involved	0.16 (0.47)	0.08 (0.28)	-0.55 [ns]
Criterion 2: The scenario describes a real-world problem to be solved through thinking critically and creatively	0.48 (0.59)	0.19 (0.39)	-2.34 [ns]
Criterion 3: The scenario includes problem-solving activities with the GINOBOT robot	0.48 (0.77)	0.00 (0.00)	-4.12***
Criterion 4: The scenario seeks to engage girls as much as boys	0.44 (0.51)	0.10 (0.31)	-3.27***

Note Standard deviations are given in parentheses; ns = non-significant; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$

in total scores between the first and second round of expert assessment, total peer assessor scores, and the absolute value of the difference in total scores between peer assessors.

Figure 6.2 presents the tree computed. At each split, the significant determinants of scenario selection are shown with the values which partitioned the sample at each branch (i.e., there is a left and a right branch in each split). The result of partitioning is depicted at nodes, where one can see the number of scenarios, which were selected or not (n), and the percentage of that number in the total sample. Partitioning is terminated at end nodes. Reading the tree from the top downwards, the first determinant in the first split is whether scenarios had been delivered by pre-service teachers who had been granted a scenario badge. If scenarios belonged to pre-service teachers who had not received such a badge, then these were most probably not selected for developing a lesson plan (first split, left branch, Node 1). Among scenarios delivered by pre-service teachers with a scenario badge (first split, right branch, Node 2), those selected were the ones with a clear improvement measured as difference in total expert assessor scores between the first and second round of expert assessment.
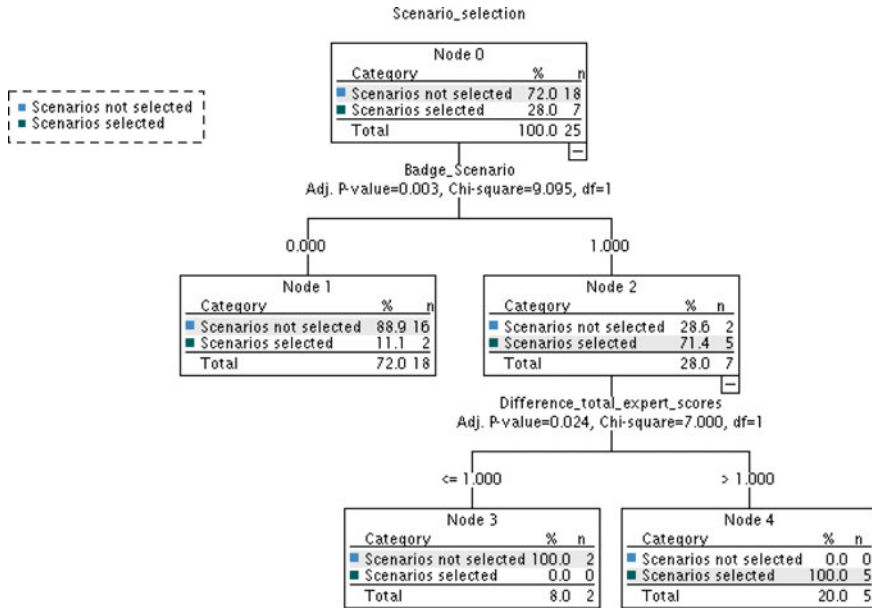
Fig. 6.2 Tree model for selection of pedagogical scenarios by peer assessees to develop a lesson plan in integrated STEAM education. Significant determinants are shown at each split with thresholds for partitioning the sample at left and right branches. Each node depicts the number of scenarios selected or not (n) and their percentage in the total sample. Overall percentage of cases correctly classified = 92.0%

6.4   Discussion

The significant correlations computed as global measures of validity (correlations between total scores of expert and peer assessors) and reliability (correlations between total scores of different peer assessors for the same pedagogical scenario) indicate that peer assessment can be employed in the case of pedagogical design of pre-service teachers in integrated STEAM education. Another strength of peer assessment in our study was that peer assessors were able to include in their feedback to peers at least one item for justifying their quantitative scores in each assessment criterion. The above findings corroborate the few studies available on peer assessment for pedagogical design, according to which, formative peer assessment can improve pedagogical design delivered by pre-service teachers (Fang et al., 2021; Lin, 2018; Ng, 2016; Tsai et al., 2002). There were, however, assessment criteria for which requirements for either validity (STEAM integration) or reliability (STEAM integration; female engagement) were not met. In the case of STEAM integration, there was also a significant difference in items for justifying scores between experts and peers, with the later presenting a lower average. It seems that peer assessors would need much more support and guidance in the training sessions preceding the enactment of peer assessment in order to secure

the validity and reliability of their quantitative scores for STEAM integration. This should refer to a concrete anchoring of STEAM disciplines in current curricula as well as a thorough exemplification of possible synergies between STEAM disciplines within the frame of educational robotics involving, for instance, engineering design, programming, and mathematics. Another concern for pre-service teacher training for peer assessment should concentrate on the use of mathematics in integrated STEAM education. As we have seen from an additional qualitative analysis of the pedagogical designs delivered by participants in our study, mathematics were embedded in their designs as simple mathematical operations and not as comprehensive mathematical thinking processes. Analogous weaknesses have been reported in recent research in integrated STEAM education for primary school teachers (Roehrig et al., 2021).

With regard to female engagement, it was quite interesting that reliability for this assessment criterion was not satisfactory despite the fact that a substantial majority of participants were women. This may imply that there was considerable heterogeneity among female participants in approaches on how to engage female students as well as in judging the effectiveness of these approaches. Female engagement seems to have been the criterion where participants confronted the most challenges in pedagogical design. This criterion had the lowest average expert score in both rounds of expert assessment, and presented the lowest score among criteria for peer assessors as well. Given these shortcomings of pedagogical design for female engagement, and given that there are urgent calls for addressing the gender gap in STEAM (Zacharia et al., 2020), much more attention should be paid for engaging girls as much as boys in pedagogical design for integrated STEAM education. Although several options have been suggested for initiating and sustaining girls' interest in STEAM, such as spatial tools (Moè et al., 2018) and role models (Barabino et al., 2020), not all of them are readily compatible with educational robotics. What is more, the selection of robotic kits for constructing artefacts, which will be the organizing principles of pedagogical design, seems to be quite crucial. A major concern here is that the motive structures, according to which female students operate, do not always overlap with male motivation, especially with regard to speed, power and competition (Johnson, 2003). Although there do not seem to be differences in learning outcomes between boys and girls in educational robotics (Zhong & Xia, 2020), girls may be more committed to follow teacher instructions (Lindh & Holgersson, 2007; Shih et al., 2012), but for that to happen, girls should first be adequately motivated and engaged. More research will be needed in this direction to support female engagement in integrated STEAM education through pedagogical design.

Average scores for each assessment criterion provided by peers were higher than expert scores. Peer over-scoring is common in peer assessment in higher education (Lu & Chiu, 2021; Panadero et al., 2013). It may be enhanced in the case of female peers, who were the majority in our sample, and who may receive higher scores than male peers, not due to gender bias, but because female peers may be assumed to perform better than males (Baker, 2008; Falchikov & Magin, 1997; May & Gueldenzoph, 2006; Tucker, 2014). This positive bias needs

to be addressed in future training sessions, especially when implementing peer assessment in pedagogical design for integrated STEAM education, since it would detract from the opportunities for improvement, which peer assessment may introduce. Indeed, this was reflected in our study in the difference between expert and peer feedback in the number of changes suggested to peers for improving pedagogical scenarios. An option to address over-scoring may be anonymity of peer assessors and assessees, although it has not always delivered the expected outcomes (Yu & Sung, 2016). For pre-service teachers in primary education, the option of anonymity would probably not contribute to tackling the positive bias since females outnumber their male peers by a wide margin. Anonymity may result in more critical feedback including changes recommended to peers (Howard, 2010; Lin, 2018), but it may severely compromise genuine and constructive peer interaction (Rotsaert et al., 2018). Indeed, it has been found that peer collaboration, when combined with peer assessment, yielded better outcomes as compared to peer assessment alone (Fang et al., 2021). Moreover, training was found to counteract the negative effects of non-anonymous peer assessment (Li, 2017). An option could be to plan a transition from anonymous to non-anonymous peer assessment, which was reported to lead through iterations to equal feedback quality with anonymous peer assessment (Rotsaert et al., 2018). Furthermore, since the concentration on the implementation of specific assessment criteria has not been enough in our study, pre-service teacher training for peer assessment in pedagogical design needs to incorporate a stronger component of the interrelationship between the peer assessor and peer assessee role, e.g., what is expected from peer assessors and what is needed by peer assessees in peer feedback to improve their designs. Reflective focus group discussions among peers may foster this exchange.

The selection process by peer assessees after receiving peer feedback, where they collectively decided which pedagogical scenario to single out and fully develop into a lesson plan, was determined by recognition of excellence in pedagogical design (scenario badge) and improvement in pedagogical design between the two rounds of expert assessment. On the one hand, this finding would imply that pre-service teacher training may benefit from exploiting performance badges and letting pre-service teachers use these badges in their social media and networks. On the other hand, we need to highlight that no aspect of peer assessment was included among the determinants of the tree model, which may imply that peer scores and feedback may not be as valued as much as expert scores and feedback. Previous research showed that pre-service teachers, despite being familiarized through peer discussion and elaboration with peer assessment formats and assessment criteria, may be still dependent upon expert (teacher) advice for the use of assessment criteria (Ng, 2016) or they may still prefer instructor feedback over peer feedback (Seroussi et al., 2019). Such an attitude may have been exacerbated by the female majority of our sample, since female prospective teachers have been found to be more reluctant to give and receive peer feedback than their male peers (Evans & Waring, 2011; Peled et al., 2014). Overall, pre-service teachers may remain ambivalent as to how peer feedback could improve their pedagogical design as long as they lack confidence in their peers' abilities to act as

competent assessors. Future research should focus on the potential contribution of peer assessment for empowering pre-service teachers in pedagogical design for integrated STEAM education. Consolidating pre-service teachers' peer assessment skills would support teacher collaboration in formal and informal teacher networks and communities of practice as well as promote distributed leadership.

References

Anker-Hansen, J., & Andrée, M. (2019). Using and rejecting peer feedback in the science classroom: A study of students' negotiations on how to use peer feedback when designing experiments. *Research in Science & Technological Education, 37*, 346–365. https://doi.org/10.1080/02635143.2018.1557628

Baker, D. F. (2008). Peer assessment in small groups: A comparison of methods. *Journal of Management Education, 32*, 183–209. https://doi.org/10.1177/1052562907310489

Barabino, G., Frize, M., Ibrahim, F., Kaldoudi, E., Lhotska, L., Marcu, L., Stoeva, M., Tsapaki, V., & Bezak, E. (2020). Solutions to gender balance in STEM fields through support, training, education and mentoring: Report of the international women in medical physics and biomedical engineering task group. *Science and Engineering Ethics, 26*, 275–292. https://doi.org/10.1007/s11948-019-00097-0

Brown, R. E., & Bogiages, C. A. (2019). Professional development through STEM integration: How early career math and science teachers respond to experiencing integrated stem tasks. *International Journal of Science and Mathematics Education, 17*, 111–128. https://doi.org/10.1007/s10763-017-9863-x

Cestone, C. M., Levine, R. E., & Lane, D. R. (2008). Peer assessment and evaluation in team-based learning. *New Directions for Teaching and Learning, 116*, 69–78. https://doi.org/10.1002/tl.334

de Jong, T., Gillet, D., Rodríguez-Triana, M. J., Hovardas, T., Dikke, D., Doran, R., Dziabenko, O., Koslowsky, J., Korventausta, M., Law, E., Pedaste, M., Tasiopoulou, E., Vidal, G., & Zacharia, Z. C. (2021). Understanding teacher design practices for digital inquiry–based science learning: The case of Go-Lab. *Educational Technology Research & Development, 69*, 417–444. https://doi.org/10.1007/s11423-020-09904-z

Evans, C., & Waring, M. (2011). Student teacher assessment feedback preferences: The influence of cognitive styles and gender. *Learning and Individual Differences, 21*, 271–280. https://doi.org/10.1016/j.lindif.2010.11.011

Falchikov, N., & Magin, D. (1997). Detecting gender bias in peer marking of students' group process work. *Assessment & Evaluation in Higher Education, 22*, 385–396. https://doi.org/10.1080/0260293970220403

Fang, J.-W., Chang, S.-C., Hwang, G.-J., & Yang, G. (2021). An online collaborative peer-assessment approach to strengthening pre-service teachers' digital content development competence and higher-order thinking tendency. *Educational Technology Research and Development, 69*, 1155–1181. https://doi.org/10.1007/s11423-021-09990-7

Hovardas, T. (2016). A learning progression should address regression: Insights from developing non-linear reasoning in ecology. *Journal of Research in Science Teaching, 53*, 1447–1470. https://doi.org/10.1002/tea.21330

Hovardas, T., Tsivitanidou, O., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers & Education, 71*, 133–152. https://doi.org/10.1016/j.compedu.2013.09.019

Hovardas, T., Pedaste, M., Zacharia, Z., & de Jong, T. (2018). Model-based science inquiry in computer-supported learning environments: The case of Go-Lab. In Azad, A. K. M., Auer, M., Edwards, A., & de Jong, T. (Eds.), *Cyber-physical laboratories in engineering and science education* (pp. 241–268). Springer. https://doi.org/10.1007/978-3-319-76935-6_10

Hovardas, T., Xenofontos, N., Irakleous, M., Pavlou, Y., Kouti, G., & Zacharia, Z. C. (2020). Specifications for learning scenarios and instructional approaches. Deliverable D3.4. GINOBOT project, Research Promotion Foundation Proposal Number INNOVATE/0719/0098.

Howard, C. D., Barrett, A. F., & Frick, T. W. (2010). Anonymity to promote peer feedback: Pre-service teachers' comments in asynchronous computer-mediated communication. *Journal of Educational Computing Research, 43*, 89–112. https://doi.org/10.2190/EC.43.1.f

Iacovou, K. (2021). Literature review on integrated STEAM education. Master Thesis, University of Cyprus, Department of Education.

Johnson, J. (2003). Children, robotics, and education. *Artificial Life and Robotics, 7*, 16–21. https://doi.org/10.1007/BF02480880

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education, 3*, 11. https://doi.org/10.1186/s40594-016-0046-z

Li, L. (2017). The role of anonymity in peer assessment. *Assessment & Evaluation in Higher Education, 42*, 645–656. https://doi.org/10.1080/02602938.2016.1174766

Lin, G.-Y. (2018). Anonymous versus identified peer assessment via a Facebook-based learning application: Effects on quality of peer feedback, perceived learning, perceived fairness, and attitude toward the system. *Computers & Education, 116*, 81–92. https://doi.org/10.1016/j.compedu.2017.08.010

Lindh, J., & Holgersson, T. (2007). Does lego training stimulate pupils' ability to solve logical problems? *Computers & Education, 49*, 1097–1111. https://doi.org/10.1016/j.compedu.2005.12.008

Lu, M., & Chiu, M. M. (2021). Do teamwork guidelines improve peer assessment accuracy or attitudes during collaborative learning? *IEEE Transactions on Education.* https://doi.org/10.1109/TE.2021.3130242

Margot, K. C., & Kettler, T. (2019). Teachers' perception of STEM integration and education: A systematic literature review. *International Journal of STEM Education, 6*, 2. https://doi.org/10.1186/s40594-018-0151-2

May, G. L., & Gueldenzoph, L. E. (2006). The effect of social style on peer evaluation ratings in project teams. *Journal of Business Communication, 43*, 4–20. https://doi.org/10.1177/0021943605282368

McGuire, L., Mulvey, K. L., Goff, E., Irvin, M. J., Winterbottom, M., Fields, G. E., Hartstone-Rose, A., & Rutland, A. (2020). STEM gender stereotypes from early childhood through adolescence at informal science centers. *Journal of Applied Developmental Psychology, 67*, 101109. https://doi.org/10.1016/j.appdev.2020.101109

Moè, A., Jansen, P., & Pietsch, S. (2018). Childhood preference for spatial toys. Gender differences and relationships with mental rotation in STEM and non-STEM students. *Learning and Individual Differences, 68*, 108–115. https://doi.org/10.1016/j.lindif.2018.10.003

Ng, E. M. V. (2016). Fostering pre-service teachers' self-regulated learning through self- and peer assessment of wiki projects. *Computers & Education, 98*, 180–191. https://doi.org/10.1016/j.compedu.2016.03.015

O'Dea, R. E., Lagisz, M., Jennions, M. D., & Nakagawa, S. (2018). Gender differences in individual variation in academic grades fail to fit expected patterns for STEM. *Nature Communications, 9*, 1–8. https://doi.org/10.1038/s41467-018-06292-0

Panadero, E., Romero, M., & Strijbos, J. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation, 39*, 195–203. https://doi.org/10.1016/j.stueduc.2013.10.005

Peled, Y., Bar-Shalom, O., & Sharon, R. (2014). Characterisation of pre-service teachers' attitude to feedback in a wiki-environment framework. *Interactive Learning Environments, 22*, 578–593. https://doi.org/10.1080/10494820.2012.731002

Roehrig, G., Dare, E., Ring-Whalen, E., & Wieselmann, J. (2021). Understanding coherence and integration in integrated STEM curriculum. *International Journal of STEM Education, 8*, 2. https://doi.org/10.1186/s40594-020-00259-8

Rotsaert, T., Panadero, E., & Schellens, T. (2018). Anonymity as an instructional scaffold in peer assessment: Its effects on peer feedback quality and evolution in students' perceptions about peer assessment skills. *European Journal of Psychology of Education, 33*, 75–99.

Sadler, P. M., Sonnert, G., Hazari, Z., & Tai, R. (2012). Stability and volatility of STEM career interest in high school: A gender study. *Science Education, 96*, 411–427. https://doi.org/10.1002/sce.21007

Selimbegović, L., Karabegović, M., Blažev, M., & Burušić, J. (2019). The independent contributions of gender stereotypes and gender identification in predicting primary school pupils' expectancies of success in STEM fields. *Psychology in the Schools, 56*, 1614–1632. https://doi.org/10.1002/pits.22296

Seroussi, D.-E., Sharon, R., Peled, Y., & Yaffe, Y. (2019). Reflections on peer feedback in disciplinary courses as a tool in pre-service teacher training. *Cambridge Journal of Education, 49*, 655–671. https://doi.org/10.1080/0305764X.2019.1581134

Shih, B. Y., Chang, C. J., Chen, Y. H., Chen, C. Y., & Liang, Y. D. (2012). LEGO NXT information on test dimensionality using Kolb's innovative learning cycle. *Natural Hazards, 64*, 1527–1548. https://doi.org/10.1007/s11069-012-0318-y

Siddiq, F., & Scherer, R. (2019). Is there a gender gap? A meta-analysis of the gender differences in students' ICT literacy. *Educational Research Review, 27*, 205–217. https://doi.org/10.1016/j.edurev.2019.03.007

Tasiopoulou, E., Myrtsioti, E., Niewint Gori, J., Xenofontos, N., Hovardas, T., Cinganotto, L., Anichini, G., Garista, P., & Gras-Velazquez, A. (2020). STE(A)M IT—An interdisciplinary STEM approach. Integrated STEM teaching State of Play. European Schoolnet, Brussels. http://steamit.eun.org/files/D2.1_STEAM_IT_State_of_play_final.pdf

Thibaut, L., Knipprath, H., Dehaene, W., & Depaepe, F. (2018). How school context and personal factors relate to teachers' attitudes toward teaching integrated STEM. *International Journal of Technology and Design Education, 28*, 631–651. https://doi.org/10.1007/s10798-017-9416-1

Topping, K. J. (2021). Digital peer assessment in school teacher education and development: A systematic review. *Research Papers in Education.* https://doi.org/10.1080/02671522.2021.1961301

Tsai, C.-C., Lin, S. S. J., & Yuan, S.-M. (2002). Developing science activities through a networked peer assessment system. *Computers & Education, 38*, 241–252. https://doi.org/10.1016/S0360-1315(01)00069-0

Tsivitanidou, O., Zacharia, Z. C., & Hovardas, T. (2011). Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction, 21*, 506–519. https://doi.org/10.1016/j.learninstruc.2010.08.002

Tucker, R. (2014). Sex does not matter: Gender bias and gender differences in peer assessments of contributions to group work. *Assessment & Evaluation in Higher Education, 39*, 293–309. https://doi.org/10.1080/02602938.2013.830282

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction, 20*, 280–290. https://doi.org/10.1016/j.learninstruc.2009.08.010

van Zundert, M., Sluijsmans, D. M. A., & Van Merriënboer, J. J. G. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20*, 270–279. https://doi.org/10.1016/j.learninstruc.2009.08.004

Wang, M. T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review, 29*, 119–140. https://doi.org/10.1007/s10648-015-9355-x

Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *Internet and Higher Education, 11*, 186–193. https://doi.org/10.1016/j.iheduc.2008.06.005

Yu, F.-Y., & Sung, S. (2016). A mixed methods approach to the assessor's targeting behavior during online peer assessment: Effects of anonymity and underlying reasons. *Interactive Learning Environments, 24*, 1674–1691. https://doi.org/10.1080/10494820.2015.104140

Zacharia, Z. C., Hovardas, T., Xenofontos, N., Pavlou, I., & Irakleous, M. (2020). Education and employment of women in science, technology and the digital economy, including AI and its influence on gender equality. Policy Department for Citizens' Rights and Constitutional Affairs, European Parliament (Report prepared at the request of the FEMM Committee, Policy Department for Citizens' Rights and Constitutional Affairs, Directorate-General for Internal Policies). https://www.europarl.europa.eu/RegData/etudes/STUD/2020/651042/IPOL_STU(2020)651042_EN.pdf

Zhong, B., & Xia, L. (2020). A systematic review on exploring the potential of educational robotics in mathematics education. *International Journal of Science and Mathematics Education, 18*, 79–101. https://doi.org/10.1007/s10763-018-09939-y

Zhou, S. N., Zeng, H., Xu, S. R., Chen, L. C., & Xiao, H. (2019). Exploring changes in primary students' attitudes towards science, technology, engineering and mathematics (STEM) across genders and grade levels. *Journal of Baltic Science Education*, *18*, 466. https://doi.org/10.33225/jbse/19.18.466

Part III

Technological Contributions on Peer Learning

Constructing Computer-Mediated Feedback in Virtual Reality for Improving Peer Learning: A Synthesis of the Literature in Presentation Research

7

Stan van Ginkel and Bo Sichterman

7.1 Introduction

Within presentation research, presenting is frequently defined as *"a combination of knowledge, skills, and attitudes needed to speak in public in order to inform, self-express, to relate and to persuade"* (De Grez, 2009, p. 5). Following this definition, an important notion is the interrelatedness of the cognitive, behavioural and affective domains considering the concept of oral presentation competence, since students' public speaking performance can be enhanced or inhibited by any or all of these competencies (Van Ginkel et al., 2015). Further, this competence is regarded as crucial for working in varying professional environments, career success and effective participation in the democratic society (e.g. De Grez et al., 2009; Van Ginkel et al., 2015; Van Konsky & Oliver, 2012). Therefore, teaching this competence is considered as a crucial objective in higher education (Van Ginkel et al., 2015).

Although the provision of curricula towards developing presentation skills remains crucial in higher education, several challenges appear for curriculum designers and teachers. First of all, developing presentation competence is widely regarded as a time-consuming activity (Van Ginkel et al., 2015). This perspective does not correspond to the current trend in education in which student numbers rise, while possibilities for teacher-student interactions diminish. Consequently, there is a pressure on curricula to integrate both effective as well as efficient

S. van Ginkel (✉) · B. Sichterman
Research Group Digital Ethics, University of Applied Sciences Utrecht, Padualaan 99, 3508 AB Utrecht, The Netherlands
e-mail: stan.vanginkel@hu.nl

evidence-based approaches, including instructions, learning activities and feed-back strategies, for teaching oral presentation competence. Second, this challenge is even strengthened given the fact that students should also develop several other academic, communication and domain-specific competencies in limited time frames during their educational lives, which even further increases the pressure on presentation curricula (Pittenger et al., 2004).

One of the crucial educational design principles for effective learning environments fostering students' oral presentation competence is peer learning (Van Ginkel et al., 2015). Although previous studies addressed both the effectivity of peer feedback for encouraging public speaking performances in higher education as well as the efficiency by adopting peers in formative assessment processes, teachers outperformed peers in terms of the impact on students' development of presentation competence. Follow-up studies experimented with VR technologies as an alternative feedback source in presentation courses and revealed significant effects on student learning comparable to teacher feedback. Recent developments regarding this innovative technology could potentially support peer and self-learning, since the VR systems can nowadays produce feedback messages on non-verbal communication aspects such as eye contact and use of voice that directly relate to the standards of high-quality feedback (Van Ginkel et al., 2019). However, it remains unclear how such messages should be formulated, how feedback messages are perceived by students in higher education and to what extent these messages, produced in VR systems, can be considered as effective for peer and self-learning.

This chapter synthesizes previous studies in presentation research with the aim to construct a research agenda on computer-mediated feedback in VR for peer learning fostering students' oral presentation competence. In the first three sections, an overview is given of research focusing on the role, the effectiveness and the quality of peer feedback in presentation education. Subsequently, the following two sections discuss the potentials of VR, AI and computer-mediated feedback in such educational trajectories. Finally, a research agenda has been constructed focusing on computer-mediated feedback in VR and AI for improving peer learning in presentation education.

7.2   The Role of Peer Feedback in Presentation Research

A systematic review on the development of learning environments fostering oral presentation competence in higher education revealed that, besides principles relating to instruction, presentation tasks, behaviour modeling and the opportunity to practice, three out of the seven crucial educational design principles address the essence of feedback. Moreover, peer feedback can be considered as one of these seven crucial educational design principles (Van Ginkel et al., 2015). In specific, based on empirical research and arguments grounded in theory, it is concluded that feedback should be explicit, contextual, adequately timed and of suitable intensity in order to improve students' oral presentation competence (Mitchell & Bakewell,

1995). Moreover, it has been highlighted that involving peers in formative assessment processes supports students' development in presentation competence and attitudes towards presenting.

Feedback provided by peers is frequently positioned within the process of formative assessment (e.g. Baker & Thompson, 2004; Carroll, 2006; Hattie & Timperley, 2007; Noroozi & Hatami, 2019; Noroozi & Mulder, 2017; Shaw, 2001). According to Falchikov (2005), formative assessment is intended to monitor and improve student learning through providing students with feedback. Regarding publications in presentation research, several scholars claim the need to triangulate multiple feedback sources, such as teachers, peers and the self, for guaranteeing that reflective learning takes place (e.g. Carroll, 2006). Additionally, others emphasize that the adoption of peers encourages a higher sense of feedback sensitivity (e.g. Econopouly et al., 2010), increases active learning (Shaw, 2001) and collaborative learning (Kolber, 2011). Another argument for peer feedback within formative assessment relates to the point that assessing other students' presentations helps students to be more aware of the presentation criteria which encourages their own public speaking performances (De Grez et al., 2012). Finally, the perceived responsibility by peers in giving and receiving feedback enhances their willingness to speak in public which as a consequence impacts their presentation performances (Mitchell & Bakewell, 1995).

Moving from conceptual arguments for peer feedback to empirical evidence for the effectiveness of this feedback source for developing presentation competence, several researchers claim the impact of feedback from peers on students' speaking skills (e.g. Chang & Warren, 2005). However, only a few studies based their claims on experimental study designs (Van Ginkel et al., 2015). One example of such an experimental study demonstrated the superiority of peer feedback when combined with tutor feedback over a condition with solely tutor feedback (Mitchell & Bakewell, 1995). Nevertheless, it remains questionable what the impact of the peer as feedback source actually was, since the quantity of the feedback was not taken into consideration. Further, empirical results showed a fragmented picture regarding the impact of peer feedback on students' attitudes towards feedback. Although some studies address positive perceptions of students towards peer evaluations, other studies highlight that certain students do not prefer peer feedback if they feel incompetent considering the predefined assessment criteria for presenting (e.g. Cheng & Warren, 2005). This is an important reason why peers should be trained in providing and receiving feedback by making use of feedback instruments, such as rubrics, prior to formative assessment processes in classroom settings.

Concluding, conceptual arguments embedded in theory, encompassing reflective, active and collaborative learning, support the involvement of peers in feedback processes in presentation education (Hattie & Timperley, 2007; Van Ginkel et al., 2015). Empirical evidence is found in peer learning studies for increasing students' oral presentation competence and students' attitudes towards presenting (De Grez et al., 2009). However, high quality evidence for the effectiveness of peer feedback in presentation research and conditions under which this feedback source is successful demonstrated ambiguous results. Therefore, more empirical and, more

importantly, experimental study designs are needed to verify the effectiveness of peer feedback and the quality of peer feedback in presentation contexts. The following chapter focuses on the potential differential effectiveness of peer feedback and other commonly used feedback sources in higher education and their impact on students' oral presentation competence.

7.3 The Differential Effectiveness of Feedback Sources

Over the last decades, the impact of peer feedback on students' development of competence has received much attention in higher education research (see Latifi et al., 2020, 2021; Noroozi et al., 2012, 2018; Taghizadeh et al., 2022). These studies tended to focus solely on peer feedback or on the combination of peer feedback with other feedback sources, such as the teacher or the self. To illustrate, research has demonstrated that students' knowledge about psychological concepts was improved when peer feedback was involved in the learning process (Kelly et al., 2010). Additionally, peer feedback improved the language skills and transferable skills of students (Tsaushu et al., 2012). Moreover, in regard to the combination of peer feedback with other feedback sources, studies demonstrated a positive effect on the development of scientific writing skills (Clarke et al., 2013).

While studies revealed a positive impact of peer feedback, as an individual feedback source or combined with other sources (such as the teacher), on the development of students' cognition, skills and attitudes, it has been reported that different feedback sources, such as the peer or the teacher, potentially have a differential impact on learning (Hattie & Timperley, 2007). Moreover, empirical findings addressing this potential differential effect were lacking. Therefore, Van Ginkel et al. (2017a) aimed to investigate the impact of different feedback sources, that is the teacher, the peer, the peer guided by a tutor and the self, on the development of students' oral presentation competence. In this study, a pre-test post-test quasi-experimental design was adopted and students' presentation performances, in terms of cognition, behaviour and attitude towards presenting, were assessed using multiple-choice tests and a rubric. Results of this study showed a substantial overall progression in each of these components of students' oral presentation competence. Interestingly, with respect to presentation behaviour, the impact of teacher feedback was significantly higher than the instructional conditions that involved the peer or the self. Moreover, the effect of self-assessment on students' progression of presentation behaviour and attitude towards presenting was smaller compared to the other feedback sources.

The findings of the experimental study highlight the superiority of feedback provided by the teacher over peer feedback and peer feedback guided by a tutor. This, therefore, supports the idea of the differential impact of these different feedback sources on students' learning. Results of this study are in line with literature that emphasizes the essence of the teacher, and their function as a role-model, for students' learning within the context of higher education (Van Haaren & Van der

Rijst, 2014). Moreover, it has been stated in research focusing on constructing educational design principles for peer feedback that the teacher fulfills an essential role as designer and facilitator within the peer feedback process (Van den Berg et al., 2006). Taken this together, although various studies revealed a positive impact of peer feedback on students' development of competence, it is recommended to optimize the feedback of this source to make it as effective for learning as teacher approaches. However, this requires in-depth knowledge about underlying feedback processes, including the quality of feedback and differences in quality between the teacher and the peer.

7.4   Quality Criteria for Developing Effective Feedback Messages

Although the experimental field study focused on the impact of peer feedback in comparison to other commonly used feedback sources, such as the teacher, and students' presentation performance, insights into the underlying feedback processes remain unclear. As such, it is questionable to what extent the quality of feedback differs between the teachers and peers. Regarding the gaps in the feedback and presentation literature, more knowledge is needed on how teachers, peers and peers guided by tutors deliver their feedback. Additionally, more research needs to be carried out to determine the aspects of feedback they focus on and how feedback processes relate to theoretical and empirical insights considering feedback quality criteria (Boud & Molloy, 2013; Price et al., 2010). Therefore, a follow-up study focused on analyzing the feedback processes, since these are considered as essential in student learning (Asghar, 2010; Falchikov, 2005), and may influence students' oral presentation performance. Specifically, the empirical study examined the feedback processes initiated directly after five minute pitches of 95 undergraduate students in realistic university presentation courses.

In order to analyze the feedback processes of teachers and peers, a coding scheme was composed that included crucial feedback quality criteria based on the literature. To illustrate, the earlier studies addressed both content as well as form-related characteristics of feedback that influence students' learning and performance. To start with, feedback should be specifically related to pre-defined assessment criteria (Moreno, 2004). In the context of presentation skills development, the content of the presentation, the structure of the presentation, the interaction with the audience and the presentation delivery (i.e. use of voice, eye contact and posture and gestures) should be included in the feedback. Moreover, feedback should also include content-related arguments that directly relate to the assessment criteria (Topping, 1998). Further, the following three criteria relate directly to the directions of feedback that are emphasized by Hattie and Timperley (2007). Feedback should incorporate information about students' actual performance, the ideal or desired level of performance and opportunities to bridge the gap between the actual and desired performance. Besides content-related characteristics, form-related criteria are especially essential in the delivery of feedback

messages from the teacher or peer to the individual student. In line with this, feed-back should be delivered in manageable units in order to prevent cognitive overload (Mayer & Moreno, 2002). Subsequently, these messages should be formulated in a positive and constructive manner to increase the likelihood that students will uptake their feedback and to persist in learning (Kluger & DeNisi, 1996).

The analyses revealed that on all seven quality feedback criteria significant dif-ferences existed between the teacher, peers and peers guided by tutor (Van Ginkel et al., 2017b). The teacher scored higher than peers on all quality criteria of feed-back and the teacher performed better than peers guided by tutor on six out of the seven quality criteria. Further, peers guided by tutor scored higher than peer feedback only on the content-related criteria. Reflecting these results with the pre-vious experimental study on the feedback source, it can be concluded the feedback quality could be argued as the essential explanation for the earlier identified differ-ences in impact between the teacher and the peer in presentation education. Both feedback quality as well as teachers as experts are highly emphasized as valuable in formative assessment processes in the literature (e.g. Shute, 2008).

Taking a closer look at the gathered results of this empirical study, it should be noted that also significant differences exist between peers and peers guided by tutors purely related to the content-characteristics of feedback. This might be caused by the fact that a tutor (a student-assistant) was present to guide the feedback processes by questioning and intervening. However, it remains remark-able that the previous experimental study did not reveal any significant difference between the peers and peers guided by tutor conditions regarding their impact on students' oral presentation performances. This might be explained by the cru-cial role of form-related characteristics, such as the stepwise manner in which the feedback is presented and formulated, as being conditional for delivering a mes-sage effectively. Although other factors, for example the authority of the feedback provider, are not taken into consideration in this study, the quality of the feed-back can be considered as crucial for student learning in presentation education. However, peers should be explicitly trained before entering feedback processes in classrooms. And, as addressed in this chapter, innovative technologies might also be valuable in feedback processes. Regarding the delivery of computer-mediated feedback messages in the presentation context, both content- as well as form-related should be critically be incorporated in the construction and composition processes of these messages.

7.5 Virtual Reality as an Alternative Feedback Source for Peer Learning

Previous experimental studies revealed that peer feedback, when adopted as an individual feedback source, had a limited impact on students' development of presentation competence. Moreover, a lack of quality in peer feedback has been established. Subsequently, it has been recommended that students should be edu-cated in providing peer feedback. Additionally, the triangulation of feedback

sources was suggested to be potentially effective in enhancing reflective learning. Concerning the latter, it remained questionable whether innovative technologies, such as VR, might be a valuable contribution in peer feedback processes by delivering computer-mediated feedback aiming to foster students' presentation skills. Recent studies in closely related fields revealed the potentials of integrating peer learning in VR-based technologies (e.g. Chang et al., 2020; Chien et al., 2020). In this study, we specifically focus on the field of presentation research.

As addressed in several domains, such as the medical, engineering, leisure and flight industry sectors, virtual learning environments are increasingly being adopted for practicing delicate surgeries for medical students, educating engineering students in spatial thinking skills, providing images of destinations for travelers and training pilots for real-life flying tasks (e.g. Coller & Scott, 2009; Hawkins, 1995; Merchant et al., 2014; Van Ginkel et al., 2019). However, it remained unclear whether learning environments adopting VR-based technologies can also be applied for developing academic and communication skills. These systems are potentially relevant, since they are able to imitate real-life situations and could deliver computer-mediated feedback from the VR system to the user (e.g. Boetje & Van Ginkel, 2021; LaViola et al., 2017; Van Ginkel et al., 2019).

Seeing the potentials of the VR technology, an experimental field study was conducted to examine to what extent there are significant differences in students' presentation development between a VR and a traditional face-to-face condition. Additionally, this study intended to learn from perceptions of students regarding working with such an innovative tool as a potential replacement for a face-to-face presentation rehearsal in terms of practicing and receiving feedback (Van Ginkel et al., 2019). Therefore, in a realistic university presentation skill course, students were randomly assigned to one of the following conditions. In the first condition, students had to present a five-minute pitch to a VR audience and received quantitative feedback on eye contact, use of voice and posture and gestures traced by the VR system and explained by an expert. In the second condition, students had to present face-to-face and received feedback from a presentation teacher.

Within this experiment, comparable instruments were adopted for measuring students' presentation skills, knowledge and attitude towards presenting as in an earlier described study in this chapter. Results showed that students' developed these components of oral presentation competence significantly from pre-test to post-test without a difference between the VR and face-to-face condition. Further, the self-evaluation tests revealed that students in both conditions highly appreciated the feedback they received. However, the arguments they provided differ between the two groups. Students in the traditional setting who received feedback from the presentation trainer addressed the value of its feedback because of the positive and constructive comments, while students who presented in VR appreciated the—by experts—interpreted quantitative computer-mediated feedback regarding the detailed and analytical characteristics. More specifically, students who pitched in VR emphasized they never received such a detailed feedback on their skills

in previous educational programs. Moreover, the objective character of the feedback, as perceived by students, was also highlighted as a valuable component for developing their presentation skills in a VR environment (Van Ginkel et al., 2019).

The lack of difference in impact between the conditions on developing students' presentation competence might be explained by the opinions of students with regard to their rehearsal and feedback experiences in this experiment. Although arguments for their perceptions differed between students in the VR and face-to-face conditions, no differences in scores were found for two crucial educational design principles fostering presentation skills relating to both practicing as well as receiving feedback. The findings of this study, therefore, suggest that the incorporation of a VR-based presentation task in presentation education including computer-mediated feedback is effective for students' development of presentation competence. However, based on this experiment, VR is not necessarily more efficient, since experts had to be involved in order to translate the quantitative feedback reports provided by the VR system to the students, and it remained questionable to what extent this alternative feedback source could contribute in peer learning. On the other hand, following technological developments, VR technologies also facilitate the delivery of immediate feedback during presentation performances in which the presence of an expert is not required. Moreover, even computer-mediated feedback, delivered after students' presentations, is on the agenda of raped transitions in educational technology (Van Ginkel et al., 2020).

7.6   Two Recent VR Experiments: Students' Perceptions on Computer-Mediated Feedback

In order to verify to what extent VR feedback could be valuable for peer learning, two additional VR field experiments were conducted focusing on (1) the effects of immediate feedback in VR on presentation skills development (Van Ginkel et al., 2020) and (2) the perceptions of students regarding the value of qualitative computer-mediated delayed feedback in a VR presentation environment (Sichterman et al., 2021). The first study focused explicitly on the role of immediate feedback, since VR offers the opportunity to deliver feedback directly during presentations of students on aspects such as eye contact and use of voice. The second study explored the value of qualitative computer-mediated delayed feedback messages following students' perceptions, since this factor can be considered as a crucial intermediate variable for encouraging or inhibiting students' presentation competence development (Van Ginkel et al., 2015). Based on these insights, follow-up studies should be formulated focusing on the role of VR feedback for peer learning, which will be used to construct a future research agenda on peer and learning in the field of presentation education.

Regarding the first field experiment, the effects of immediate computer-mediated feedback in VR were tested by comparing the impact of immediate feedback on students' presentation development with a control group of delayed expert-mediated feedback in a realistic presentation course setting. The target

aspects were eye contact and speech pace, since these components of non-verbal communication are frequently selected by students for formulating personal learning goals in secondary and higher education presentation curricula. Immediate feedback for eye contact was provided by making use of time icons, provided by the VR system, that appeared if the eye contact of the speaker began to linger. For example, if the presenter focused for more than five seconds on their slides, the icon, projected in VR, turned red, advising the student to re-focus their eye contact and to re-engage their audience members. For speech pace, a comparable icon was used to inform the speaker to slow down if their speech rate exceeded 160 words per minute. These timings are based on the validation of a presentation rubric in the scientific literature (Van Ginkel et al., 2017c). The results of the experiment revealed no difference in impact between the immediate feedback and expert feedback condition on presentation performance. Further, students characterized the VR environment as an effective and motivating platform for practicing presentation skills. Findings from this study facilitate the expansion of opportunities for students to use immediate feedback as an alternative form of feedback, for example in peer feedback, for their presentation skills development. Moreover, adopting such a type of feedback in education, without making use of experts, could result in less pressure on resources, including time and staffing (Van Ginkel et al., 2020).

Besides insights considering the value of immediate feedback in VR for students' learning, recent technological and pedagogical developments allow for composing qualitative delayed feedback messages based on the earlier used quantitative feedback reports produced by the VR system in presentation education (see Van Ginkel et al., 2019). The conversion of quantitative feedback, which had to be interpreted by an expert, to qualitative feedback messages might suggest that there is no expert intervention needed anymore and that students could interpret the feedback messages individually or with their peers. Consequently, a preliminary study, in which 27 university students were involved, explored the perceived value of automated, qualitative feedback messages in a VR-system for developing students' presentation skills development (Sichterman et al., 2021). In this experimental study, students' perceptions on the qualitative automated feedback messages (i.e. the experimental condition) were compared with a situation in which quantitative feedback reports were produced by the VR system and interpreted by an expert (i.e. the control condition). The formulation of the feedback messages in the experimental condition was constructed by adopting (1) the seven feedback quality criteria as earlier explained in this chapter (Van Ginkel et al., 2017b) and (2) two crucial presentation criteria for non-verbal behaviour, relating to eye contact and use of voice, as emphasized in a previously validated rubric oral presentation skills (Van Ginkel et al., 2017c).

Considering students' perceptions of feedback within this VR experiment, the following groups of items were selected: (1) aspects regarding the value of feedback (such as the perceived relevance of feedback, sensitivity of feedback and quality of the feedback messages) and (2) aspects regarding students' development of presentation skills after receiving computer-mediated delayed feedback

(such as perception of competence, presentation anxiety and attitude towards presenting). Starting with the perception of feedback, students highly appreciated the relevance of the feedback they received in both the experimental group (M = 4.01, SD = 0.79) as well as in the control group (M = 4.00, SD = 0.80). However, no differences between the conditions were found ($t(25)$ = 0.05, p = 0.96). Further, students also perceived the feedback they received as constructive and non-confrontational, encompassing students' feedback sensitivity, in both the experimental (M = 4.03, SD = 0.58) as well as the control condition (M = 4.02, SD = 0.59). Again, no significant differences were determined between the two groups ($t(25)$ = 0.06, p = 0.96). Further, the quality of feedback was highly appreciated on six out of the seven quality criteria of feedback in both conditions without significant differences (see Table 7.1). However, only the feedback criterion relating to 'opportunities to bridge the gap between the actual and desired performance' was scored lower than '4.0' in both conditions, which can therefore not be considered as 'sufficient' (Van Ginkel et al., 2017c). Despite of a lack of differences between the conditions, both in the qualitative VR feedback (M = 3.00, SD = 1.23) as well as in the quantitative VR feedback condition (M = 3.64, SD = 1.12) the scores on this feedback criterion were relatively lower. This might suggest that in follow-up experiments specific attention should be devoted not only on how feedback is provided to the actual presentation behaviour, but especially towards how feedback messages can be constructed in such a manner that they support strategies to develop presentation performances relating to the ideal or desired presentation behaviour.

Subsequently, students perceived their own development of presentation skills as more than sufficient, revealing the scores in the qualitative VR feedback condition (M = 6.57, SD = 1.27) and the quantitative VR condition (M = 5.75, SD = 1.36). Although no significant differences between these conditions were found on this perception of presentation skills ($t(25)$ = 1.61, p = 0.12), interestingly, differences exist between the two groups for the component of presentation anxiety. Within the no intervention expert condition with qualitative feedback messages, students scored significantly lower on their perceived presentation anxiety ($t(25)$ = −2.24, p = 0.034) after training in VR (M = 2.37, SD = 0.69) in comparison to the expert intervention condition with quantitative feedback reports (M = 3.08, SD = 0.92). This could be explained by the notion that students experience more pressure and perceive more stage fright after receiving feedback from a teacher. Therefore, these findings might suggest that training in VR, while receiving automated feedback without the intervention of an expert, can be considered as an effective strategy for reducing presentation anxiety in the stage of rehearsing speeches before presenting in front of real audiences and receiving feedback from experts. However, it remains questionable whether students experience similar levels of anxiety when peers are involved in the feedback process. Therefore, future studies will be undertaken in this area.

Another significant difference in this preliminary research was found between students of different domains regarding their attitude towards presenting (F(3, 23)

Table 7.1 Mean scores, SDs and N related to closed questions (5-point Likert scale) about perceptions of the feedback quality for students within the control condition (intervention expert) and the experimental condition (no intervention expert)

Items	Control condition (+ expert)	Experimental condition (− expert)	Difference between conditions
1. The feedback I received after my presentation is related to the pre-defined assessment criteria of the presentation task			
Mean	4.73	4.80	0.07
SD	0.47	0.45	0.25
N	11	5	16
2. I received valuable content-related arguments about how to improve my non-verbal communication aspects during my presentation			
Mean	4.55	4.20	−3.5
SD	0.69	0.84	0.40
N	11	5	16
3. I received valuable feedback on my actual behaviour (e.g. non-verbal communication) that I have shown during my presentation			
Mean	4.64	3.80	−0.84
SD	0.51	1.30	0.60
N	11	5	16
4. I received valuable feedback about the behaviour (e.g. non-verbal communication) that I should have shown during my presentation			
Mean	4.55	4.20	−0.35
SD	0.52	1.10	0.52
N	11	5	1
5. The feedback contained valuable tips and tricks to improve my actual presentation behaviour to the type of behaviour I should have shown during my presentation			
Mean	3.64	3.00	−0.64
SD	1.12	1.23	0.62
N	11	5	1
6. The type of feedback (e.g. form and length) is usable to me			
Mean	4.45	3.80	−0.66
SD	0.52	0.84	0.34
N	11	5	1
7. The feedback is formulated positively and constructively			
Mean	4.64	4.20	−0.44
SD	0.51	1.30	0.60
N	11	5	1

Table 7.2 Mean scores, SDs and N related to closed questions about students' attitudes towards presenting for different educational domains

Educational domain	M	SD	N
ICT	3.56	0.92	9
Education and pedagogy	4.52	0.33	10
Healthcare	3.73	0.60	6
Engineering	3.90	0.42	2

$= 3.86$, $p = 0.022$), which includes the perception of students regarding the relevance to acquire presentation skills and their motivation to train these skills. A Tukey post hoc test revealed that students' attitudes towards presenting were significantly lower for students within the ICT-domain ($M = 3.56$, $SD = 0.92$) compared to students within the educational and pedagogy domains ($M = 3.73$, $SD = 0.60$). The difference in self-perceived performance between the domains (see also Table 7.2) might refer to technical curricula focusing more on teaching domain-specific skills instead of integrating soft skills, such as presentation competencies, in their educational programs (e.g. Belboukhaddaoui & Van Ginkel, 2019). However, several recent studies in presentation research describe developing presentation skills in technical curricula (e.g. Mitrovic et al., 2017; Mohamed et al., 2015). Another argument for the lack of perceived presentation skills amongst technical students might relate to the idea that technical students naturally possess fewer communication competencies in comparison to students from non-technical curricula. Since there is a lack of evidence in empirical presentation studies regarding this issue, more research is needed towards (1) the integration of presentation environments in technical curricula and (2) the role of students' traits, prior competencies and perceptions towards presenting in relation to presentation performances (see also Van Ginkel et al., 2015).

In retrospective, besides varying perceptions of students regarding their presentation anxiety and attitude towards presenting depending on conditions and/or domains, students appreciated the value and relevance of the feedback they received in both the non-expert as well as the expert intervention condition. In follow-up projects, insights from these studies are being used to compose and construct feedback messages for analyzing and evaluating 'posture and gestures' in presentation education, since this is regarded as another essential component of non-verbal communication in presentations (Van Ginkel et al., 2015). VR technologies can support the provision of feedback on eye contact and use of voice. However, for monitoring body language, Artificial Intelligence (AI) technologies are more suited to monitor detailed posture and gestures of presenters. Therefore, a current project focuses on constructing an application for the smartphone that supports students' development in posture and gestures independently of time and place. By using AI technology, data about body language is converted into automatically generated feedback messages that supports students in their presentation development (see Fig. 7.1). Moreover, this application, entitled Honest Mirror, is aimed to meet design criteria regarding scalability, mobility, effectiveness and adoption in education. In order to guarantee the effectiveness of the app,

(1) The student is presenting

(2) The presentation is video recorded with a smartphone

(3) The neural network is analyzing the video

(4) The neural network is comparing effective and ineffective postures and gestures based on scientific literature

(5) The application is providing personalized & automated feedback on delivered postures and gestures
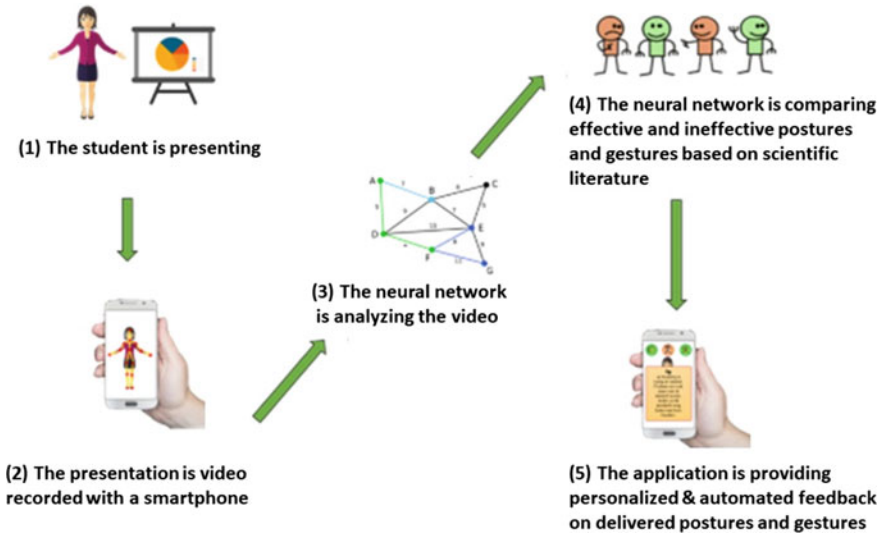
Fig. 7.1   The feedback model of the AI-driven app fostering students' body language during presentation rehearsals

the lessons learned from the earlier discussed VR studies are used for composing automatically generated feedback messages. Therefore, validated effective and ineffective postures and gestures were selected from the presentation literature (Schneider et al., 2017). Further, criteria for effective feedback in presentation research were adopted for constructing effective feedback messages (Van Ginkel et al., 2017a, 2017b). An example of such a message is: *"You used your hands during your presentation. If used effectively, this can reinforce the message. Still, in a subsequent presentation you could try not to put your hands in your pockets. This attitude can come across to the audience as casual and uninterested. Therefore, try to keep your hands relaxed next to the body or use supporting gestures to convey a message more powerfully. In that case, make sure you have open hands to make those gestures possible."* In order to encourage the adoption of this app in education as an alternative feedback source in peer learning, it will be published open source and the app will be connected to the previously constructed VR system, which is already adopted in higher education presentation curricula.

7.7   A Future Research Agenda on Computer-Mediated Feedback for Peer Learning in Presentation Research

After synthesizing varying review and empirical publications in the field of presentation competence development, it can be concluded that peer learning is considered as one of the crucial educational design principles for developing students' public speaking performances in higher education. However, it is also stated

that peer feedback is not yet as effective as teacher feedback, due to a lack of feedback quality in peer learning within current presentation curricula. From an educational technological point of view, VR technologies are regarded as valuable alternative feedback sources, since they can provide effective feedback comparable to teacher or expert feedback. However, while adopting VR technologies in presentation education, the role of teachers in guiding students, while guaranteeing high levels of feedback quality, should not be underestimated. Nevertheless, recent VR studies reveal that immediate feedback, without any support of teachers, is as effective as delayed feedback explained by teachers. Further, other studies revealed that computer-mediated delayed feedback messages, provided within VR systems without the support of teachers, are perceived as constructive and valuable by higher education students.

From a scientific perspective, based on synthesizing literature in the field of presenting and feedback, insights from this chapter might further refine the educational design principle regarding peer learning for developing students' oral presentation competence, since empirical evidence from recent studies emphasized the value of computer-mediated delayed feedback messages within VR regarding students' perceptions. However, it remains questionable to what extent combining different forms of feedback, such as immediate and delayed feedback, ánd combining different forms of technologies, such as VR and AI, could further optimize the effectivity of peer learning for developing varying aspects of oral presentation competence. Especially combining VR and AI could support the provision of such feedback messages on the most crucial non-verbal communication aspects, such as eye contact, use of voice (both supported by VR) and posture and gestures (supported by AI).

From an educational practice perspective, developing, testing and optimizing computer-mediated feedback messages by making use of innovative technologies in presentation education might release the pressure on teachers' tasks in providing effective and efficient presentation courses, since such feedback opportunities might increase the value of peer feedback while solely using teacher feedback in these stages of the learning process or for specific learning objectives of students when it is needed the most. In line with supporting students' learning processes and even for educating teachers, UNESCO emphasized the adoption of VR and AI technologies as crucial in the light of the global teacher shortage (Adubra et al., 2019; Parmigiani et al., 2020). If learners are able to individually interpret feedback messages without the intervention of a teacher, it could enrich the quality of feedback in peer and self-learning and further increase students' development in a wide range of academic, communication, digital literacy and domain-specific competencies.

However, several limitations still exist with regard to the earlier discussed studies, which should be taken critically into consideration while constructing a future research agenda on the topic of computer-mediated feedback in VR for improving peer learning in presentation research. First of all, although recent studies revealed positive perceptions of computer-mediated feedback messages with regard to the relevance and value of feedback for developing students' learning processes, it

is questionable to what extent these feedback messages can also be considered as effective for developing presentation performances. Second, although previous studies revealed effects of innovative technologies for delivering feedback, such as VR or AI, on developing public speaking competencies, the N-values of these studies are relatively low. Experimental follow-up studies should therefore incorporate higher numbers of students in order to detect significant results in presentation developments or potential differences between VR or teacher intervention conditions. Third, most of the publications on feedback in VR contexts fostering presentation competencies report on relatively short-term experiments. In line with this, it remains questionable what the effects of peer or self-learning in VR contexts are on the long term when students have the opportunity to rehears their presentations several times in VR and also have the opportunity to develop themselves based on computer-mediated feedback messages in multiple occasions.

A future research agenda on computer-mediated feedback for peer learning in presentation research should incorporate the following studies. First, an experimental study should be conducted focusing on the effects of computer-mediated delayed feedback on developing students' oral presentation competence. In such a study, the experimental condition should focus on the effects of students who individually interpret feedback messages without the support of teachers in VR, while the control condition consists of a situation in which students learn from feedback messages that are interpreted and provided by teachers. Such a study should reveal whether students do not only positively interpret earlier constructed feedback messages, as suggested in previous empirical studies, but to what extent these messages are also effective for developing their presentation competencies. Second, a follow-up study should concentrate on the effects of adopting computer-mediated feedback messages in peer learning in order to verify whether peer feedback can be optimized in terms of effects on developing students' presentation competencies. Previous studies revealed that the quality of peer feedback is lacking in comparison to feedback provided by teachers and feedback quality standards. However, it remains questionable whether peer feedback, supported by VR and AI technologies, could help to optimize this learning environment characteristics in presentation education. Such a study should also incorporate procedures of peer assessment by taking into account the complexity of peer feedback processes through integrating specific feedback stages for combining face-to-face and computer-mediated feedback in formative assessment (e.g. Baartman & Gulikers, 2017). Third, another follow-up empirical study should follow students in their learning processes from a longitudinal perspective while rehearsing presentations in VR and/or with the support of AI, learning from interpreting feedback messages and formulating new learning objectives towards presenting. As such, results might reveal not only the possibilities of such technologies for peer and self-learning, but also provide insights about the sustainability of adopting AI technologies in higher education curricula in times when education is under pressure due to teachers shortages and in times of pandemics that force learners to optimize their learning processes by embracing online education.

References

Adubra, E., Da Silva, I., Dhungana, B., Mohan, N., Saltsman, G., & Van Ginkel, S. (2019). Exploring the use of Artificial Intelligence to support teachers and teacher development. *UNESCO International Task Force on Teachers, 1*(19), 1–16.

Asghar, A. (2010). Reciprocal peer coaching and its use as a formative assessment strategy for first-year students. *Assessment and Evaluation in Higher Education, 35*(4), 403–417.

Baartman, L., & Gulikers, J. (2017). Assessment in Dutch vocational education: Overview and tensions of the past 15 years. In *Enhancing Teaching and Learning in the Dutch Vocational Education System* (pp. 245–266).

Baker, W. H., & Thompson, M. P. (2004). Teaching presentation skills. *Business Communication Quarterly, 67*(2), 216–219.

Belboukhaddaoui, I., & van Ginkel, S. (2019). Fostering oral presentation skills by the timing of feedback: An exploratory study in virtual reality. *Research on Education and Media, 11*(1), 25–31.

Boetje, J., & van Ginkel, S. (2021). The added benefit of an extra practice session in virtual reality on the development of presentation skills: A randomized control trial. *Journal of Computer Assisted Learning, 37*(1), 253–264; 1–12.

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment and Evaluation in Higher Education, 38*(6), 698–712.

Carroll, C. (2006). Enhancing reflective learning through role-plays: The use of an effective sales presentation evaluation form in student role-plays. *Marketing Education Review, 16*(1), 9–13.

Chang, S. C., Hsu, T. C., & Jong, M. S. Y. (2020). Integration of the peer assessment approach with a virtual reality design system for learning earth science. *Computers and Education, 146*, 103758.

Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing, 22*(1), 93–121.

Chien, S. Y., Hwang, G. J., & Jong, M. S. Y. (2020). Effects of peer assessment within the context of spherical video-based virtual reality on EFL students' English-Speaking performance and learning perceptions. *Computers and Education, 146*, 103751.

Clarke, P., Schull, D., Coleman, G., Pitt, R., & Manathunga, C. (2013). Enhancing professional writing skills of veterinary technology students: Linking assessment and clinical practice in a communications course. *Assessment and Evaluation in Higher Education, 38*(3), 273–287.

Coller, B. D., & Scott, M. J. (2009). Effectiveness of using a video game to teach a course in mechanical engineering. *Computers and Education, 53*(3), 900–912.

De Grez, L. (2009). *Optimizing the instructional environment to learn presentation skills* (Dissertation). Universiteit Gent.

De Grez, L., Valcke, M., & Roozen, I. (2009). The impact of an innovative instructional intervention on the acquisition of oral presentation skills in higher education. *Computers and Education, 53*(1), 112–120.

De Grez, L., Valcke, M., & Roozen, I. (2012). How effective are self-and peer assessment of oral presentation skills compared to teachers' assessments? *Active Learning in Higher Education, 13*(2), 129–142.

Econopouly, B. F., Byrne, P. F., & Johnson, M. A. (2010). Incorporating case studies into a world food and population course. *Journal of Natural Resources and Life Sciences Education, 39*, 79–83.

Falchikov, N. (2005). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education.* RoutledgeFalmer.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Hawkins, D. G. (1995). Virtual reality and passive simulators: The future of fun. In F. Biocca & M. R. Levy (Eds.), *Communication in the age of virtual reality* (pp. 159–189). Lawrence Erlbaum.

Kelly, D., Baxter, J. S., & Anderson, A. (2010). Engaging first-year students through online collaborative assessments. *Journal of Computer Assisted Learning, 26*(6), 535–548.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254.

Kolber, B. (2011). Extended problem-based learning improves scientific communication in senior-biology students. *Journal of College Science Teaching, 41*(1), 32–39.

Latifi, S., Noroozi, O., & Talaee, E. (2020). Worked example or scripting? Fostering students' online argumentative peer feedback, essay writing and learning. *Interactive Learning Environments*, 1–15. https://doi.org/10.1080/10494820.2020.1799032

Latifi, S., Noroozi, O., Hatami, J., & Biemans, H. J. A. (2021). How does online peer feedback improve argumentative essay writing and learning? *Innovations in Education and Teaching International, 58*(2), 195–206. https://doi.org/10.1080/14703297.2019.1687005

LaViola Jr, J. J., Kruijff, E., McMahan, R. P., Bowman, D., & Poupyrev, I. P. (2017). *3D user interfaces: Theory and practice.* Addison-Wesley Professional.

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12*(1), 107–119.

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers and Education, 70*, 29–40.

Mitchell, V. W., & Bakewell, C. (1995). Learning without doing—Enhancing oral presentation skills through peer-review. *Management Learning, 26*(3), 353–366.

Mitrovic, A., Gostomski, P., Herritsch, A., & Dimitrova, V. (2017). Improving presentation skills of first-year engineering students using active video watching. In *28th Annual Conference of the Australasian Association for Engineering Education* (p. 809).

Mohamed, A. A., Asmawi, A., Hamid, M. R. A., & Mustafa, Z. B. (2015). Descriptive analysis of context evaluation instrument for technical oral presentation skills evaluation: A case study in English techncal communication course. *AIP Conference Proceedings, 1643*(1), 536–540.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science, 32*, 99–113.

Noroozi, O., & Hatami, J. (2019). The effects of online peer feedback and epistemic beliefs on students' argumentation-based learning. *Innovations in Education and Teaching International, 56*(5), 548–557. https://doi.org/10.1080/14703297.2018.1431143

Noroozi, O., & Mulder, M. (2017). Design and evaluation of a digital module with guided peer feedback for student learning biotechnology and molecular life sciences, attitudinal change, and satisfaction. *Biochemistry and Molecular Biology Education, 45*(1), 31–39. https://doi.org/10.1002/bmb.20981

Noroozi, O., Kirschner, P. A., Biemans, H. J. A., & Mulder, M. (2018). Promoting argumentation competence: Extending from first- to second-order scaffolding through adaptive fading. *Educational Psychology Review, 30*(1), 153–176. https://doi.org/10.1007/s10648-017-9400-z

Noroozi, O., Weinberger, A., Biemans, H. J. A., Mulder, M., & Chizari, M. (2012). Argumentation-based computer supported collaborative learning (ABCSCL). A systematic review and synthesis of fifteen years of research. *Educational Research Review*, *7*(2), 79–106. https://doi.org/10.1016/j.edurev.2011.11.006

Parmigiani, D., van Ginkel, S., Saltsman, G., & Dhungana, B. (2020). Innovation in teacher education. *Italian Journal of Educational Technology, 28*(2), 103–105.

Pittenger, K., Miller, M., & Mott, J. (2004). Using real-world standards to enhance students' presentation skills. *Business Communication Quarterly, 67*(3), 327–336.

Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: All that effort, but what is the effect? *Assessment and Evaluation in Higher Education, 35*(3), 277–289.

Reis, S. M., & Renzulli, J. S. (2010). Is there still a need for gifted education? An examination of current research. *Learning and Individual Differences, 20*(4), 308–317.

Schneider, J., Börner, D., van Rosmalen, P., & Specht, M. (2017). Presentation trainer: What experts and computers can tell about your nonverbal communication. *Journal of Computer Assisted Learning, 33*(2), 164–177.

Shaw, V. (2001). Training in presentation skills: An innovative method for college instruction. *Education, 122*(1), 140–144.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.

Sichterman, B., Schipper, M., Verstappen, M., Waisvisz, P., & van Ginkel, S. (2021). The impact of computer-mediated delayed feedback on developing oral presentation skills: An experimental study in virtual reality. In A. Pucihar, M. Kljajic Borstnar, R. Bons, H. Cripps, A. Sheombar, & D. Vidmar (Eds.), *Digital support: From crisis to progressive change* (pp. 683–690). University Press.

Taghizadeh Kerman, N., Noroozi, O., Banihashem, S. K., Karami, M. & Biemans, Harm. H. J. A. (2022). Online peer feedback patterns of success and failure in argumentative essay writing. *Interactive Learning Environments*, 1–10. https://doi.org/10.1080/10494820.2022.2093914

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276.

Tsaushu, M., Tal, T., Sagy, O., Kali, Y., Gepstein, S., & Zilberstein, D. (2012). Peer learning and support of technology in an undergraduate biology course to enhance deep learning. *CBE-Life Sciences Education, 11*(4), 402–412.

Van den Berg, I., Admiraal, W., & Pilot, A. (2006). Design principles and outcomes of peer assessment in higher education. *Studies in Higher Education, 31*(3), 341–356.

Van Ginkel, S., Gulikers, J., Biemans, H., & Mulder, M. (2015). Towards a set of design principles for developing oral presentation competence: A synthesis of research in higher education. *Educational Research Review, 14,* 62–80.

Van Ginkel, S., Gulikers, J., Biemans, H., & Mulder, M. (2017a). The impact of the feedback source on developing oral presentation competence. *Studies in Higher Education, 42*(9), 1671–1685.

Van Ginkel, S., Gulikers, J., Biemans, H., & Mulder, M. (2017b). Fostering oral presentation performance: Does the quality of feedback differ when provided by the teacher, peers or peers guided by tutor? *Assessment and Evaluation in Higher Education, 42*(6), 953–966.

Van Ginkel, S., Laurentzen, R., Mulder, M., Mononen, A., Kyttä, J., & Kortelainen, M. J. (2017c). Assessing oral presentation performance: Designing a rubric and testing its validity with an expert group. *Journal of Applied Research in Higher Education, 9*(3), 474–486.

Van Ginkel, S., Ruiz, D., Mononen, A., Karaman, C, De Keijzer, A., & Sitthiworachart, J. (2020). The impact of computer-mediated immediate feedback on developing oral presentation skills: An exploratory study in virtual reality. *Journal of Computer Assisted Learning,* 412–422.

Van Ginkel, S., Gulikers, J., Biemans, H., Noroozi, O., Roozen, M., Bos, T., Van Tilborg, R., Van Halteren, M., & Mulder, M. (2019). Fostering oral presentation competence through a virtual reality-based task for delivering feedback. *Computers and Education, 134*, 78–97.

Van Haaren, M., & Van der Rijst, R. (2014). Teaching talented students at a research-intensive university. In M. V. C. Wolfensberger, L. Drayer & J. J. M. Volker (Eds.), *Pursuit of excellence in a networked society: Theoretical and practical approaches* (pp. 39–46). Waxmann.

Von Konsky, B. R., & Oliver, B. (2012). The iPortfolio: Measuring uptake and effective use of an institutional electronic portfolio in higher education. *Australasian Journal of Educational Technology, 28*(1), 67–90.

Web-Based Peer Assessment Platforms: What Educational Features Influence Learning, Feedback and Social Interaction?

8

José Carlos G. Ocampo and Ernesto Panadero

8.1 Introduction

The use of web-based peer assessment has exponentially increased in the last couple of decades due to its benefits for both instructors and students. Among these benefits, it is usually argued that web-based peer assessment lessens instructors' workload by automatically managing peer assessment data (e.g., ratings, feedback) (Bouzidi & Jaillet, 2009), helps in conducting formative assessment (Søndergaard & Mulder, 2012), can help develop students' motivation (Lai & Hwang, 2015), critical thinking (Wang et al., 2017), and positive affect (Chen, 2016). Nevertheless, there are also challenges to the implementation of web-based peer assessment. Some students perceive that web-based peer assessment is unfair (Kaufman & Schunn, 2011), academics find it challenging to create online learning environments (Adachi et al., 2018b), and some features of these web-based peer assessment platforms might limit the interpersonal and collaborative nature of peer assessment (Panadero, 2016; van Gennip et al., 2009) since some platforms are not capable of transmitting non-verbal cues needed for interaction (Phielix et al., 2010). Thus, as much as web-based peer assessment has great potentials, it also

J. C. G. Ocampo (✉) · E. Panadero
Facultad de Educación y Deportes, ERLA Research Group, Universidad de Deusto, Bilbao, España
e-mail: jc.ocampo@deusto.es

E. Panadero
Ikerbasque, Basque Foundation for Science, Bilbao, Spain

O. Noroozi and B. de Wever (eds.), *The Power of Peer Learning*,
Social Interaction in Learning and Development,
https://doi.org/10.1007/978-3-031-29411-2_8

brings challenges and a key aspect for the success or failure in the implementation are the features offered in the web-based peer assessment platforms.

Because of this, our aim is to evaluate the characteristics and features of web-based peer assessment platforms to explore whether they facilitate students learning, feedback and social interaction. There are a number of reviews that have already compared and contrasted the different features and tools embedded in different web-based peer assessment platforms (e.g., Babik et al., 2016; Luxton-Reilly, 2009; Søndergaard & Mulder, 2012). However, we believe that there is a need to look at these platforms from an educational assessment lens, as previous reviews have reviewed the platforms from a computer science education or software engineering education lens. In doing this, we can determine its potential benefits and potential constraints in instruction and student interaction. With this in mind we used the peer assessment design elements framework (i.e., Adachi et al., 2018a) to determine how features of these platforms can affect three variables: students' learning, the feedback provided through the platform, and the dynamics of student interaction online.

8.1.1   Web-Based Peer Assessment Platforms

The mode in which peer assessment is carried out is one of the most important decisions that teachers and instructors have to take if they want to implement peer assessment (Topping, 1998). One of the main decisions is whether to use a web-based platform or a more traditional paper-based approach. A recent meta-analysis that reviewed close to 60 studies found that web-based peer assessment shows larger effect size than paper-based peer assessment ($g = 0.452$ vs $g = 0.237$), which means that web-based may be preferable (Li et al., 2020). Similarly, web-based peer assessment was also deemed to be more convenient and flexible than paper-based peer assessment (Chen, 2016; Wen & Tsai, 2008) since it can be used synchronously or asynchronously with any web-connected device (e.g., computers, mobile devices, etc.) in different environments (e.g., classroom, or home) (Fu et al., 2019). Also, web-based peer assessment has specific features that might be too laborious to do in paper-based peer assessment, like allocating different grading weights at different stages of peer assessment to maintain reliability and validity of peer scores, algorithm-based pairing for assessors and assessees, or maintaining double-blind anonymity during the peer assessment process (Cho & Schunn, 2007; Patchan et al., 2018), or just simply aggregating and managing peer scores and peer feedback data in big courses. Because of the variety of tools and features available in different peer assessment platforms, a number of articles have reviewed different computer-supported and/or web-based peer assessment platforms available in published literature or in the educational technology market. Next, we discussed the three most relevant of these reviews.

First, Luxton-Reilly (2009), looked at the common features as well as the differences of various peer assessment platforms in a systematic review. He compared web-based peer assessment platforms based on: rubric design (i.e., if it is fixed or

modifiable); rubric criteria (i.e., if it supports Boolean criteria like checkboxes; discrete choices; numeric scales; and, textual comments); possibility of discussion (i.e., dialogue between assessor and assessee); option to give backward feedback (i.e., if students and or instructors can assess the quality of feedback); flexibility of workflow (i.e., if the platform allows instructors to organise peer assessment workflow); and, evaluation (i.e., if there is a post evaluation performed in the study). Additionally, he categorized the platforms in three groups based on their context: generic, domain-specific, and context-specific systems. He grouped six peer assessment platforms under the "generic systems" where most features and activities in the platform can be configured by the instructor to cater to different disciplines and contexts. Seven platforms were grouped under "domain-specific systems" that were designed for specific disciplines (e.g., programming, essay writing). Finally, five platforms were grouped under "context-specific systems", for platforms programmed solely for specific courses. Moreover, he expressed the need to further improve, or develop, web-based peer assessment platforms since the majority of the platforms he reviewed (13 of 18) were limited to computer science courses and settings. This review intended to serve as a helpful guide for developers in improving the design and features of existing and subsequent web-based peer assessment platforms.

Second, Søndergaard and Mulder (2012) evaluated peer assessment platforms based on four characteristics: (1) the ease of automation: automatic anonymisation and distribution of outputs and notification of instructors and students; (2) simplicity: convenience of the interface, ease of managing student data and integration with other learning management systems, and availability of resources for teachers and students; (3) customisability: flexibility to configure based on course needs; and, (4) accessibility: subscriptions and availability of a system online. They also analysed other features that might be essential to different contexts, like guidelines in pairing assessors and assessees, student assessor training/calibration, built-in plagiarism checks, and reporting tools to monitor the quality of feedback. Additionally, they categorised four web-based peer assessment platforms based on their focus, such as being training oriented, similarity checking oriented, customisation oriented, or writing skills oriented. This work provided an interesting framework for educators to evaluate if a web-based peer assessment platform is an appropriate formative and collaborative tool that support learning and student interaction, rather than a mere tool that collects peer scores or feedback.

Third and last, Babik et al. (2016) developed a peer-to-peer focused framework for evaluating the affordances and limitations of web-based peer assessment platforms based on an informal focused group discussion with instructors using web-based peer assessment in their courses and guided by the relevant practices of the peer assessment studies they reviewed from academic papers. Based on the categorized discussion of instructors' practices, they listed five primary objectives for web-based peer assessment: (1) eliciting evaluation; (2) assessing achievement and generating learning analytics; (3) structuring automated peer assessment workflow; (4) reducing or controlling for evaluation biases; and (5) changing social

atmosphere of the learning community as the main objectives for the use of web-based peer assessment. They viewed these objectives as "system independent", where instructors determine what they need for instruction outside of the platform. While the functions and design in the platforms are categorized under "system-dependent" features. This study is important because they looked at platforms from the point-of-view of individuals making decisions on how web-based peer assessment is implemented in various courses—the instructors.

Taken altogether, these three reviews of web-based peer assessment platforms were made to assist instructors in planning their lessons to integrate student-centred assessment practices such as peer assessment. Additionally, they provided an overview of the technological advances in implementing web-based peer assessment. Nonetheless, there is still a need to investigate web-based peer assessment platforms from the point-of-view of a peer assessment design elements perspective, since the reviews we just presented were construed from a computer science education or software engineering context. Moreover, there is an increase in the number of platforms developed and updated since the last review, which poses the need to further investigate and determine the current directions of web-based peer assessment platforms. In the next section, we will describe the framework we utilised in evaluating each platform.

8.1.2   Peer Assessment Design Elements Framework

Topping (1998) wrote one foundational study to clarify how peer assessment can be carefully carried out in classrooms and research. He proposed a typology including seventeen variables, which were: (1) curriculum area; (2) objectives; (3) focus; (4) product/output; (5) relation to staff assessment; (6) official weight; (7) directionality; (8) privacy; (9) contact; (10) year; (11) ability; (12) constellations assessors; (13) constellations assessees; (14) place; (15) time; (16) requirement; and, (17) reward. The typology gave way for instructors and assessment researchers to construe peer assessment in an organised and systematic manner even if, unfortunately, it is still under-used and under-reported (Panadero, 2016). Importantly, since the original typology by Topping, there has been a number of new proposals that reorganize or amplify the original categories. For instance, van den Berg et al. (2006), categorised Topping's variables into four clusters to respond to their course context, while van Gennip et al. (2009) classified the variables into three clusters considering how social interactions occurs between students in peer assessment.

More recently, Adachi et al. (2018a) added an additional dimension to Gielen et al.'s (2011) five-cluster work that reviewed and organised earlier ideas on peer assessment, which covered: (1) the decisions concerning peer assessment use; (2) peer assessment's link to other elements in the learning environment; (3) interaction between peers; (4) composition of assessment groups; (5) management of assessment procedures, and (6) contextual elements. This peer assessment design

elements framework is composed of 19 design elements that consider the diversity of peer assessment strategies, which were obtained from literature synthesis and their interview with academics from different disciplines. The design elements in this framework modified previous frameworks (e.g., Gielen et al., 2011; Topping, 1998) by collapsing, combining, and adding elements to form a unified one. For example, some elements were combined into one (i.e., requirement + reward into 'formality and weighting'), while others were added into the framework (e.g., feedback utilisation). We have decided to use this framework as it covered design elements that are useful in future studies (i.e., Cluster VI: Contextual Elements).

In sum, the multiple iterations of peer assessment typologies suggest the idea that there is no "one size fits all" approach in implementing in the classroom and doing research in peer assessment. Also, it suggests that peer assessment is a complex process that requires further investigation due to rapid changes in the educational landscape. Therefore, there is a need to explore web-based peer assessment platform features to determine how it can affect students' learning, the feedback that students provide and receive, and the dynamics of student interaction online. Examining the features of web-based peer assessment platforms that provide support to various interpersonal and intrapersonal factors that students go through during peer assessment is crucial since evidence has mentioned that it helps in promoting positive educational and affective outcomes (Chen, 2016; Lai & Hwang, 2015; Wang et al., 2017). Also, it is important to look at these interpersonal and intrapersonal factors because the interaction that occurs between students in web-based environments as a result of the features of web-based peer assessment may generate different social and human factors (i.e., thoughts, emotions, actions) that affect peer assessment outcomes (Panadero, 2016). Given that, there is a need to investigate the features of web-based peer assessment platforms from a peer assessment design elements framework to ascertain how these platforms can support peer assessment and student interaction online. Thus, we decided to perform a systematic review of platforms.

8.1.3 Search, Screening and Access to the Platforms, and Review Criteria

We used two approaches to identify the platforms. First, we extracted names of web-based peer assessment platforms from a parallel systematic review on intrapersonal and interpersonal variables in peer assessment. Second, a peer assessment expert was consulted for web-based peer assessment platform recommendations. In total, we identified 31 web-based platforms.

In screening the platforms, we visited each platform's website to evaluate its availability. From this, 8 platforms were excluded (i.e., social media site, company tool, website in foreign language, website was unavailable or ceased to operate). Subsequently, the developers of the remaining platforms were contacted to request for complementary access to their platform if no free sign-up was available, as some required payment or licensing, or were offered exclusively for a

select number of institutions. From this, 6 platforms were excluded (i.e., developers did not grant access or were unresponsive, platform was made for a specific course/commercially unavailable).

Finally, 17 web-based peer assessment platforms were evaluated in this study, which are: Aropä (United Kingdom); Blackboard Learning Management System (United States); Canvas Learning Management System (United States); CATME (United States); CritViz (United States); Crowd Grader (United States); Eduflow (Denmark); Eli Review (United States); Expertiza (United States); Kritik (Canada); Mobius SLIP (United States); Moodle Learning Management System (Australia); Peerceptiv (United States); Peergrade (Denmark); PeerMark (United States); PeerScholar (Canada); and, TEAMMATES (Singapore).

In evaluating the features of each web-based peer assessment platform, we extracted nine peer assessment design elements from Adachi et al.'s (2018a) framework covering three different areas. First, we evaluated the features that might have a direct influence in *students' learning*, since a number of studies have expressed that some features of computer-supported collaborative learning environments (e.g., web-based peer assessment platforms) affects learning and performance (Janssen et al., 2007; Phielix et al., 2010, 2011; Zheng et al. 2020). Second, we evaluated the features that influence the *feedback* that student provide and receive when peer assessing since feedback is an essential component of peer assessment for both assessors and assessees (Gielen & De Wever, 2015; Patchan et al., 2016; Voet et al., 2018). Third, we evaluated aspects of *social interaction* between students since peer assessment is essentially a social and interpersonal process (Panadero, 2016; van Gennip et al., 2009). Table 8.1 shows the peer assessment design elements we selected and corresponding descriptions.

We coded the relevant information from each platform to a standard data extraction template. In most web-based peer assessment platforms, we created a standard sample activity where peer assessment was the main focus. Then, we looked at the feature options available when designing the activity which would relate to a certain design element (e.g., choosing "enable self-evaluation?" would relate to design element number 8; choosing "enable anonymity?" would relate to design element 9). When the information about certain design elements was unclear, we used the search function in the help centre or search bar available in the platform. To assess the validity of the coding, an external researcher conducted an independent coding of three of the 17 platforms included in this study, which resulted to 91.2% agreement. In the next sections, we will examine how the features of the 17 web-based peer assessment platforms influences learning, feedback, and social interaction.

8.2   Web-Based Peer Assessment Features Influencing Student Learning

In this section, we will analyse features of the web-based peer assessment platforms in terms of how they might influence student learning based on the

Table 8.1   Peer assessment design elements

Categories	Cluster	Design Elements	Description
Student learning: Web-based peer assessment features that affects student learning	Cluster 1: Decisions concerning the use of peer assessment	(2) Intended learning outcomes (for students)	• What should students achieve through this activity? • Peer assessment/self-assessment/teamwork?
	Cluster 2: Link between peer assessment and other elements in the learning environment	(8) Link to self-assessment	• Can students also do self-assessment at any stage?
	Cluster 5: Management of assessment procedure	(15) Calibration and task scaffolding	• How are students oriented to standards prior to using them? • Is there a peer assessment training for students?
Feedback: Web-based peer assessment features that influences assessor's and assessee's feedback	Cluster 3: Interaction between peers	(10) Feedback information type	• Quantitative or qualitative; written, recorded? • Scores and/or comments? • Multimedia? Voice recorded? Video recorded?
		(11) Feedback utilization	• How is the feedback information used by the peer? • Opportunity for resubmission?
	Cluster 5: Management of assessment procedure	(16) Moderation of feedback	• Is feedback checked prior to communication? • Is there a way to examine the accuracy of scores or validity of feedback?
Social interaction: Web-based peer assessment features that affects student interaction	Cluster 3: Interaction between peers	(9) Anonymity	• Do students know who gave them feedback? • Single-blind/double-blind/non-anonymous?
	Cluster 4: Composition of groups	(12) Peer configuration	• Individual or group assessments/submission?
		(13) Peer matching	• How are students matched? • System generated matching or staff matching?

following design elements: intended learning outcomes (for students), link to self-assessment, and calibration and scaffolding.

8.2.1   Intended Learning Outcomes for Students

With regard to the intended learning outcomes for students, Adachi et al. (2018a) construed it as a range of possible outcomes (e.g., transferable skills) as a result of peer assessment. In this case, we regarded it as the possible assessment activities that can be paired with peer assessment in the web-based peer assessment platform. From the platforms we reviewed, it was possible for 8 (47%) of the platforms to combine peer assessment, self-assessment, and team member evaluation in the design of an activity (i.e., Blackboard; Eduflow; Expertiza; Kritik; Moodle; Peerceptiv; PeerMark; and, PeerScholar). On the other hand, 4 (23.5%) of the platforms allowed instructors to include both peer assessment and self-assessment when setting up an activity (i.e., Aropä; Eli Review; Mobius SLIP; and, PeerGrade). Also, 3 (17.6%) platforms allowed teachers to arrange peer assessment of submitted outputs at the time of our data collection (i.e., Canvas, CritViz, and Crowd Grader), while 2 (11.8%) platforms were designed for team member evaluation in group works (i.e., CATME and TEAMMATES). Generally, the majority of the platforms can be used in a variety of educational fields and levels due to its flexible and modifiable nature. This flexibility allows the instructors to mix and match features that they wish to integrate in their class based on their intended learning outcomes for students. Such option is especially powerful given that instructors obviously play a central role in implementing new assessment designs in their courses, particularly peer assessment (Panadero & Brown, 2017).

8.2.2   Link to Self-assessment

Previous studies have acknowledged the benefits of the intertwined roles of peer assessment and self-assessment (Boud, 2013; Dochy et al., 1999; To & Panadero, 2019). Therefore, it was not surprising that the majority of the platforms had a self-assessment feature. To illustrate, there were 12 (70.6%) platforms where self-assessment (or self-critique, self-review; self-evaluation; self-check, etc.) was integrated in the design of the web-based peer assessment platform (i.e., Aropä; Blackboard; CATME; Eduflow; Expertiza; Kritik; Mobius SLIP; Peerceptiv; Peer-Grade; PeerMark; PeerScholar; and TEAMMATES). Also, 1 (5.9%) platform did not appear to have 'self-assessment' as a named feature, but it has a different feature (e.g., Revision Notes) which can be considered as self-assessment (i.e., Eli Review). There were also 2 (11.8%) platforms that facilitated self-assessment, but it required instructors to set it up in a different feature (e.g., plug-in installation; as a quiz or survey; adding questions) (i.e., Canvas; Moodle). Finally, 2 (11.8%) platforms did not appear to have a self-assessment feature when we extracted information (i.e., CritViz; Crowd Grader).

Therefore, it can be said that in most of the platforms instructors would just have to click a few options to enable students to self-assess. Other platforms on the other hand, require self-assessment to be in an external activity, which may

require a little work for instructors to set up. It is important to note that self-assessment was called with various terms in most of the platforms. Given that self-assessment has become an integral part of the platforms, it is important for instructors to carefully plan how self-assessment and peer assessment would be combined to reap the benefits of it. More than just simply making students rate the quality of their work or asking surface questions about students' perception of their submission, it would be more powerful if students could assess their work against concrete standards and criteria to facilitate better reflection during self-assessment (Panadero et al., 2016).

8.2.3   Calibration and Task Scaffolding

In terms of calibration and task scaffolding, 4 (23.5%) platforms had a built-in training and/or practice feature that students had to go through before proceeding with the peer assessment exercise (i.e., CATME; Kritik; PeerScholar; and, Expertiza). Additionally, in three of these platforms students could practice their peer scoring skills on fictitious team members or sample outputs before proceeding with the actual peer assessment (i.e., CATME; Kritik; Expertiza). There is also an option in a platform where an instructor could embed "Microlearning Experience" videos about giving effective peer feedback and accurate peer scores before assessing peer's outputs (i.e., PeerScholar). While external training may depend on the instructor in the rest of the platforms (e.g., setting up an additional practice assessment activity before the actual peer assessment activity), the majority of the platforms offered a support page in their website with materials about how to give helpful feedback or accurate scores to peers (e.g., YouTube videos, guide prompts; articles). In such instances, it may require some effort for the students to navigate around the website to find such resources. Therefore, it would be more helpful if web-based platform developers integrate features for scaffolding and training prior to the actual peer assessment activity since providing students with sufficient and proper scaffolding to perform peer assessment, through multiple training and practice sessions, improves their assessment skills (Double et al., 2020; Li et al., 2020).

8.3   Web-Based Peer Assessment Features Influencing Feedback

In this section, we will analyse how the features of web-based peer assessment platforms influenced the feedback that students give to each other, based on the following design elements: feedback information type, feedback utilization, and moderation of feedback.

8.3.1 Feedback Information Type

In terms of feedback information type, all the 17 (100%) platforms supported both quantitative (e.g., peer scores) and qualitative (e.g., peer feedback) peer assessment. It is also important to note that some platforms also allowed the provision of multimedia recorded feedback (via audio or video). Also, the platforms offered a flexible way for instructors to set up their rubrics for peer scoring and prompts for peer feedback. For instance, these platforms allowed instructors to upload or create their rubrics or write their prompts in the website, or to adapt existing rubrics or prompts. These are important features since having students give and receive both quantitative and qualitative feedback are obviously the central actions of peer assessment (Topping, 1998).

In relation to platforms that support multimedia recorded feedback, evidence has shown that such feedback delivery approach helps in promoting deeper learning for assessors and assessees (Filius et al., 2019). However, although evidence showed that students provided better quality peer feedback in audio recorded mode than in a written mode, students perceived that preparing audio recorded peer feedback was not efficient (Reynolds & Russell, 2008). Importantly, students still preferred receiving written peer feedback over audio recorded peer feedback in a writing task (Reynolds & Russell, 2008). Granting that listening to recorded peer feedback may appear to be beneficial, the additional preparation involved might bring more work for students. Also, it might present challenges for instructors to manage students' multimedia feedback since they also have to keep track of, not just the peer feedback messages itself, but also each assessor's non-verbal gestures for video feedback, or prosodic features for audio feedback (e.g., intonation, stress, rhythm, etc.). Nonetheless, multimedia recorded peer feedback is important since it overcomes the limitations of text-based communication (e.g., absence of non-verbal cues) (Phielix et al., 2010). Therefore, further studies should consider looking at how the features of these multimedia recorded feedback can influence the dynamics between assessors and assessees in a web-based peer assessment environment.

8.3.2 Feedback Utilization

The uptake or utilization of feedback that students receive from various sources (e.g., peer, self, instructor) has been one of the focus of many feedback models (see Lipnevich & Panadero, 2021 for a review). Many web-based peer assessment platforms materialize this by integrating a resubmit function in their platforms. For instance, 14 (82.35%) of the platforms allowed students to submit multiple revisions of their work after peer assessment (i.e., Aropä; Blackboard; Canvas; CritViz; Eduflow; Eli Review; Expertiza; Kritik; Mobius SLIP; Moodle; Peerceptiv; PeerGrade; PeerMark; and, PeerScholar), while 1 (5.9%) platform did not seem to have a resubmission feature, but instructors may set up another assignment to allow resubmission (i.e., Crowd Grader). On the other hand, resubmission

was not applicable for 2 (11.8%) platforms since it was developed to evaluate team members in a group task (i.e., CATME and TEAMMATES). Allowing students to resubmit their output, whether in-class or online, after receiving feedback facilitates assessment *for* learning, which can be beneficial for students (Black & Wiliam, 1998; Panadero et al., 2016). This provides students multiple opportunities to improve their work, while it also gives instructors multiple indices to determine how students are learning.

8.3.3 Feedback Moderation

In terms of the moderation of feedback, 10 (59%) platforms had a built-in mechanism for assessees responses to assessor's judgements, disputing the peer scores received, or complain about inappropriate feedback (i.e., Aropä; Crowd Grader; Eduflow; Eli Review; Kritik; Mobius SLIP; Moodle; Peerceptiv; PeerGrade; and, Peer Scholar). To illustrate, these platforms allowed assesses to rate assessor's feedback based on a variety of criteria (e.g., helpfulness, motivating, etc.), which instructors may integrate in the final grade. Also, there are features where assesses can "return the feedback" (or back-evaluate/back-review) on assessors' feedback by giving suggestions on how the feedback can be improved, engaging in anonymous collaboration to ask for further advice, or simply ask for clarification if assessors' feedback was vague (i.e., Aropä; Crowd Grader; Eduflow; Eli Review; Peerceptiv; PeerGrade; and, Peer Scholar). Additionally, some platforms also allowed students to flag inappropriate feedback or inaccurate scores, where the instructor would have to mediate to settle differences (i.e., Kritik; Moodle; PeerGrade). Besides assessees' ratings of each feedback, it was also possible to automatically compare an assessor's rating based on several indices (e.g., against other assessors of the same output) (i.e., Mobius SLIP). On the other hand, there are 3 (17.6%) platforms where the instructor may choose to censor or rate a feedback if it is inappropriate or inaccurate (i.e., Blackboard; PeerMark; and, TEAMMATES). The other 4 (23.5%) platforms relied on instructor's manual monitoring of the process to moderate peer assessment (i.e., Canvas; CATME; CritViz; and, Expertiza).

Since students generate various thoughts, feelings, and actions in peer assessment (Panadero, 2016; Topping, 2021), it is not surprising that students may be concerned about retaliation when giving peers a critical feedback or low score (Patchan et al., 2018). Therefore, promoting student accountability in peer assessment is vital given that most web-based peer assessment activities in the majority of the platforms are anonymous. Although this feature may facilitate a discussion between assessors and assessees by allowing them to interact during the feedback process, such as in back-evaluations, we believe that investing more time in training students' assessment skills—whether in web-based or face-to-face settings—will be more fruitful than encouraging students to do well in peer assessments because their peers would rate the quality or accuracy of their feedback, or because their peer's rating of their feedback would be part of their course grade.

Developing students' assessment skills will enhance their evaluative judgement, which may be useful beyond schooling (Tai et al., 2018). Thus, finding the right balance between developing students' assessment skills, and making them accountable for the feedback they give is an area that should be considered by instructors and platform developers.

8.4   Web-Based Peer Assessment Features Influencing Social Interactions

In this section, we will analyse how the features of the web-based peer assessment platforms affect interaction between students based on the following design elements: anonymity, peer configuration, and peer matching.

8.4.1   Anonymity

Implementing anonymity in peer assessment has been the subject of intensive discussion in recent years (see Panadero & Alqassab, 2019 for a review). Some studies suggest that anonymity is beneficial for students' performance (Li, 2017; Lu & Bol, 2007) and their affect (Raes et al., 2015; Rotsaert et al., 2018; Vanderhoven et al., 2015), while others questioned its role in formative peer assessment activities since assessors and assesses are supposed to know each other to process feedback (Strijbos et al., 2009). In the web-based peer assessment platforms we reviewed, there were 15 (88.2%) platforms with a double-blind anonymity feature (e.g., completely unidentifiable, assignment of a number or pseudonym) and most of these platforms have options to remove the double-blind anonymity feature to make assessors and assessee identifiable (i.e., Aropä; Blackboard; Canvas; CritViz; Crowd Grader; Eduflow; Eli Review; Expertiza; Kritik; Mobius SLIP; Moodle; Peerceptiv; PeerGrade; PeerMark; and, PeerScholar). Finally, 2 (11.8%) platforms had a single-blind anonymity since it was designed for team member evaluation, where assessors know the identity of the assessee (typically their groupmate) they are assessing (i.e., CATME and TEAMMATES).

Since peer assessment is an interpersonal and social activity (Panadero, 2016; van Gennip et al., 2009), it is important to carefully plan the activities, so that students feel comfortable and safe. This was also noted in a previous web-based peer assessment platform review, which suggested to software developers to consider various institutional regulations in managing student privacy during peer assessment (Luxton-Reilly, 2009). The platforms we evaluated considered this aspect of peer assessment by integrating a number of flexible anonymity settings in their system. Then again, decisions lie with the instructors since there might be activities where peer assessment should be anonymised, and activities where putting off anonymity might be more beneficial for student interaction.

8.4.2 Peer Configuration

In terms of peer configuration, Gielen et al. (2011) notes that peer assessment can be done individually between students, between groups, or a combination of both. All of the 17 (100%) platforms we evaluated allowed a purely individual peer assessment between students (e.g., one assessor and one assessee). Additionally, there were 2 (11.8%) platforms designed for team member evaluation in group tasks (e.g., members of a group assesses each other in terms of helpfulness, contributions, etc.) which may not require a group submission of an output (i.e., CATME and TEAMMATES). Since the majority of the platforms allowed group submission, one would assume that they also allowed inter-group peer assessment (e.g., one group would assess another group's output). Of the 15 platforms that allow group submission, there were 8 (53.3%) platforms which allowed both individual submission and individual peer assessment, as well as group submission and inter-group peer assessment (i.e., Aropä; Canvas; Eduflow; Kritik; Mobius SLIP; Peerceptiv; PeerGrade; and, Peer Scholar). Finally, there were 7 (46.7%) platforms which supported individual submission and individual peer assessment, as well as group submission, but it was unclear if they also supported inter-group peer assessment during the period of our data extraction (i.e., Blackboard; CritViz; Crowd Grader; Eli Review; Expertiza; Moodle; and, PeerMark).

In a recent article, Topping (2021) noted that the constellation of assessors and assessees during peer assessment can be a complicated decision to make. For instance, instructors have to consider: how students (or groups) created the output to be assessed? If peer assessment is to be done individually, by pairs, or by groups? Will peer assessment be reciprocal? While these questions can be answered by the instructor's objectives in performing peer assessment, the platforms we evaluated offered an array of options in configuring students to perform various peer assessment activities. Thus, choosing the right option to support better student interaction is crucial when planning peer assessment activities.

8.4.3 Peer Matching

With regards to how assessors and assessees are matched in the platforms we evaluated, 14 (82.4%) offered both system and instructor matching, where the platform allocated assessors and assesses based on its own algorithms for the prior, and a manual matching of assessors and assessees for the latter (i.e., Aropä; Blackboard; Canvas; CATME; CritViz; Eduflow; Eli Review; Expertiza; Mobius SLIP; Moodle; Peerceptiv; PeerGrade; PeerMark; and, Peer Scholar). The majority of these platforms also offered flexibility for instructor to set the minimum or maximum number of assessors per output. There were also platforms where the instructor may choose to keep the same matching of assessor and assessee per draft submission, or randomize the pairing every submission, or manually match students per submission (i.e., Peerceptiv). Apart from system and instructor matching, some

platform also offered students to self-select outputs they want to assess (i.e., Peer-Mark). There were also unique platforms where the students fill a survey based on several aspects (e.g., schedule, sex, race, etc.), and the instructor may choose to system match students based on survey result similarity or diversity (i.e., CATME). On the other hand, 1 (5.9%) platform used artificial intelligence to match students based on several factors (e.g., equal distribution of weak and strong assessors for an output; or matching students based on output similarity) (i.e., Kritik). Also, 1 (5.9%) platform gave students a choice if they wish to participate in the peer assessment process, where they may choose to *decline review* or *request review* in the platform (i.e., Crowd Grader). It is also important to note that students who wish to participate in those reviews were incentivised through grades. Finally, 1 (5.9%) platform only supported instructor matching since it was designed for team member rating, and it could happen that students had already formed the group outside the system (i.e., TEAMMATES).

The variety of new approaches in matching assessors and assessees in peer assessment provides the chance for instructors to match the students based on several parameters. This is particularly useful for courses with a high number of students. While instructor matching is a "time-tested approach" of matching students and system generated matching is a "newer approach" of matching students, future studies should explore how these two approaches affect student interaction in peer assessment outcomes.

8.5   Conclusions

In this chapter, we investigated the features of 17 web-based peer assessment platforms to determine how they can potentially affect learning, students' feedback exchange, and the social interaction. We used nine peer assessment design elements from Adachi et al.'s (2018a) framework. Overall, we deem that the majority of the analyzed platforms offered features in support of students' learning, to generate positive feedback exchange between assessors and assessees, and a productive social interaction between students, but all depend on the configuration chosen by the instructor. The question of whether some features are helpful or detrimental is beyond the scope of this study. However, we provided a set of categories that researchers and instructors may use to further examine platform features. Also, these features will be put to waste if students and instructors do not receive ample training on how to use and take advantage of the features embedded in these platforms along with training on peer assessment itself (Panadero et al., 2016; Panadero & Brown, 2017). Regarding the platforms and as we mentioned earlier, students should be trained on how to provide and process feedback, while instructors should also be onboarded on how to plan and properly harness the features embedded in the peer assessment platforms. In sum, there is large potential for web-based peer assessment platforms in having a significant impact on students' peer assessment and academic performance, and on facilitating instructors' implementation of peer assessment. Researchers, instructors,

educational technologists, and programmers should work together to seamlessly integrate web-based peer assessment platforms in more settings to cater to different courses and educational contexts.

References

Adachi, C., Tai, J., & Dawson, P. (2018a). A framework for designing, implementing, communicating and researching peer assessment. *Higher Education Research and Development, 37*(3), 453–467. https://doi.org/10.1080/07294360.2017.1405913

Adachi, C., Tai, J.H.-M., & Dawson, P. (2018b). Academics' perceptions of the benefits and challenges of self and peer assessment in higher education. *Assessment and Evaluation in Higher Education, 43*(2), 294–306. https://doi.org/10.1080/02602938.2017.1339775

Babik, D., Gehringer, E., Kidd, J., Pramudianto, F., & Tinapple, D. (2016). Probing the landscape: Toward a systematic taxonomy of online peer assessment systems in education. In *CSPRED 2016: Workshop on Computer-Supported Peer Review in Education.* https://digitalcommons.odu.edu/teachinglearning_fac_pubs/22

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Boud, D. (2013). *Enhancing learning through self-assessment*. Routledge. https://doi.org/10.4324/9781315041520

Bouzidi, L., & Jaillet, A. (2009). Can online peer assessment be trusted? *Journal of Educational Technology and Society*, *12*(4), 257–268. http://www.jstor.org/stable/jeductechsoci.12.4.257

Chen, T. (2016). Technology-supported peer feedback in ESL/EFL writing classes: A research synthesis. *Computer Assisted Language Learning, 29*(2), 365–397. https://doi.org/10.1080/09588221.2014.960942

Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers and Education, 48*(3), 409–426. https://doi.org/10.1016/j.compedu.2005.02.004

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education, 24*(3), 331–350. https://doi.org/10.1080/03075079912331379935

Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review, 32*(2), 481–509. https://doi.org/10.1007/s10648-019-09510-3

Filius, R. M., Kleijn, R. A. M., Uijl, S. G., Prins, F. J., Rijen, H. V. M., & Grobbee, D. E. (2019). Audio peer feedback to promote deep learning in online education. *Journal of Computer Assisted Learning, 35*(5), 607–619. https://doi.org/10.1111/jcal.12363

Fu, Q.-K., Lin, C.-J., & Hwang, G.-J. (2019). Research trends and applications of technology-supported peer assessment: A review of selected journal publications from 2007 to 2016. *Journal of Computers in Education, 6*(2), 191–213. https://doi.org/10.1007/s40692-019-00131-x

Gielen, M., & De Wever, B. (2015). Scripting the role of assessor and assessee in peer assessment in a wiki environment: Impact on peer feedback quality and product improvement. *Computers and Education, 88*, 370–386. https://doi.org/10.1016/j.compedu.2015.07.012

Gielen, S., Dochy, F., Onghena, P., Struyven, K., & Smeets, S. (2011). Goals of peer assessment and their associated quality concepts. *Studies in Higher Education, 36*(6), 719–735. https://doi.org/10.1080/03075071003759037

Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007). Visualization of participation: Does it contribute to successful computer-supported collaborative learning? *Computers and Education, 49*(4), 1037–1065. https://doi.org/10.1016/j.compedu.2006.01.004

Kaufman, J. H., & Schunn, C. D. (2011). Students' perceptions about peer assessment for writing: Their origin and impact on revision work. *Instructional Science, 39*(3), 387–406. https://doi.org/10.1007/s11251-010-9133-6

Lai, C.-L., & Hwang, G.-J. (2015). An interactive peer-assessment criteria development approach to improving students' art design performance using handheld devices. *Computers and Education, 85*, 149–159. https://doi.org/10.1016/j.compedu.2015.02.011

Li, L. (2017). The role of anonymity in peer assessment. *Assessment and Evaluation in Higher Education, 42*(4), 645–656. https://doi.org/10.1080/02602938.2016.1174766

Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment and Evaluation in Higher Education, 45*(2), 193–211. https://doi.org/10.1080/02602938.2019.1620679

Lipnevich, A. A., & Panadero, E. (2021). A review of feedback models and theories: Descriptions, definitions, and conclusions. *Frontiers in Education, 6*. https://www.frontiersin.org/article/10.3389/feduc.2021.720195

Lu, R., & Bol, L. (2007). A comparison of anonymous versus Identifiable e-Peer review on college student writing performance and the extent of critical feedback. *Journal of Interactive Online Learning, 6*(2), 100–115.

Luxton-Reilly, A. (2009). A systematic review of tools that support peer assessment. *Computer Science Education, 19*(4), 209–232. https://doi.org/10.1080/08993400903384844

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In *Handbook of Human and Social Conditions in Assessment* (pp. 247–266). Routledge.

Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment and Evaluation in Higher Education, 44*(8), 1253–1278. https://doi.org/10.1080/02602938.2019.1600186

Panadero, E., & Brown, G. T. L. (2017). Teachers' reasons for using peer assessment: Positive experience predicts use. *European Journal of Psychology of Education, 32*(1), 133–156. https://doi.org/10.1007/s10212-015-0282-5

Panadero, E., Jonsson, A., & Strijbos, J.-W. (2016). Scaffolding self-regulated learning through self-assessment and peer assessment: Guidelines for classroom implementation. In D. Laveault & L. Allal (Eds.), *Assessment for Learning: Meeting the challenge of implementation* (pp.311–326). Springer.

Patchan, M. M., Schunn, C. D., & Clark, R. J. (2018). Accountability in peer assessment: Examining the effects of reviewing grades on peer ratings and peer feedback. *Studies in Higher Education, 43*(12), 2263–2278. https://doi.org/10.1080/03075079.2017.1320374

Patchan, M. M., Schunn, C. D., & Correnti, R. J. (2016). The nature of feedback: How peer feedback features affect students' implementation rate and quality of revisions. *Journal of Educational Psychology, 108*(8), 1098–1120. https://doi.org/10.1037/edu0000103

Phielix, C., Prins, F. J., & Kirschner, P. A. (2010). Awareness of group performance in a CSCL-environment: Effects of peer feedback and reflection. *Computers in Human Behavior, 26*(2), 151–161. https://doi.org/10.1016/j.chb.2009.10.011

Phielix, C., Prins, F. J., Kirschner, P. A., Erkens, G., & Jaspers, J. (2011). Group awareness of social and cognitive performance in a CSCL environment: Effects of a peer feedback and reflection tool. *Computers in Human Behavior, 27*(3), 1087–1102. https://doi.org/10.1016/j.chb.2010.06.024

Raes, A., Vanderhoven, E., & Schellens, T. (2015). Increasing anonymity in peer assessment by using classroom response technology within face-to-face higher education. *Studies in Higher Education, 40*(1), 178–193. https://doi.org/10.1080/03075079.2013.823930

Reynolds, J., & Russell, V. (2008). Can you hear us now? A comparison of peer review quality when students give audio versus written feedback. *The WAC Journal*, *19*(1), 29–44. https://doi.org/10.37514/WAC-J.2008.19.1.03

Rotsaert, T., Panadero, E., & Schellens, T. (2018). Anonymity as an instructional scaffold in peer assessment: Its effects on peer feedback quality and evolution in students' perceptions about peer assessment skills. *European Journal of Psychology of Education, 33*(1), 75–99. https://doi.org/10.1007/s10212-017-0339-8

Søndergaard, H., & Mulder, R. A. (2012). Collaborative learning through formative peer review: Pedagogy, programs and potential. *Computer Science Education, 22*(4), 343–367. https://doi.org/10.1080/08993408.2012.728041

Strijbos, J. -W., Ochoa, T. A., Sluijsmans, D. M. A., Segers, M. S. R., & Tillema, H. H. (2009). Fostering interactivity through formative peer assessment in (web-based) collaborative learning environments: In C. Mourlas, N. Tsianos, & P. Germanakos (Eds.), *Cognitive and emotional processes in web-based Education* (pp. 375–395). IGI Global. https://doi.org/10.4018/978-1-60566-392-0.ch018

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3

To, J., & Panadero, E. (2019). Peer assessment effects on the self-assessment process of first-year undergraduates. *Assessment and Evaluation in Higher Education, 44*(6), 920–932. https://doi.org/10.1080/02602938.2018.1548559

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276. https://doi.org/10.3102/00346543068003249

Topping, K. (2021). Peer assessment: Channels of operation. *Education Sciences, 11*(3), 91. https://doi.org/10.3390/educsci11030091

van den Berg, I., Admiraal, W., & Pilot, A. (2006). Design principles and outcomes of peer assessment in higher education. *Studies in Higher Education, 31*(3), 341–356. https://doi.org/10.1080/03075070600680836

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review, 4*(1), 41–54. https://doi.org/10.1016/j.edurev.2008.11.002

Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers and Education, 81*, 123–132. https://doi.org/10.1016/j.compedu.2014.10.001

Voet, M., Gielen, M., Boelens, R., & De Wever, B. (2018). Using feedback requests to actively involve assessees in peer assessment: Effects on the assessor's feedback content and assessee's agreement with feedback. *European Journal of Psychology of Education, 33*(1), 145–164. https://doi.org/10.1007/s10212-017-0345-x

Wang, X. -M., Hwang, G. -J., Liang, Z. -Y., & Wang, H. -Y. (2017). Enhancing students' computer programming performances, critical thinking awareness and attitudes towards programming: An online peer-assessment attempt. *Journal of Educational Technology and Society*, *20*(4), 58–68. https://www.jstor.org/stable/26229205

Wen, M. L., & Tsai, C.-C. (2008). Online peer assessment in an inservice science and mathematics teacher education course. *Teaching in Higher Education, 13*(1), 55–67. https://doi.org/10.1080/13562510701794050

Zheng, L., Zhang, X., & Cui, P. (2020). The role of technology-facilitated peer assessment and supporting strategies: A meta-analysis. *Assessment and Evaluation in Higher Education*, *45*(3), 372–386. https://doi.org/10.1080/02602938.2019.1644603

Feed-Back About the Collaboration Process from a Group Awareness Tool. Potential Boundary Conditions for Effective Regulation

9

Sebastian Strauß and Nikol Rummel

9.1   Introduction

Collaborating with peers can be an effective arrangement for learning, however, research has shown that we cannot expect all learners to collaborate effectively without further support (Rummel & Spada, 2005). Among the core challenges for collaborative learning are coordination and regulation of the group's interaction. Research on computer-supported collaborative learning (CSCL) has investigated different means of scaffolding in order to support groups. In this chapter, we focus on social interaction in groups and present social group awareness tools as a means to support groups in regulating their interaction processes. We conceptualize group awareness tools as sources for feed-back because they provide groups with information regarding the interaction between their members. Groups can then use this information (i.e., feedback) to adapt the interaction between their group members, that is, improve the future interaction in the group. While previous studies have accumulated evidence for the effectiveness of group awareness tools (Janssen & Bodemer, 2013), the mechanisms behind their effectiveness are not yet well-understood, and a framework of the respective mechanisms is lacking.

Unlike other means of collaboration support such as collaboration scripts (Kollar et al., 2018), group awareness tools provide groups with feed-*back* on their past performance or interaction without explicitly suggesting potential regulatory

S. Strauß (✉) · N. Rummel

Institute for Educational Research, Ruhr-University Bochum, Bochum, Germany
e-mail: sebastian.strauss@rub.de

183

actions (i.e., feed-*forward*). Similar to using instructional feed-back or peer feed-back, students need to actively engage with the feed-back from group awareness tools to benefit from it (Lipnevich & Panadero, 2021). In line with this assumption, Janssen et al. (2011) found that the amount of time that groups attended to the group awareness tool, affected the effect of the tool on the distribution of participation in the group. Other studies found that groups require additional help in interpreting the feed-back by the group awareness tool (e.g., Jermann & Dillenbourg, 2008) or that some groups may require additional support in deriving effective regulatory actions (Dehler et al., 2009; Strauß & Rummel, 2021b). With this in mind, we seek to identify potential boundary conditions that may help or prevent groups from leveraging the feed-back from group awareness tools. Afterwards, we review previous studies on group awareness tools and present two small-scale field experiments from our own research in which we explored the processes of feed-up, feed-back and feed-forward. We conclude this chapter by discussing potential factors that may affect whether will a group is motivated and is able to leverage the feed-back provided by a group awareness tool effectively.

9.2 Supporting Collaboration with Group Awareness Tools

9.2.1 Feed-Back on Interaction: How Group Awareness Tools Guide Collaboration

Collaborative learning refers to "*a situation in which two or more people learn or attempt to learn something together*" (Dillenbourg, 1999, p. 1, emphasis in original), or more specifically "[…] a coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem" (Roschelle & Teasley, 1995, p. 70). Years of research have shown the benefits of collaboration for domain-specific knowledge as well as for collaboration skills (Chen et al., 2018; Hattie, 2009; Jeong et al., 2019; Pai et al., 2015; Tenenbaum et al., 2020).

The effectiveness of collaboration for learning stems from productive interaction between the group members. This includes interactions that serve processing of information to solve the joint task (e.g., giving explanations, cognitive modeling, see King, 2007), processes that allow a group to monitor and regulate collaboration processes, as well as the group members' motivation and affective states (Järvelä et al., 2016; Kirschner et al., 2015). Collaborating with others increases the demands for regulation because the members of a group not only need to regulate their own learning (self-regulated learning, SRL), but, in addition, the group members need to support each other during their regulation (co-regulation), and all members of the group explicitly need to align their perception and regulate as a group (socially-shared regulation) (Järvelä et al., 2016).

Soller et al. (2005) proposed a model of how groups regulate their collaboration and how technology can scaffold this regulation. Their proposed model draws on the cybernetic idea of homeostasis (Umpleby & Dent, 1999; Wiener, 1949) which

conceptualizes a group as a system that seeks to achieve an equilibrium. The group reacts to imbalance (disequilibrium), that is, a difference between the current state and a desired goal-state, with regulation. This regulation aims at returning the system to an equilibrium. According to the model by Soller et al. (2005), regulation of collaboration occurs in five phases. During the first phase, the group collects data on the current state of a relevant aspect of the system, such as information on the participation of the individual group members. In the second phase, the group develops a model of the interaction by aggregating the data into indicators that characterize the current state of the collaboration in terms of the desired aspect (e.g., the distribution of participation). In the third phase, the group uses these indicators to compare the current state with a desired goal-state (e.g., eqtrauual participation). The desired goal-state can be set by the group itself (descriptive collaboration norm) or by an external agent such as a teacher (prescriptive collaboration norm). In the fourth phase, the group is expected to regulate if an imbalance had been detected, that is, if the current state and the desired state differ. For example, a group could redistribute tasks so that all group members participate. In the fifth and final phase, the group evaluates the success of the regulation, that is, if the regulatory action restored the equilibrium and the desired-goal state is achieved. If this is not the case, the group will repeat the cycle until it achieves an equilibrium.

Given the central role of interaction for the effectiveness of collaboration and collaborative learning it is important to note that fostering the regulating of the interaction can be a target of instructional support (see Meier et al., 2007; Rummel, 2018). Regulation on this social plane of collaboration requires information about past and current states of the interaction in the group (i.e., feed-back), for example information about the knowledge and skills of the other group members, or on who is currently working on which part of the joint task. Once gathered, this information (i.e., feed-back) can serve for the group as a basis to coordinate their interaction and improve the quality of their interaction. The notion of gathering information about the actions of the other team members can be found in the concept of *group awareness* (Schnaubert & Bodemer, 2022) which Dourish and Bellotti (1992) defined as "an understanding of the activities of others, which provides a context for [one's] own activity" (Dourish & Bellotti, 1992, p. 107). If the intention of the information is to increase a teams' performance, small group researchers refer to it as "team feed-back" while if the feed-back focuses on processes or psychological states in the team is termed "team mediator feed-back"(Handke et al., 2022).

The concept of group awareness has been introduced to the field of CSCL (Schnaubert & Bodemer, 2022) where it lead to the development of so-called group awareness tools (GATs) (Bodemer et al., 2018). These tools collect data from the collaboration environment (e.g., keystrokes, logged actions, self-reports) and visualize the data for the group (Buder, 2011). Research on CSCL has investigated different types of GATs. While *cognitive group awareness tools* visualize different aspects of the knowledge that is available in the group (e.g., Dehler et al., 2011; Engelmann & Hesse, 2011; Ollesch et al., 2021), *social group awareness tools*

provide information about processes and states of group members such as participation (Bachour et al., 2010; Janssen et al., 2011; Ollesch et al., 2021; Strauß & Rummel, 2021b), how much information has been shared (Kimmerle & Cress, 2009) or how the members of a group perceive each other (Phielix et al., 2011). In this chapter, we focus on social group awareness tools. Specifically, group awareness tools that provide groups with information about the interaction within the group, for example by visualizing the distribution of participation (i.e., the result of individual participation during collaboration).

GATs provide groups with a visualization which can be characterized as team (mediator) feed-*back* (Handke et al., 2022), that is, information regarding past performance or the current state of the collaboration. The group can utilize this feed-back to improve the group's performance, that is, the quality of the interaction in the group (Carless & Boud, 2018; Handke et al., 2022; Hattie & Timperley, 2007; Lipnevich & Panadero, 2021). The information provided by a GAT does not contain information about potential desired goal-states (feed-*up*), or guidance regarding potentially helpful strategies (feed-*forward*). In this regard, unlike other means of directive collaboration support such as collaboration scripts (Kollar et al., 2018), GATs provide "tacit guidance" (Bodemer, 2011). Based on research on students' use of feed-back for learning (Lipnevich & Panadero, 2021; Winstone et al., 2017), we assume that the learners of a group also need to take an active role and process the information from a GAT, in order to determine whether regulation is necessary and what actions may help them achieve a more desirable state.

Thus far, research on GATs has not provided a comprehensive framework that specifies how GATs support collaboration. Therefore, we will briefly summarize potential mechanisms that are mentioned throughout the GAT literature.

First, a GAT visualizes information for the group, such as the distribution of participation. This visualization makes a particular aspect of the collaboration more salient and thus draws the learners' attention to it (Bachour et al., 2010; Carless & Boud, 2018; Pea, 2004). Thus, it is expected that the GAT increases the possibility that a group focuses its efforts on regulating this aspect. Second, group awareness information serves as negative feed-back for a group (Jermann & Dillenbourg, 2008) which allows the group to assess whether and to which degree the current state of the collaboration deviates from a desired goal-state. A discrepancy between the current state of the collaboration and the desired goal-state may then lead to a reflection process within the group which can eventually trigger regulation. Especially continuous feed-back can be expected to facilitate monitoring the progress towards the desired goal-state (e.g., Carless & Boud, 2018; Soller et al., 2005; Webb & de Bruin, 2020) which helps groups regulate their collaboration (Harkin et al., 2016; Webb & de Bruin, 2020).

Finally, a graphical representation of the group members' behavior makes the individual group members more visible and increases individual accountability which has been shown to be an important predictor of effective collaboration (Handke et al., 2022; Johnson & Johnson, 2009). Feed-back regarding participation also reduces learners' uncertainty about their peers' activity and thus supports

trust (Robert, 2020; Walther & Bunz, 2005), may reduce social loafing (e.g., Johnson & Johnson, 2009; Price et al., 2006), and promote social comparison which motivates group members to contribute (Michinov & Primois, 2005). Despite a lack of a comprehensive framework that covers the mechanisms behind GATs, there are several studies that investigated their effects on collaboration which we will summarize in the following section.

9.2.2   Prior Research on Social Group Awareness Tools

Social GATs provide groups with information regarding the functioning of the group, that is, the behavior of the group members, their presence and their perception of the group (Bodemer & Dehler, 2011; Janssen & Bodemer, 2013). While previous studies did not find positive effects on group performance when a GAT visualized the distribution of participation (Janssen et al., 2007a, 2007b, 2011) (cf. Jongsawat & Premchaiswadi, 2009 for a contrary result), research found positive effects of GATs on the collaboration *processes*. For example, when receiving a GAT that visualized participation, group members authored longer dialogue acts (Janssen et al., 2007a, 2007b; Kimmerle & Cress, 2009; Kimmerle et al., 2007; Lin & Tsai, 2016) (cf. Jermann & Dillenbourg, 2008; Jongsawat & Premchaiswadi, 2009 for different results), showed more coordination of social activities (Janssen et al., 2007a, 2007b), or reported higher group cohesion (Leshed et al., 2009) than groups without the GAT. Interestingly, research did not find direct effects of GATs that provide information about the participation on the distribution of participation (Janssen et al., 2007a, 2007b, 2011; Strauß & Rummel, 2021b). Instead, this effect is mediated by the time that the students in the group have the GAT open (Janssen et al., 2011). Similarly, Bachour et al. (2010) found that groups achieved a more equal participation when the group members perceived equal participation as important. These results suggest that groups can leverage feed-back on their interaction and adapt their collaboration. However, as mentioned above, research on feed-back has shown that simply providing students with access to feedback does not guarantee positive effects (Lipnevich & Panadero, 2021; Winstone et al., 2017). In line with this circumstance, studies on social GATs suggest not all groups appear to benefit from a GAT (Dehler et al., 2009) and may require additional guidance (Clarebout & Elen, 2006; Janssen et al., 2011). In our own research, we therefore investigated whether additional explicit support helps groups activate adequate regulation strategies. We offered groups a combination of a GAT and adaptive collaboration prompts that both targeted regulating the distribution of participation in the group (Strauß & Rummel, 2021b). Our analyses revealed that the distribution of participation became more even over time but that groups that received the combination of a GAT and prompts did not achieve a significantly more even distribution of participation. Exploring students' perceptions of the support further suggested that students rather used the GAT to regulate their own participation instead of discussing the distribution of participation with the group. In addition, students reported that the feed-back from the GAT was useful

but that it was difficult to regulate on a group's level. These results lead us to the general question of boundary conditions for social GATs. The results of our study specifically highlighted two questions. The first question concerned, whether students require a dedicated opportunity to process the information displayed by the GAT with the goal to assess whether regulation is required and how this can be achieved. Second, the results call into question whether using the number of words to operationalize participation and using this metric for the GAT may fall short to capture the phenomenon of participation. If the GAT does not provide groups with a useful indicator, they may struggle with taking up the feed-back and translating it into productive interaction. To shed some light on these questions we conducted two small-scale field experiments which we summarize below.

9.3   Our Research: Scaffolding Collaborative Reflection and Using Self-reports to Assess Participation

In this section, we present two small-scale field experiments and summarize the central findings. These two field experiments were based on findings from an earlier study (Strauß & Rummel, 2021b) and explored two hypotheses concerning potential boundary conditions for the effectiveness of a social GAT. The first field experiment addressed the question whether groups benefit from additional guidance for feed-back take-up and reflection, the second experiment explored the effects of two different data sources for the GAT, that is, a system-generated indicator of participation (number of words) and a peer-generated indicator (self-report of own participation).

A premise underlying our studies was that equal participation is crucial for collaborative learning as the effectiveness of collaboration for learning and problem solving is based on interaction between the members of a group. Productive interactions are less likely to occur if only a few group members actively participate in the collaboration. As a result, less active group members will benefit less from the collaboration. As outlined earlier, effective collaboration requires interaction to serve goals such as achieving a shared understanding (Baker et al., 1999; Clark & Brennan, 1991), pooling unshared information (Deiglmayr & Spada, 2011; Stasser & Titus, 1985) or regulating the interaction (Järvelä et al., 2016; Panadero & Järvelä, 2015). If not all learners participate evenly in these processes, a group may not achieve its goal, for example because not all group members shared their information which were required for finding a good solution to the joint problem. In addition, studies report that learners experience frustration when not all members of their group are actively contributing to the joint task (Strauß & Rummel, 2021a) as well as dissatisfaction with the collaboration the more the participation is unevenly distributed in the group (Strauß & Rummel, 2021b).

Thus, we sought to support groups in regulating the distribution of participation. With regard to facilitating monitoring and regulation of collaboration, GATs have been used in the past, also for fostering the regulation of participation (Janssen et al., 2011; Jermann & Dillenbourg, 2008; Strauß & Rummel, 2021b). The notion

underlying a GAT that visualizes the current distribution of participation to the group is that the group can take up this feed-back and compare the current distribution of participation to a desired distribution. Previous research showed that an uneven distribution of participation is a source of frustration for students (for an overview see Strauß & Rummel, 2021a). Hence, it can be assumed that students aim to achieve an equal distribution of participation, thus trying to regulate their interaction in a way that all group members contribute equally. However, most studies did not find direct effects of a GAT that displays group members' participation on the distribution of participation in the group. In a recent study we investigated whether groups may benefit from explicit guidance (i.e., adaptive prompts) in addition to the tacit guidance of a GAT (Strauß & Rummel, 2021b). The results of our study left open whether a combination of collaboration prompts and a GAT helps groups to regulate the distribution of participation. Exploratory analyses of students' use and perception of the support indicated that groups may need additional support for leveraging the feedback provided by the GAT, instead of explicit guidance regarding which actions may be useful given an uneven distribution of participation. Further, while using the number of words as an indicator for participation in an online environment is widespread, we acknowledge Hrastinski's (2008) argument that operationalizing participation solely as the number of words that each group member had contributed may be an incomplete view of what participation includes. Against this background, we derived two small field experiments that explored the effect of additional guidance that targets the process of taking up the information from the GAT and reflecting on them, and that explored the effect of a more holistic operationalization of participation.

Both experiments were conducted in an online course for university students that went over fourteen weeks. On the university's Moodle, students could access the learning materials for each of the six course topics (two weeks per topic), such as a lecture video, literature, and a quiz. During each topic, students worked in small groups to solve a collaborative task and create a joint answer text. Groups used a private group forum on Moodle for coordination and a private wiki to formulate their answer to the collaborative task.

Both studies took place during one course topic (i.e., two weeks), during which the students collaborated in groups of four to solve a collaborative task. In total, 104 students enrolled in the course and 84 (80.8%) agreed to participate in the study for monetary reward. During collaboration, groups received a group awareness tool which was constantly available on every page of the learning environment (main page, group's discussion forum, the group's wiki) on the right-hand margin of the Moodle environment.

To engage students with the GAT we followed the guidelines offered by Wise (2014) and Wise and Vytasek (2017) on how to implement learning analytics interventions in learning settings. First, the analytics should be integrated into the course and students should understand the goal of the analytics. Specifically, students need to be aware of the pedagogical goal of the current learning activity, understand what is considered effective engagement in this activity, and learn how the analytics help them monitor productive activity. We offered this information to

the students in a familiarization message that explained the tool's pedagogical intent, that is, that active and equal participation during collaborative assignments is important for successful problem solving. Further, we explained that the GAT provided an up-to-date visual representation of the current distribution of participation.

Second, learners should be free to interpret the analytics and choose regulatory behavior. Specifically, learners should be able to set goals individually and assess whether they were able to attain them. In our studies, we implemented a collaborative reflection activity (see below) which required students to set goals, monitor their progress towards these goals and decide whether regulation of the collaboration was necessary.

A third aspect that is expected to enhance learning analytics interventions is that students need a frame of reference that helps them interpret the analytics. In our studies, this frame of reference was created during the collaborative reflection activity, that is, the general instruction (i.e., being active and achieving equal participation) and by the goals set by the individual groups.

Finally, learners should have the freedom and opportunities to negotiate the analytics, either with the teacher or with peers. This was a core aspect in our studies as the students worked in groups and received feed-back on their collaboration, which they could freely view and discuss.

9.4   Field Study 1: Collaborative Reflection to Scaffold Feed-Up, Feed-Back, and Feed-Forward

In the first small-scale field experiment, we implemented a GAT together with collaborative reflection activity as part of the regular group task, which was expected to help groups actively engage with the feed-back from the GAT. To learn more about the effects of additional co-reflection, we compared groups that only received a GAT with groups that received the GAT and also collaboratively processed the information provided by the tool.

From a theoretical point of view, one important step during regulation is reflection (Butler & Winne, 1995). In collaborative settings, peers can serve as resources for critically questioning experiences and developing alternative perspectives (Kori et al., 2014). Yukawa (2006) defined of collaborative reflection (co-reflection) as "cognitive and affective interactions between two or more individuals who explore their experiences in order to reach new intersubjective understandings and appreciation" (Yukawa, 2006, p. 206). Gabelica et al. (2014) refer to this concept as "team reflexivity". A reflective team exhibits three behaviors. First, the team uses feed-back to evaluate the group's past performance, for example by collectively discussing the performance on a joint task. Second, the team searches for alternative ways to perform such a task in the future, and eventually, the team arrives at a shared decision on which strategies should be enacted in the future (Gabelica et al., 2014).

Given that reflection has been conceptualized as a key process during regulation, as well as based on findings that emphasize that a designated phase for reflection benefit the collaboration, we expected that providing groups with a collaborative reflection activity helps groups make use of the feed-back from a GAT. For our studies, we adapted the co-reflection activity from Phielix et al. (2011) who designed their co-reflection activity by integrating the suggestions by Hattie and Timperley (2007). This activity tasked the groups to clarify their goals of the current activity (feed-up), decide whether progress is being made towards this goal (feed-back), and eventually decide which activities are needed to progress towards the goal (feed-forward).

9.4.1   Sample, Materials, Procedure, Measures

We conducted the field experiment in a university online course. In total, 104 students enrolled in the course, of which 84 (80.8%) agreed to participate in the study. The study took place during the fourth topic of the course (i.e., week six of the course). By the time of the data collection, 51 participants (58.6% of the initial sample; 64.7% female; age: M = 24.00; SD = 3.45) were still active in the course. During the collaborative task, students collaborated for two weeks in groups of four. All groups received a GAT that visualized the participation of each group member as a bar graph (see Fig. 9.1).

Each bar in the GAT represented the number of words that each group member had contributed (group's forum and wiki) and was updated automatically whenever a student submitted a new contribution. A legend below the GAT identified the individual group members. On mouse-over, the GAT displayed the absolute number of words for each group member. Through a collapsible text box below the GAT students could access a brief explanation of the GAT like the one they had received in the familiarization email. In addition, students could view the deadline of the current task and set individual to-dos by clicking on the buttons above the bar graph.

For the experiment, the students were randomly assigned to one of two conditions. Twenty-six students (six groups) received only the GAT during collaboration, while the remaining 25 students (six groups) received the GAT during collaboration and additionally performed the co-reflection activity.

At the beginning of the study, students in both conditions received a familiarization message which informed them about the role of active participation and how the GAT can assist them achieving equal participation. Afterwards, students worked on the collaborative task for two weeks. At the end of the first week, half of the groups performed the collaborative reflection activity.

The reflection activity was designed similar to the one presented by Phielix et al. (2011). We implemented the process of feed-up, feed-back and feed-forward in the form of four questions that students answered in Moodle: (1) "In your opinion: How should participation be distributed during collaboration in a team like yours? Explain." *(feed-up, goal-setting)* (2) "Take a look at the visualization: How

Fig. 9.1 Group awareness tool displaying the number of words of a fictitious group

well is the participation in your team currently distributed? Give a rating (scale 1 (bad) – 5 (good)) and explain your rating." *(feed-back, reflection)* (3) "Examine the visualization again and post your rating of the current participation into the forum. Discuss together the ratings of the team members and agree on a rating." *(feedback, reflection)* (4) "Is it necessary to change the way you participate? Develop a plan and set specific goals for your team regarding the distribution of participation (Who? What? When?). Write down your plan in the Etherpad" *(feed-forward, goal-setting)*. Students answered the first two questions individually to prepare for the following co-reflection and subsequently answered the last two questions collaboratively in the group's discussion forum.

Over the weekend of the first week, the students in each group (1) individually set a goal for the distribution of participation in their group, and (2) individually reflected on the current distribution of participation as displayed by the GAT. At the beginning of the second week, the members of each group negotiated (3) whether

regulation was necessary, and (4) how they can regulate their collaboration in terms of the distribution of participation.

To assess the *distribution of participation*, we used two measures. First, we calculated the gini-coefficient based on the number of words that each group member had contributed to the group's discussion forum and the group's wiki where they created a text that included the solution to the collaborative problem. The gini-coefficient uses the number of words that each group member had contributed and returns a value that represents the distribution participation for this group. This coefficient represents the distribution of participation for each group as a value ranging from 0 (perfect balance) and 1 (perfect imbalance) (Dorfman, 1979).

Second, to acknowledge students' perception of the participation, we assessed perceived social loafing by asking students to rate how the participation was distributed during the collaborative task (-5, only one group member contributed; $+5$, every group member contributed equally) (Aggarwal & O'Brien, 2008). As a proxy for *engagement with the GAT*, we asked students to indicate how frequently they had looked at the GAT on average during the two weeks of the collaborative task.

9.4.2 Results

Our manipulation check indicated that students complied with the co-reflection activity. Specifically, students who co-reflected contributed significantly more words to their group's forum (M = 256.57; SD = 176.91) than students who only received a GAT (M = 79.76; SD = 68.85; U = 127.00, Z = $-$ 3.73, $p <$ 0.05), and also reported that they used the GAT to regulate the collaboration more intensely (M = 3.00; SD = 1.18) than their counterparts (M = 1.82; SD = 0.81; U = 77.50, Z = $-$ 3.01, $p <$ 0.05). However, against our assumptions, students who performed the co-reflection activity (M = 6.33; SD = 4.51) did not report having looked at the GAT more frequently than students in the GAT condition (M = 5.76; SD = 2.71; U = 176.00; Z = $-$ 0.07, $p >$ 0.05).

We hypothesized that the additional co-reflection activity helps groups achieve a more even distribution of participation. Our analyses revealed tentative evidence for this hypothesis, as the 17 students in the GAT condition rated the distribution neither as unevenly distributed nor evenly distributed (M = 0.94; SD = 3.09), while the 21 students who performed the additional co-reflection reported that the participation was more evenly distributed, as indicated by a larger positive value (M = 1.81; SD = 2.94). While this difference in means pointed into the hypothesized direction, it was not statistically significant (U = 140.00, Z = $-$ 1.14, $p >$ 0.05). Further, we analyzed the distribution of the number of words that the students had contributed. Since the gini-coefficient is calculated for each group (i.e., the level of analysis are the groups of students, not the individual students), the number of cases that enter the analysis decreases. Since the remaining sample of 12 groups does not allow for inferential statistics, we report descriptive statistics (see Table 9.1). Groups in the two conditions only differed slightly in terms of the

Table 9.1 Distribution of participation in both conditions

Gini-coefficients	GAT (n = 6 groups)			GAT + Co-reflection (n = 6 groups)		
	M (SD)	Min	Max	M (SD)	Min	Max
Total number of words	0.45 (0.28)	0.11	0.93	0.42 (0.17)	0.15	0.56
Word count group wiki	0.55 (0.23)	0.37	1.00	0.43 (0.17)	0.15	0.60
Word count group forum	0.53 (0.19)	0.32	0.83	0.49 (0.14)	0.35	0.69

total number of words (i.e., contributions in the group's wiki and forum combined). On average, groups in both conditions reached a rather even distribution of overall participation as indicated by gini-coefficients below 0.5.

Carefully inspecting the distribution of participation, we found that there were groups in both conditions that achieved an almost perfect balance of participation (i.e., minimal values close to 0), as well as groups that did not achieve an even distribution of participation (i.e., maximum values tending towards 1). One group achieved a gini-coefficient of 1 as only one group member had contributed. It is important to note that this group may have been an outlier as the group with the next lower value yielded a gini-coefficient of 0.64. In comparison, the least successful groups that performed the co-reflection achieved a more even distribution of participation. Overall, the groups in this condition reached lower minima and maxima which indicates a more even distribution of participation. In sum, our data indicate a trend that is congruent with our expectation that groups would benefit from a collaborative reflection, however, our results were not statistically significant.

In a subsequent step, we explored the answers that students provided during the individual goal-setting activity (step 1) of the co-reflection task to learn more about students' collaboration norms. Therefore, we coded the answers regarding the optimal distribution of participation that the 25 students provided during the individual part of the co-reflection activity. During coding, we assigned a label to each response, grouped similar responses, and eventually aggregated them along overarching themes.

The individual answers revealed that students generally valued equal participation. However, we identified four nuances of this collaboration norm. We used representative quotes from the students to name these nuances. We termed the first nuance "*The participation in a team should be evenly distributed*". Six students (24%) stated that all group members should contribute evenly to the joint task, however "minimal differences [in participation]" are still acceptable. One student reasoned that participation should be evenly distributed since the requirements for all students in the group are the same. Students mentioned no further boundary conditions or possible compromises.

We summarized the second nuance of this collaboration "*It's normal that not everyone contributes the exact same amount, but the proportions should be right*". Most students (n = 14; 56%) noted that the distribution of participation may differ

among the members of the group. Unlike students from the first category, students in this category included qualifiers such as "roughly" or "if possible". For example, one student proposed dividing the work equally by the number of group members: "Everyone should contribute a part to the task. We are four people so we should divide the workload roughly (!) by four and then look through the results together."

The third nuance can be summarized as "*Essentially, the distribution should be even, but…*". Students who fell into this category (n = 3; 12%) advocated equal participation but also specified boundary conditions. While "participation should be equal by default" and also "fair and just", multiple factors affect how evenly participation should be distributed. Students mentioned that the task, group members' capacities, inactive group members, as well as the remaining time until the deadline should be considered. In addition, group members should get the chance to work on tasks that they can excel at. Students argued that uneven participation would be acceptable if a group member signaled early enough that they will not be able to contribute their fair share. In this case, workload could be redistributed. Finally, students acknowledged that asynchronous tasks allow team members to work at their own pace which may lead to uneven participation during the process but should even out towards the deadline.

Finally, we termed the fourth nuance "[…] *it should become visible that every team member at least tried to contribute to the final result*". So far, most of the responses focused on the *amount* of participation. However, one student argued that "while the number of words does not indicate quality, a basic level of participation is required". Specifically, the student noted that "every participant should say something" and "while not everyone needs to perform exactly equally, or write, that is, it should become visible that every member of the team tries to participate and contribute to the final result," and had "[looked] into the topic". In other words, *any visible participation* by the group members is appreciated.

Discovering these nuances lead us to assume that not all groups may strive for an exact even distribution of participation. Further investigating students' collaboration norms help us understand under which conditions groups initiate regulation and which goal-state they aim for. For example, these collaboration norms may serve as mediating or moderating variables for regulation and explain differences in the degree to which groups are motivated to achieve an even distribution of participation.

To summarize, we conducted this first small-scale field experiment based on the assumption that groups may require additional guidance on how to engage with the information provided by GAT, instead of actional suggestions for effective regulation. In this first small-scale field experiment we investigated whether a collaborative reflection activity supports groups in leveraging the information from a GAT (i.e., information regarding the interaction in the group). Contrary to our expectations the results of our field experiment indicate that triggering co-reflection (i.e., a sequence of feed-up, feed-back and feed-forward) does not significantly affect the distribution of participation during online-collaboration. While descriptive trends point into the hypothesized direction, the data reported above need to

be interpreted with great care due to the limited sample size. In addition to comparing means between two experimental conditions, we further identified different collaboration norms that students may hold about the distribution of participation. We hypothesize that these different collaboration norms affect under which circumstances and to which goal-state the members of a group will regulate the distribution of participation within the group.

9.5   Field Study 2: Contrasting System-Generated Feed-Back and Peer-Generated Feed-Back

The second question that arose from our field experiment (Strauß & Rummel, 2021b) concerned the operationalization of participation. As discussed above, using the number of words contributed by each group member only captures one dimension of participation (Hrastinski, 2008). We explored this question with a second field experiment that we conducted in the same online course. Based on the promising results of the first study, we carefully assumed that groups benefit from a co-reflection activity and thus required students to answer the four reflection questions outlined above. To address the question of the operationalization of *participation during collaboration*, we developed a second version of the GAT that asked students to provide their peers with information about their own participation (i.e., peer-generated feed-back).

9.5.1   Using Peer-Generated Feed-Back to Include a More Holistic Operationalization of Participation

One potential limitation of the design of our earlier study (Strauß & Rummel, 2021b) was that we used the number of words as an indicator for participation during web-based collaboration. While this operationalization is common in research on e-learning and computer-mediated collaboration, Hrastinski (2008) argues that participation can be viewed more holistically. In his review, he identified six concepts of online learner participation: (1) Participation as accessing the e-learning environment, (2) participation as writing, (3) participation as quality writing, (4) participation as writing and reading, (5) participation as actual and perceived writing (i.e., a student makes contributions that are perceived as useful), (6) participation as taking part and joining in a dialogue. Further, he acknowledges that participation may also occur off-system (i.e., offline), for example when students research and read material, or make notes outside the e-learning environment. Importantly, some of these dimensions can be captured by computer systems (e.g., access to the collaboration environment, contributing words) while the remaining dimensions either require more complex computations (e.g., assessing quality writing, having read a contribution) or occur off-system and thus cannot be assessed automatically. If the indicator that is being used in the GAT does not suit the needs of a group, the group may not be able to assess the need for regulation.

For example, the number of words provides information regarding the quantity of participation but does not capture the quality of the contributions which may stem from a group member investing a lot of their time into working through the learning materials.

Against the background of Hrastinski's review, we explored the effect of incorporating a more holistic view of participation in the GAT. Since not all dimensions of participation can be captured through logged events from the learning management system Moodle, we decided to ask the members of the group to display their participation by filling in a short questionnaire on their participation during the collaborative task. Using self-reports as a data source for a GAT is more closely connected to the original idea of group members displaying important information to their peers in order to promote group awareness (Buder, 2011) and can be found to varying degrees in prior studies, for example as peer-assessment of social performance (Phielix et al., 2011), individual task perception (Hadwin et al., 2018) or meta-cognitive judgements (Schnaubert & Bodemer, 2019). Therefore, in this second field experiment, students provided the system with self-reports regarding their own participation which was then visualized in the GAT. Thus, the GAT included feed-back regarding the distribution of participation which consisted of students' perception of their own behavior. In our field experiment, we contrasted this source of feed-back with providing groups with the number of words that each group member had contributed (i.e., system-generated feed-back). Again, we assumed that groups may use the feed-back regarding the distribution of participation to the current distribution of participation with a desired distribution (i.e., equal participation).

9.5.2 Sample, Procedure, and Materials

The study was conducted in the same course in which we conducted the first field-experiment. This second study began in week eight of the course. By then, 50 participants (59.5% of the initial sample; age: M = 23.96; SD = 3.48) who agreed to participate and were still active in the course. Again, the participants were randomly assigned to one of two conditions. Twenty-three students (six groups) received a GAT that displayed the number of words that each group member contributed (system-generated feed-back). The remaining 27 students (seven groups) were asked to provide information on their own participation through a short questionnaire. This information was then visualized in the GAT (peer-generated feed-back).

Depending on the condition, the bars in the GAT represented the number of words that each group member had contributed (group's forum and wiki), or the results of the group members' self-reports, respectively. The GAT that visualized system-generated feed-back was identical to the one used in the previous field study. The GATs that visualized peer-generated feed-back as well as the pop-up for the participation questionnaire are shown in Fig. 9.2.

Fig. 9.2 GAT that visualizes peer-generated feed-back on participation (right) and pop-up for participation questionnaire (center) for a fictitious group

The bars updated automatically whenever a student posted a new contribution, or when a student filled in the participation-questionnaire. The participation-questionnaire was presented as a pop-up window in Moodle and contained three questions that the students rated on a 5-point Likert scale: (1) "I have been reading the posts of my team mates", (2) "I have been working on the team task by preparing contributions, reading or by thinking about the topics", (3) "I have contributed (both online and offline) in a way that brought my team forward". The participation-questionnaire was displayed every time a student logged into Moodle for the first time each day; and returned each time a student returned to the main course page. If a student had answered the questionnaire, it would not appear for the rest of the day. Students could always update their participation via a button on the GAT. As in the previous field experiment, groups performed the co-reflection activity after the first week of the collaborative task.

9.5.3   Results

The average distribution of the total number of words (gini-coefficient) within the six groups that received a GAT with system-generated feed-back was more equal (M = 0.34; SD = 0.25) than the distribution in groups that received peer-generated feed-back (M = 0.46; SD = 0.16). Again, due to the small sample size of six groups per condition and non-response at the questionnaires, we did not conduct inferential statistics.

From the 50 students who participated in this field experiment, sixteen students from each condition (32; 64%) responded to the questionnaire. The 16 students in the system-generated feed-back condition rated the distribution of participation as rather evenly distributed (M = 3.13; SD = 1.54), while 16 students who received the GAT based on peer-generated feed-back perceived the participation as significantly less evenly distributed, as indicated by a value closer to zero (M = 1.94; SD = 1.88; U = 76.5; Z = − 1.98, $p < 0.05$).

We further compared students' perception of the different GATs (Table 9.2). Students who worked in a group that received a GAT that visualized the number of words rated the information in the GAT significantly more helpful than students who worked in groups that received a visualization of self-reported participation (U = 55.50; Z = -2.90; $p < 0,05$). Similarly, students in the system-generated condition rated the visualization of participation as more realistic than students in the peer-generated condition (U = 68.00; Z = −2.475; $p < 0,05$).

Altogether, exploring the data of our study revealed a trend that system-generated feed-back led to a more even distribution of participation in contrast to peer-generated feed-back. Interestingly, the group members perceived the peer-generated feed-back as less helpful and as a less realistic representation of the distribution of participation. Again, caution is warranted when interpreting the results of the field trial due to the small sample size. Nonetheless, we identified trends that indicate a coherent picture, that is, that while participation may encompass more than simply providing a certain number of words, students perceive this metric as helpful and more realistic than their peers' self-reports.

Table 9.2 Mean ratings of perceived helpfulness and perceived realism

	System-generated (n = 16)	Peer-generated (n = 16)
	M (SD)	M (SD)
Feed-back was helpful	4.06 (0.85)	2.88 (1.15)
Realistic representation of participation	4.50 (0.52)	3.56 (1.26)

9.6   Discussion: What Are Boundary Conditions for the Effective Use of Feed-Back Regarding Collaboration?

For collaborative learning to unfold its potential, groups need to monitor their collaboration and assess the interaction in their group. To this end, they collect feed-back. In this chapter we argued that group awareness tools support groups in collecting feed-back on their collaboration, monitoring their collaboration, and adapting their interaction. However, groups do not benefit from the mere presence of these tools, neither can we take for granted that groups possess effective strategies to make use of the support.

We conceptualized social GATs as a means for feedback, specifically, feedback regarding the interaction. Groups can take up this feedback to improve their collaborative interaction. It should be noted, however, that not all tools that have been characterized as GATs may be conceptualized as source for feedback, for example cognitive GATs that display the knowledge held by the group members (e.g., Engelmann & Hesse, 2011). Prior research suggests that boundary conditions exist which affect the effectiveness of these tools (e.g., Dehler et al., 2009; Janssen et al., 2011; Strauß & Rummel, 2021b). To shed light on potential boundary conditions, we presented two small-scale field experiments that explored different ways of promoting regulation of participation. These field experiments were designed to explore questions that arose from our field experiment (Strauß & Rummel, 2021b) and other studies (Dehler et al., 2009; Janssen et al., 2007a, 2007b, 2011). Specifically, we explored whether groups benefit from instruction for collaborative reflection, and whether an indicator for participation that goes beyond the number of words provides groups with more useful feedback for their regulation. The results of our studies indicate a trend that a collaborative reflection activity may help groups achieve a more even distribution of participation, however, the analyses lack statistical power. Analyzing students' perceptions of an "optimal" distribution of participation showed that students prefer an even distribution of participation, however, different notions may exist. Finally, the results of our second field experiment suggest that students perceive self-reported participation as less valid than a system-generated visualization of the number of words.

A major limitation of the two small-scale field experiments reported above is the small sample size. Further, the items used to assess the individual group members' participation during the self-reports should further validated in more details in futures studies. Thus, the results can only serve to develop hypotheses that can be tested in studies with a larger sample. During the remainder of this chapter we will tie together the results of the two field-experiments reported in this chapter as well as the results from our first field-experiment (Strauß & Rummel, 2021b) and point out factors that may influence that process of taking up and processing feedback regarding the current state of the interaction in the group. While we use our studies as examples, we assume that the boundary conditions will apply to other types of GATs and other sources of visual feedback on collaborative interaction. We organize these factors along the phases of the collaboration management cycle

(Soller et al., 2005) and try to ground them in prior research. We hope that this overview can serve as a starting point for future studies that investigate the role of these factors during collaboration.

Figure 9.3 shows the collaboration management cycle (Soller et al., 2005). Like other cyclical models of self-regulation (e.g., Butler & Winne, 1995; Zimmerman, 2000) the collaboration management cycle is based on the cybernetic notion of a system that seeks to achieve an equilibrium between its current state and a desired goal-state. To reach this desired goal-state, the system (i.e., a group) uses its sensors (i.e., senses, collaboration support) to collect feed-back on the current state of the system, and then processes this feed-back to compare the current state with a set desired state. In case of a discrepancy, the system tries to transform the current state into the goal-state. The original model only contains the phases and examples of supporting technologies for each phase. In Fig. 9.3 we added factors that may affect whether groups will or can take up the feedback from a GAT, process it effectively and perform adequate regulatory actions. Specifically, we propose processes that appear to be potential blockades for continuing monitoring and regulation of collaboration. Additionally, we propose properties of the learning environment that affect whether and how groups will engage in active monitoring and regulation. Finally, the knowledge, perception and motivation of the individual group members affect monitoring and regulation.
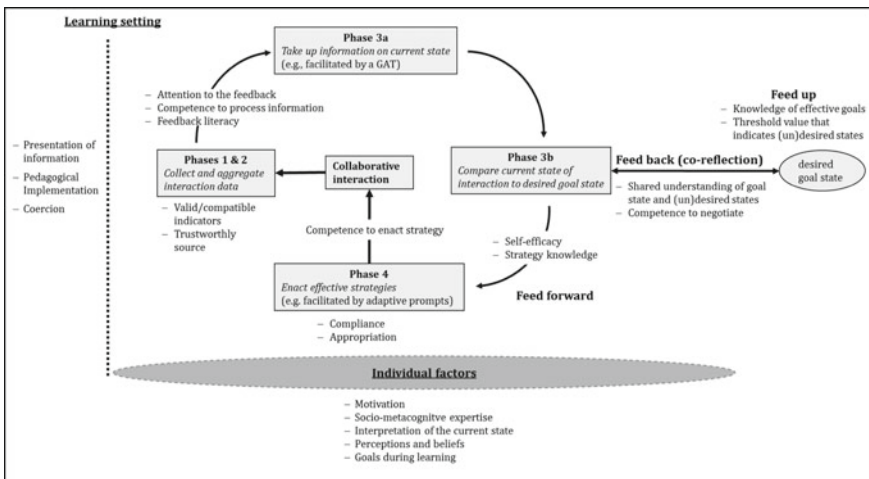

Fig. 9.3 Collaboration management cycle (Soller et al., 2005) and potential boundary conditions for the effective use of feedback on the collaboration

9.6.1 Phases 1 and 2: Collecting and Aggregating Data

The general competence to monitor and regulate the collaboration can be termed socio-metacognitive expertise (Borge & White, 2016). In the first and second phase of the collaboration management cycle, a group (i.e., its members) or a support system (e.g., a GAT) collects and aggregates information about the current state of the group. Being able to do this requires that the learners of a group to look for cues, that is, feed-back. In the case of a GAT, this includes noticing the feed-back and *paying attention* to it. In our previous field experiment (Strauß & Rummel, 2021b) as well as in the two small-scale field experiments reported in this chapter we found that students reported having paid attention to the GAT, however, the number of times that students reported having looked at the visualization did not affect the groups' regulation (i.e., achieving a more even distribution of participation). In this regard, the results reported by Janssen et al. (2011) suggest that the duration of interaction with the GAT is a better predictor of regulation based on the GAT than the mere frequency of interaction with the GAT. Obviously, the mere time spent on the GAT is only a correlate of (socio)cognitive processes that occur within the (members of the) group. Instead, the way that students take up and process the feedback predicts the time spent on the feedback. Which processes may play a role during this will be discussed in the respective sections for the subsequent phases of the collaboration management cycle.

A further aspect that may affect whether a group engages with the feedback is the data and the indicators that are being used to assess the current state of the collaboration. We suggest that the indicators need to be a *valid representation* as well as *compatible with students' perceptions*. During the analyses in the first field study we found evidence that group members hold different conceptualizations what an *optimal* distribution of participation may look like in a group. If support-systems like GATs use indicators that do not align with the learners' perceptions, needs or goals, the learners may ignore the information and not engage with the support any further. For example, a group may pay less attention to the number of words if the group conceptualizes participation based on a different indicator, or if the students perceive the indicator as an unrealistic representation of their behavior. This can be linked to research on cue-utilization (e.g., de Bruin et al., 2017) which has shown that learners regulation depends on whether the learners able to use inadequate cues in order to assess the need for regulation. In this regard, future research should explore groups' needs in terms of group awareness (Schnaubert & Bodemer, 2022) and valid cues that are suited to foster the regulation of collaboration. One question that is worth investigating in this regard is the compatibility between a valid operationalization of an aspect of the interaction in the group (e.g., the distribution of participation) and group members' perception of what indicator best represents the respective aspect of the collaboration.

Furthermore, the relationship between the operationalization of the aspect of the collaboration that is being displayed in the GAT, (i.e., feedback on the collaboration) and the intended pedagogical goal of the GAT should be taken into account as well. According to Rummel (2018), one can distinguish between the

goal of the collaboration support and the aspect of the learning or collaboration that is being targeted by the support in order to achieve this goal. In our field studies, we collected the number of words that each participant had contributed to the group's forum and wiki. The total number of words from each group member was then visualized to present the group the distribution of participation in their group (i.e., the "target" of the GAT). The intended effect of this visualization was to trigger reflection processes in the group which we assumed groups to regulate the distribution of participation and achieve a more even distribution (the "goal" of the GAT). That groups would be motivated to engage in regulation of the distribution of participation was based on the finding that an uneven distribution of participation is a source for frustration (see Strauß & Rummel, 2021a for an overview). As the second field experiment reported above showed, students perceive an even distribution of participation as desirable. However, a question that remains open after our field studies and similar prior studies (e.g. Janssen et al., 2011; Janssen et al., 2007a, 2007b, is which indicators may help groups regulate their collaboration. One potential pitfall when using only behavioral indicators in a GAT is "becom[ing] what you measure" (Duval & Verbert, 2012, p. 3). For the case of our field studies this would mean that the group members would simply focus on producing words. While the results of our original field study (Strauß & Rummel, 2021b) and the content of students' collaborative reflection reported in this chapter do not suggest that students simply contributed more words in order to appear more active in the group, we found evidence of social comparison between students, especially upwards comparison. The particular case of our field experiments underscores the question how participation can best be operationalized. While the number of words is used in many studies it may fall short to capture all aspects of participation. Therefore, we explored the use of self-reported participation in our second field experiment reported in this chapter. While the relationship between the degree of participation of the individual group members, their satisfaction with the collaboration, effective interaction and eventually group performance is complex (see Strauß & Rummel, 2021b for a discussion), building on the argumentation for our second field experiment, implementing a more holistic indicator for participation that combines behavioral data from the learning environment, sensor data, the content of students' and contributions self-reports (i.e., multimodal learning analytics, Ochoa, 2017; Praharaj et al., 2021) may be worth exploring since the distribution of participation in a group is not only a source of dissatisfaction (Strauß & Rummel, 2021a) but also central for learning through interaction and the group's success.

Another potential boundary concerns students' *competence to process the feedback.* For instance, the degree to which learners can process visual information (e.g., a graph in a GAT) depends on the way that the information is presented. Given the limited working memory capacity, visual feed-back should be presented in a way that allows for easy processing. Here, research on instructional psychology (e.g., learning with multimedia, Mayer & Moreno, 2003), human–computer interaction, and human-centered design (e.g., Brandenburger et al., 2020; Jacko, 2012) can inform the design process and facilitate information processing.

Finally, it should be considered whether learners perceive the source of feed-back as *trustworthy*. Research on feed-back has not yet systematically investigated the role the feed-back source (e.g., teacher/experts, peers, task, computer system, self) (Panadero & Lipnevich, 2022), however, for example, Winstone et al. (2017) posit that signals of credibility such as expertise or experience may affect whether and to which extent learners engage with the feed-back. Our results indicate that students prefer the number of words as an indicator for participation over peer-generated feed-back, although the number of words may fall short to cover all facets of participation. This finding points to a tension between trust in computer-systems and trust in peers, or between trust in data and validity of the feed-back.

9.6.2   Phase 3: Taking up Feed-Back and Comparing It to a Desired State

During the third phase of the collaboration management cycle a group compares the current state of the collaboration to a desired goal-state. This goal may be set by the group itself or externally, for example by the task or the teacher. To analyze the relevant processes in more detail, we propose to distinguish between the process of taking up the feed-back and comparing the current state of the collaboration with the goal-state. Thus, we split phase 3 into two parts (3a and 3b, Fig. 9.3).

In the first half of phase 3 (i.e., 3a), a group deliberately *takes up* the feed-back (Hattie & Timperley, 2007) with the goal of comparing it to the desired goal-state. We assume that monitoring and reflecting upon feed-back requires more deliberate processing than merely noticing and viewing the information (phases 1 and 2). The model of regulated learning (Butler & Winne, 1995) as well as research on monitoring (Harkin et al., 2016) describe monitoring as a process that precedes regulatory action. Receiving feed-back in the form of a visualization then requires the *competence to process the information*. This may include *data literacy* (Calzada Prado & Marzal, 2013) as well as *feed-back literacy*, that is, "[…] an understanding of what feed-back is and how it can be managed effectively; capacities and dispositions to make productive use of feed-back; and appreciation of the roles of teachers and themselves in these processes" (Carless & Boud, 2018, p. 1316). These competencies enable learners to become active agents who make sense of the feed-back information and adapt their behavior (Carless & Boud, 2018).

In the second part of the third phase (3b), a group compares the current state with a desired state and assesses whether regulation is required. In case of a dis-crepancy, the collaboration management cycle predicts that the group initiates a reflection process to identify potential reasons for the discrepancy (Boud et al., 1985; Gabelica et al., 2014; Kori et al., 2014; Soller et al., 2005). Hattie and Tim-perley (2007) refer to this as feed-*back*. One potential barrier here is learners' *motivation*. Following Butler and Winne (1995), students' motivation affects how much they invest in regulation. Also, if learners do not expect that their efforts will

be beneficial for the groups' performance they are less likely to put in additional effort (e.g., collective effort model, Karau & Williams, 1993). When the group members compare the current state with a desired goal-state, their *interpretation of the current state* and their *knowledge about effective goals* (e.g., which degree of discrepancy requires regulation) are further potential boundary conditions for regulation (i.e., feed-up). In this regard, Butler and Winne (1995) stress that the configuration of the goal-state should be appropriate because otherwise regulation fails to lead to desired outcomes. For the context of our studies, the questions remain whether achieving an even distribution of words is an appropriate goal for a productive group, which degree of inequality represent an ineffective state of unequal participation, and which indicators may be the most helpful for a group to monitor and regulate their collaboration (see Strauß & Rummel, 2021b for an initial discussion of this point).

With respect to the desired goal-state, we acknowledge that the individual group members may hold different (and diverging) perspectives of effective interaction patterns and goal-states. For example, the students in the first field experiment described above held different ideas of the "optimal" distribution of participation during collaboration. This ranged from an exactly equal distribution of words to any meaningful contributions. Consequently, within a group, there may not exist a shared understanding regarding the desired goal state (Clark & Brennan, 1991; Hadwin et al., 2018). Given that goals play an important role for regulation as goals describe the desired state that should be achieved through regulation, we propose that a *shared understanding of goals and (un)desired states* is necessary to negotiate and coordinate potential regulatory actions. Given findings that a shared perception of the current task is an important factor for effective collaboration (Hadwin et al., 2018), we hypothesize that a diverging set of goals or collaboration norms may affect the motivation to regulate the collaboration. Besides having the competence to process (i.e., make sense of) the information (i.e., feed-back), the members of a group collectively need the *competence to collectively negotiate* about the current state of the collaboration and whether action is needed.

9.6.3   Phase 4: Regulating the Collaboration

In the fourth phase, a group enacts regulation strategies to transform the current state of the collaboration into the desired goal-state (i.e., feed forward). Whether individuals enact strategies or adapt their behavior depends on their *self-efficacy*, that is, their expectation that they are capable of achieving a goal and whether their actions will lead to the desired goal (outcome expectation) (Luszczynska & Schwarzer, 2020). Since striving to meet a goal is a volitional process which requires effort, Webb and de Bruin (2020) propose that individuals only invest this effort if the goal is important to them.

Further, Butler and Winne (1995) acknowledge that *students' perceptions* and *beliefs* affect whether and how students process feed-back and consequently regulate their learning. For instance, if learners hold the belief that learning progress

occurs quickly, they are more likely to employ superficial learning strategies (Butler & Winne, 1995). Whether relevant perceptions and beliefs exist is still not explored.

Once students engage in regulation, their success depends on students' *knowledge about appropriate strategies* (Butler & Winne, 1995; Carless & Boud, 2018; Webb & de Bruin, 2020) as well as their *competence to enact these strategies* (see Flavell et al., 1966; Hübner et al., 2010 for stages of strategy acquisition, and Kollar et al., 2007; Kollar et al., 2018 for internal collaboration scripts). If the learners of a group do not possess adequate strategies or lack the expertise to use them, the group may struggle to achieve the desired goal-state. At this point during the regulation, adaptive technology may scaffold the regulation process by suggesting groups with effective strategies. When designing an adaptive system that offers groups explicit guidance (i.e., a guiding system, Soller et al., 2005), designers need to consider students' internal collaboration scripts (Kollar et al., 2018) and which *threshold values indicate a problematic state* (i.e., what constitutes a "large" discrepancy between the current state and the desired goal state). This value does not necessarily have to be in line with students' perceptions but still should motivate students to follow the prompted regulation strategy. If students do not agree with the system's assessment of the current state or with the proposed strategy, they may be less *compliant with the support*. The challenge of compliance with instructional support has rarely been addressed by prior studies (some exceptions are Bannert et al., 2015; Daumiller & Dresel, 2019; Kwon et al., 2013). Again, students' trust in the feedback may influence whether they engage with it or follow suggestions made by the collaboration support. The question of compliance may further depend on the *pedagogical implementation* of the support. Wise (2014), Wise and Vytasek (2017) suggest that learning analytics interventions need to be implemented carefully. Alternatively, instead of providing learners with agency to engage with feed-back, computer support may also include *coercion* (Rummel, 2018) to achieve compliance. Previous studies (e.g., Kirschner et al., 2008) provide promising evidence that coercion benefits collaboration. However, the question remains whether students on all competence levels equally benefit from coerced support (over-scripting, Dillenbourg, 2002).

Another factor that plays a role are learners' *goals* during collaboration. While working in a group, developing group awareness is only a secondary task for the group (Gutwin & Greenberg, 2001) while the primary goal usually encompasses solving a problem or creating a joint artifact such as a presentation. According to Borge et al. (2018), groups rarely invest effort in regulating the interaction, instead, they focus on solving the joint task. Thus, during collaboration, a group may not invest much effort in achieving an even distribution of participation. Students' goals during collaboration further affect how students perceive and use collaboration environment. As a result, students may *appropriate* the support so that they can achieve their goals (Tchounikine, 2016, 2019). For example, students in our field experiment (Strauß & Rummel, 2021b) reported using the GAT to learn which group members can be trusted to be good collaborators. The observation

indicates that the original purpose of the GAT may not have covered students' needs in terms of feed-back.

9.7   Conclusion

In this chapter, we conceptualized group awareness tools (GATs) from a feed-back perspective and argued that groups may use this feedback to regulate their interaction. Improving the quality of the interaction in the group serves not only the performance of the group (e.g., successfully solving a problem) but also affects learning through interaction. As prior research on instructional feed-back and peer-feed-back has shown, there are several factors that affect whether and to which degrees students can benefit from feedback, and thus from GATs. While cybernetic models like the one proposed by Soller et al. (2005), Butler and Winne (1995) or Zimmerman (2000) are often used to describe the regulation processes, these models may falls short to model the intricate details of regulation, such as students' goals, motivation, perceptions, or competencies, and thus may fall short to predict regulation processes.

Thus far, research on GATs has not presented a comprehensive framework regarding the mechanisms underlying their effectiveness. We became sensitive to this issue because implementing GATs into authentic learning settings did not yield the expected results and our explorative analysis lead to more questions than answers (Strauß & Rummel, 2021b).

Based on the results of prior research and our studies, we propose that leveraging feed-back from GATs regarding the interaction of groups is demanding for students and that research still must identify the mechanisms and boundary conditions for this type of collaboration support. Bringing together evidence from different fields such as team feed-back, instructional feed-back, peer-feed-back, and group awareness, we located different boundary conditions during the process of computer-supported monitoring and regulation of collaboration. Since our work is only a first step towards a systematic investigation of monitoring and regulation of interaction in groups and how groups may leverage feed-back regarding the interaction, we warmly welcome future research on how groups can benefit from feed-back on their collaboration.

References

Aggarwal, P., & O'Brien, C. L. (2008). Social loafing on group projects: Structural antecedents and effects on student satisfaction. *Journal of Marketing Education, 30*(3), 255–264. https://doi.org/10.1177/0273475308322283

Bachour, K., Kaplan, F., & Dillenbourg, P. (2010). An interactive table for supporting participation balance in face-to-face collaborative learning. In *Proceedings of the 11th International Conference on Ubiquitous Computing September 30-October 3, 2009,* (Vol. 3. No. 3, pp. 203–213) Orlando, Florida, USA. https://doi.org/10.1109/TLT.2010.18

Baker, M., Hansen, T., Joiner, R., & Traum, D. (1999). The role of grounding in collaborative learning tasks. In P. Dillenbourg (Ed.), *Collaborative learning. Cognitive and computational approaches* (pp. 31–63). Pergamon.

Bannert, M., Sonnenberg, C., Mengelkamp, C., & Pieger, E. (2015). Short- and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. *Computers in Human Behavior, 52*, 293–306. https://doi.org/10.1016/j.chb.2015.05.038

Bodemer, D. (2011). Tacit guidance for collaborative multimedia learning. *Computers in Human Behavior, 27*(3), 1079–1086. https://doi.org/10.1016/j.chb.2010.05.016

Bodemer, D., & Dehler, J. (2011). Group awareness in CSCL environments. *Computers in Human Behavior, 27*(3), 1043–1045. https://doi.org/10.1016/j.chb.2010.07.014

Bodemer, D., Janssen, J., & Schnaubert, L. (2018). Group awareness tools for computer-supported collaborative learning. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences* (pp. 351–358). Routledge.

Borge, M., & White, B. (2016). Toward the development of socio-metacognitive expertise: An approach to developing collaborative competence. *Cognition and Instruction, 34*(4), 323–360. https://doi.org/10.1080/07370008.2016.1215722

Borge, M., Ong, Y. S., & Rosé, C. P. (2018). Learning to monitor and regulate collective thinking processes. *International Journal of Computer-Supported Collaborative Learning, 13*(1), 61–92. https://doi.org/10.1007/s11412-018-9270-5

Boud, D., Keogh, R., & Walker, D. (1985). Promoting reflection in learning: A model. In D. Boud, R. Keogh, & D. Walker (Eds.), *Reflection: Turning experience into learning* (pp. 18–40). Routledge.

Brandenburger, J., Constapel, M., Hellbrück, H., & Janneck, M. (2020). Analysis of types, positioning and appearance of visualizations in online teaching environments to improve learning experiences. In W. Karwowski, T. Ahram, & S. Nazir (Eds.), *Advances in intelligent systems and computing: Vol. 963. Advances in human factors in training, education, and learning sciences* (Vol. 963, pp. 355–366). Springer International Publishing. https://doi.org/10.1007/978-3-030-20135-7_35

Buder, J. (2011). Group awareness tools for learning: Current and future directions. *Computers in Human Behavior, 27*(3), 1114–1117. https://doi.org/10.1016/j.chb.2010.07.012

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245–281. https://doi.org/10.3102/00346543065003245

Calzada Prado, J., & Marzal, M. Á. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri*, *63*(2). https://doi.org/10.1515/libri-2013-0010

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Chen, J., Wang, M., Kirschner, P. A., & Tsai, C.-C. (2018). The role of collaboration, computer use, learning environments, and supporting strategies in CSCL: A meta-analysis. *Review of Educational Research, 88*(6), 799–843. https://doi.org/10.3102/0034654318791584

Clarebout, G., & Elen, J. (2006). Tool use in computer-based learning environments: Towards a research framework. *Computers in Human Behavior, 22*(3), 389–411. https://doi.org/10.1016/j.chb.2004.09.007

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on Socially Shared Cognition* (13), 127–149.

Daumiller, M., & Dresel, M. (2019). Supporting self-regulated learning with digital media using motivational regulation and metacognitive prompts. *The Journal of Experimental Education, 87*(1), 161–176. https://doi.org/10.1080/00220973.2018.1448744

de Bruin, A. B. H., Dunlosky, J., & Cavalcanti, R. B. (2017). Monitoring and regulation of learning in medical education: The need for predictive cues. *Medical Education, 51*(6), 575–584. https://doi.org/10.1111/medu.13267

Dehler, J., Bodemer, D., Buder, J., & Hesse, F. W. (2009). Providing group knowledge awareness in computer-supported-collaborative learning: Insights into learning mechanisms. *Research and*

Practice in Technology Enhanced Learning, 04(02), 111–132. https://doi.org/10.1142/S17932 06809000660

Dehler, J., Bodemer, D., Buder, J., & Hesse, F. W. (2011). Guiding knowledge communication in CSCL via group knowledge awareness. *Computers in Human Behavior, 27*(3), 1068–1078. https://doi.org/10.1016/j.chb.2010.05.018

Deiglmayr, A., & Spada, H. (2011). Training for fostering knowledge co-construction from collaborative inference-drawing. *Learning and Instruction, 21*(3), 441–451. https://doi.org/10.1016/j.learninstruc.2010.06.004

Dillenbourg, P. (1999). What do you mean by 'collaborative learning'? In P. Dillenbourg (Ed.), *Collaborative learning. Cognitive and computational approaches* (pp. 1–19). Pergamon.

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL. Can we support CSCL?* (pp. 61–91).

Dorfman, R. (1979). A formula for the gini coefficient. *The Review of Economics and Statistics, 61*(1), 146–149. https://www.jstor.org/stable/1924845

Dourish, P., & Bellotti, V. (1992). Awareness and coordination in shared workspaces. In M. Mantel & R. Baecker (Eds.), *Proceedings of the 1992 ACM conference on Computer-supported cooperative work—CSCW '92* (pp. 107–114). ACM Press. https://doi.org/10.1145/143457.143468

Duval, E., & Verbert, K. (2012). Learning Analytics. *Eleed, 8.* https://eleed.campussource.de/archive/8/3336

Engelmann, T., & Hesse, F. W. (2011). Fostering sharing of unshared knowledge by having access to the collaborators' meta-knowledge structures. *Computers in Human Behavior, 27*(6), 2078–2087. https://doi.org/10.1016/j.chb.2011.06.002

Flavell, J. H., Beach, D. R., & Chinsky, J. M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development, 37*(2), 283–299.

Gabelica, C., van den Bossche, P., Segers, M., & Gijselaers, W. H. (2014). Dynamics of team reflexivity after feedback. *Frontline Learning Research*, 64–91. https://doi.org/10.14786/flr.v2i2.79

Gutwin, C., & Greenberg, S. (2001). The importance of awareness for team cognition in distributed collaboration. In E. Salas, S. M. Fiore, & J. A. Cannon-Bowers (Eds.), *Team cognition: Process and performance at the Inter- and Intra-individual level* (pp. 177–201). APA Press.

Hadwin, A. F., Bakhtiar, A., & Miller, M. (2018). Challenges in online collaboration: Effects of scripting shared task perceptions. *International Journal of Computer-Supported Collaborative Learning, 13*(3), 301–329. https://doi.org/10.1007/s11412-018-9279-9

Handke, L., Klonek, F., O'Neill, T. A., & Kerschreiter, R. (2022). Unpacking the role of feedback in virtual team effectiveness. *Small Group Research, 53*(1), 41–87. https://doi.org/10.1177/10464964211057116

Harkin, B., Webb, T. L., Chang, B. P. I., Prestwich, A., Conner, M., Kellar, I., Benn, Y., & Sheeran, P. (2016). Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. *Psychological Bulletin, 142*(2), 198–229. https://doi.org/10.1037/bul0000025

Hattie, J. (2009). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. *Routledge.* https://doi.org/10.4324/9780203887332

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hrastinski, S. (2008). What is online learner participation? A literature review. *Computers and Education, 51*(4), 1755–1765. https://doi.org/10.1016/j.compedu.2008.05.005

Hübner, S., Nückles, M., & Renkl, A. (2010). Writing learning journals: Instructional support to overcome learning-strategy deficits. *Learning and Instruction, 20*(1), 18–29. https://doi.org/10.1016/j.learninstruc.2008.12.001

Jacko, J. A. (2012). The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications (3rd ed.). *Human factors and ergonomics.* CRC Taylor & Francis. https://ebookcentral.proquest.com/lib/subhh/detail.action?docID=911990

Janssen, J., & Bodemer, D. (2013). Coordinated computer-supported collaborative learning: Awareness and awareness tools. *Educational Psychologist, 48*(1), 40–55. https://doi.org/10.1080/00461520.2012.749153

Janssen, J., Erkens, G., & Kanselaar, G. (2007a). Visualization of agreement and discussion processes during computer-supported collaborative learning. *Computers in Human Behavior, 23*(3), 1105–1125. https://doi.org/10.1016/j.chb.2006.10.005

Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007b). Visualization of participation: Does it contribute to successful computer-supported collaborative learning? *Computers & Education, 49*(4), 1037–1065. https://doi.org/10.1016/j.compedu.2006.01.004

Janssen, J., Erkens, G., & Kirschner, P. A. (2011). Group awareness tools: It's what you do with it that matters. *Computers in Human Behavior, 27*(3), 1046–1058. https://doi.org/10.1016/j.chb.2010.06.002

Järvelä, S., Kirschner, P. A., Hadwin, A. F., Järvenoja, H., Malmberg, J., Miller, M., & Laru, J. (2016). Socially shared regulation of learning in CSCL: Understanding and prompting individual- and group-level shared regulatory activities. *International Journal of Computer-Supported Collaborative Learning, 11*(3), 263–280. https://doi.org/10.1007/s11412-016-9238-2

Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of computer-supported collaborative learning: A meta-analysis of CSCL in STEM education during 2005–2014. *Educational Research Review, 28*, 100284. https://doi.org/10.1016/j.edurev.2019.100284

Jermann, P., & Dillenbourg, P. (2008). Group mirrors to support interaction regulation in collaborative problem solving. *Computers & Education, 51*(1), 279–296. https://doi.org/10.1016/j.compedu.2007.05.012

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher, 38*(5), 365–379. https://doi.org/10.3102/0013189X09339057

Jongsawat, N., & Premchaiswadi, W. (2009). An empirical study of group awareness information in web-based group decision support system in a field test setting. In *7th International Conference on ICT and Knowledge Engineering, Bangkok, Thailand: Siam University*, pp. 15–23. https://doi.org/10.1109/ICTKE.2009.5397318

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65*(4), 681–706. https://doi.org/10.1037/0022-3514.65.4.681

Kimmerle, J., & Cress, U. (2009). Visualization of group members' participation. *Social Science Computer Review, 27*(2), 243–261. https://doi.org/10.1177/0894439309332312

Kimmerle, J., Cress, U., & Hesse, F. W. (2007). An interactional perspective on group awareness: Alleviating the information-exchange dilemma (for everybody?). *International Journal of Human-Computer Studies, 65*(11), 899–910. https://doi.org/10.1016/j.ijhcs.2007.06.002

King, A. (2007). Scripting collaborative learning processes: A cognitive perspective. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.), *Scripting computer-supported collaborative learning: Cognitive, computational and educational perspectives* (pp. 13–37). Springer US. https://doi.org/10.1007/978-0-387-36949-5_2

Kirschner, P. A., Beers, P. J., Boshuizen, H. P., & Gijselaers, W. H. (2008). Coercing shared knowledge in collaborative learning environments. *Computers in Human Behavior, 24*(2), 403–420. https://doi.org/10.1016/j.chb.2007.01.028

Kirschner, P. A., Kreijns, K., Phielix, C., & Fransen, J. (2015). Awareness of cognitive and social behaviour in a CSCL environment. *Journal of Computer Assisted Learning, 31*(1), 59–77. https://doi.org/10.1111/jcal.12084

Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*(6), 708–721. https://doi.org/10.1016/j.learninstruc.2007.09.021

Kollar, I., Wecker, C., & Fischer, F. (2018). Scaffolding and scripting (computer-supported) collaborative learning. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences* (pp. 340–350). Routledge.

Kori, K., Pedaste, M., Leijen, Ä., & Mäeots, M. (2014). Supporting reflection in technology-enhanced learning. *Educational Research Review, 11*, 45–55. https://doi.org/10.1016/j.edurev.2013.11.003

Kwon, K., Hong, R.-Y., & Laffey, J. M. (2013). The educational impact of metacognitive group coordination in computer-supported collaborative learning. *Computers in Human Behavior, 29*(4), 1271–1281. https://doi.org/10.1016/j.chb.2013.01.003

Leshed, G., Perez, D., Hancock, J. T., Cosley, D., Birnholtz, J., Lee, S., McLeod, P. L., & Gay, G. (2009). Visualizing real-time language-based feedback on teamwork behavior in computer-mediated groups. In *Conference Proceedings and Extended Abstracts; The 27th Annual CHI Conference on Human Factors in Computing Systems,* Boston, USA. http://dl.acm.org/citation.cfm?id=1518701

Lin, J.-W., & Tsai, C.-W. (2016). The impact of an online project-based learning environment with group awareness support on students with different self-regulation levels: An extended-period experiment. *Computers and Education, 99*, 28–38. https://doi.org/10.1016/j.compedu.2016.04.005

Lipnevich, A. A., & Panadero, E. (2021). A review of feedback models and theories: Descriptions, definitions, and conclusions. *Frontiers in Education*, *6*, Article 720195. https://doi.org/10.3389/feduc.2021.720195

Luszczynska, A., & Schwarzer, R. (2020). Changing behavior using social cognitive theory. In M. S. Hagger, L. D. Cameron, K. Hamilton, N. Hankonen, & T. Lintunen (Eds.), *The handbook of behavior change* (Vol. 2, pp. 32–45). Cambridge University Press. https://doi.org/10.1017/9781108677318.003

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning, 2*(1), 63–86. https://doi.org/10.1007/s11412-006-9005-x

Michinov, N., & Primois, C. (2005). Improving productivity and creativity in online groups through social comparison process: New evidence for asynchronous electronic brainstorming. *Computers in Human Behavior, 21*(1), 11–28. https://doi.org/10.1016/j.chb.2004.02.004

Ochoa, X. (2017). Multimodal learning analytics. In C. Lang, G. Siemens, A. Wise, & D. Gasevic (Eds.), *Handbook of learning analytics* (pp. 129–141). Society for Learning Analytics Research (SoLAR). https://doi.org/10.18608/hla17.011

Ollesch, L., Heimbuch, S., & Bodemer, D. (2021). Improving learning and writing outcomes: Influence of cognitive and behavioral group awareness tools in wikis. *International Journal of Computer-Supported Collaborative Learning, 16*(2), 225–259. https://doi.org/10.1007/s11412-021-09346-6

Pai, H.-H., Sears, D. A., & Maeda, Y. (2015). Effects of small-group learning on transfer: A meta-analysis. *Educational Psychology Review, 27*(1), 79–102. https://doi.org/10.1007/s10648-014-9260-8

Panadero, E., & Järvelä, S. (2015). Socially shared regulation of learning: A review. *European Psychologist, 20*(3), 190–203. https://doi.org/10.1027/1016-9040/a000226

Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review, 35*, 100416. https://doi.org/10.1016/j.edurev.2021.100416

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences, 13*(3), 423–451.

Phielix, C., Prins, F. J., Kirschner, P. A., Erkens, G., & Jaspers, J. (2011). Group awareness of social and cognitive performance in a CSCL environment: Effects of a peer feedback and reflection tool. *Computers in Human Behavior, 27*(3), 1087–1102. https://doi.org/10.1016/j.chb.2010.06.024

Praharaj, S., Scheffel, M., Drachsler, H., & Specht, M. (2021). Literature review on co-located collaboration modeling using multimodal learning analytics—Can we go the whole nine yards?

IEEE Transactions on Learning Technologies, 14(3), 367–385. https://doi.org/10.1109/TLT.2021.3097766

Price, K. H., Harrison, D. A., & Gavin, J. H. (2006). Withholding inputs in team contexts: Member composition, interaction processes, evaluation structure, and social loafing. *The Journal of Applied Psychology, 91*(6), 1375–1384. https://doi.org/10.1037/0021-9010.91.6.1375

Robert, L. P. (2020). Behavior-output control theory, trust and social loafing in virtual teams. *Multimodal Technologies and Interaction, 4*(3), 39. https://doi.org/10.3390/mti4030039

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *NATO ASI Series, Series F: Vol. 128. Computer Supported Collaborative Learning* (pp. 69–97). Springer. https://doi.org/10.1007/978-3-642-85098-1_5

Rummel, N. (2018). One framework to rule them all? Carrying forward the conversation started by Wise and Schwarz. *International Journal of Computer-Supported Collaborative Learning, 13*(1), 123–129. https://doi.org/10.1007/s11412-018-9273-2

Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *Journal of the Learning Sciences, 14*(2), 201–241. https://doi.org/10.1207/s15327809jls1402_2

Schnaubert, L., & Bodemer, D. (2022). Group awareness and regulation in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning.* Advance online publication. https://doi.org/10.1007/s11412-022-09361-1

Schnaubert, L., & Bodemer, D. (2019). Providing different types of group awareness information to guide collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 14*(1), 7–51. https://doi.org/10.1007/s11412-018-9293-y

Soller, A., Martínez, A., Jermann, P., & Muehlenbrock, M. (2005). From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. *International Journal of Artificial Intelligence in Education*(15), 261–290.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48*(6), 1467–1478. https://doi.org/10.1037/0022-3514.48.6.1467

Strauß, S., & Rummel, N. (2021a). Problematic interaction patterns during online-collaboration. A library and a survey. In C. E. Hmelo-Silver, B. de Wever, & J. Oshima (Chairs), *Proceedings of the 14th International Conference on Computer-Supported Collaborative Learning - CSCL 2021a,* Bochum.

Strauß, S., & Rummel, N. (2021b). Promoting regulation of equal participation in online collaboration by combining a group awareness tool and adaptive prompts. But does it even matter? *International Journal of Computer-Supported Collaborative Learning*, 67–104. https://doi.org/10.1007/s11412-021-09340-y

Tchounikine, P. (2016). Contribution to a theory of CSCL scripts: Taking into account the appropriation of scripts by learners. *International Journal of Computer-Supported Collaborative Learning, 11*(3), 349–369. https://doi.org/10.1007/s11412-016-9240-8

Tchounikine, P. (2019). Learners' agency and CSCL technologies: Towards an emancipatory perspective. *International Journal of Computer-Supported Collaborative Learning, 69*(1), 1–14. https://doi.org/10.1007/s11412-019-09302-5

Tenenbaum, H. R., Winstone, N. E., Leman, P. J., & Avery, R. E. (2020). How effective is peer interaction in facilitating learning? A Meta-Analysis. *Journal of Educational Psychology, 112*(7), 1303–1319. https://doi.org/10.1037/edu0000436

Umpleby, S. A., & Dent, E. B. (1999). The origins and purposes of several traditions in systems theory and cybernetics. *Cybernetics and Systems, 30*(2), 79–103. https://doi.org/10.1080/019697299125299

Walther, J. B., & Bunz, U. (2005). The rules of virtual groups: Trust, liking, and performance in computer-mediated communication. *Journal of Communication, 55*(4), 828–846. https://doi.org/10.1111/j.1460-2466.2005.tb03025.x

Webb, T. L., & de Bruin, M. (2020). Monitoring interventions. In M. S. Hagger, L. D. Cameron, K. Hamilton, N. Hankonen, & T. Lintunen (Eds.), *The handbook of behavior change* (Vol. 2, pp. 537–553). Cambridge University Press. https://doi.org/10.1017/9781108677318.037

Wiener, N. (1949). Kybernetik. *Physikalische Blätter, 5*(8), 355–362. https://doi.org/10.1002/phbl.19490050802

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist, 52*(1), 17–37. https://doi.org/10.1080/00461520.2016.1207538

Wise, A. F., & Vytasek, J. (2017). Learning analytics implementation design. In C. Lang, G. Siemens, A. Wise, & D. Gasevic (Eds.), *Handbook of learning analytics* (pp. 151–160). Society for Learning Analytics Research (SoLAR). https://doi.org/10.18608/hla17.013

Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. In M. Pistilli, J. Willis, D. Koch, K. Arnold, S. Teasley, & A. Pardo (Eds.), *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge (LAK '14)* (pp. 203–211). ACM Press. https://doi.org/10.1145/2567574.2567588

Yukawa, J. (2006). Co-reflection in online learning: Collaborative critical thinking as narrative. *International Journal of Computer-Supported Collaborative Learning, 1*(2), 203–228. https://doi.org/10.1007/s11412-006-8994-9

Zimmerman, B. J. (2000). Attaining self-regulation. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7

10

Viewbrics: A Technology-Enhanced Formative Assessment Method to Mirror and Master Complex Skills with Video-Enhanced Rubrics and Peer Feedback in Secondary Education

Ellen Rusman, Rob Nadolski, and Kevin Ackermans

'*Like master, Like man*'

10.1 Introduction and Background

The acquisition of cross-curricular complex skills (such as collaboration, presentation and information skills) is important for students in secondary education, as they are often required in their future professional life. However, looking closer at daily practice in secondary schools, it shows that they struggle with how to organize the acquisition, guidance and supervision and (formative and summative) assessment of these skills. Both schools and teachers recognize the importance of teaching cross-curricular complex skills, nevertheless they are only practiced occasionally (Rusman et al., 2014; Thijs et al., 2014). The extent to which this happens also largely depends on efforts of individual teachers. Moreover, the SLO (Dutch national institute for curriculum planning and development) points

E. Rusman (✉) · R. Nadolski · K. Ackermans
Department Technology-Enhanced Learning and Innovation (TELI), Faculty of Educational Sciences and Technology, Open University of the Netherlands, Building Chiba, Room 1.12, 6401 DL Heerlen, The Netherlands
e-mail: ellen.rusman@ou.nl

out that the training of skills is often not organized in a methodical, structured, goal-oriented and substantiated manner (Thijs et al., 2014, p. 103). When schools support the acquisition of cross-curricular complex skills, it is often done through project-based education, using textual rubrics occasionally and incidentally and in a time-and paper consuming manner. Furthermore, to streamline both acquisition of these skills as well as supervision and guidance during practice, students and teachers need to develop a concrete and consistent mental model of skills. If students know to what skill performance level they work towards (feed-up) and where they stand with regard to this level (feedback), they can better regulate their practice (feed forward) to achieve these objectives (Hattie & Timperley, 2007). An analytic assessment rubric describes skills' mastery levels, usually textually, through a set of quality criteria and descriptions for the constituent skills of a specific skill (Andrade & Du, 2005). Thus, they can become a 'mirror' to determine one's skills performance level. However, we expected that textual rubrics could still be improved, as many aspects of desired behavior can hardly be described into words. Therefore, we designed and developed a technology-enhanced and structured formative assessment method, called Viewbrics. Within the Viewbrics method, we alternatively proposed to use video-enhanced rubrics as a manner to counterbalance disadvantages of textual rubrics.

We were interested whether such a technology-enhanced structured formative assessment method, with either video-enhanced or textual analytic rubrics, could offer a more efficient and effective solution to teach, practice, achieve and formatively evaluate cross-curricular complex skills. Thus, the design-based research project Viewbrics was conceived (Rusman et al., 2019). In this chapter a (theory and practice-based) description of both the design and development as well as the characteristics of the Viewbrics technology-enhanced formative assessment method are described. Furthermore, overall results of two alternative pilot implementations of the Viewbrics method (video-enhanced or textual rubrics) on students' mental models, feedback quality and skills' performance levels regarding complex skills in two secondary schools are reported.

10.1.1 The Acquisition of Complex Skills, Formative Assessment and (Video-Enhanced) Rubrics

Complex skills consist of constituent subskills which concertation require high cognitive effort and concentration (Kirschner & Merriënboer, 2008; Van Merriënboer & Kirschner, 2017; Voogt & Pareja-Roblin, 2012). Complex generic (also 'transversal' or 'twenty-first century') skills are not specific for a domain, occupation or type of task, but important for all kinds of work, education and life in general. These skills are applicable in a broad range of situations and many subject domains (Bowman, 2010). To master a complex skill, it requires frequent, prolonged and repetitive practice, but also (timely) feedback on performances. Also modelling examples and variability in application contexts influence skills' acquisition (Kirschner & Merriënboer, 2008; Van Merriënboer & Kirschner, 2017). One

of the instruments that can be used to support skills acquisition through structured feedback and reflection during practice are rubrics. Rubrics define the features of work that are considered quality, and can be either holistic or analytic. It is a mechanism for judging the quality of a students' performance on a task (Arter & Chappuis, 2006; Sluijsmans et al., 2013). Analytic assessment rubrics describe a skill, their constituent subskills and a set of quality criteria (performance indicators) for the various mastery levels of a sub-skill (Andrade & Du, 2005) in text. Performance indicators specify aspects of variation in the complexity of a skill, constituent sub-skills and related performance levels (Rusman & Dirkx, 2017). For example; the skills mastery (as displayed and visible behaviour) ranging from a novice to that of an expert. When students acquire insight in their performance compared to the targeted mastery level of a complex skill, they can better monitor their own learning activities and communicate with teachers (Panadero & Jonsson, 2013; Schildkamp et al., 2014). Thus, rubrics provide opportunities to jointly adjust the teaching–learning process through reflection. Furthermore, an analytic rubric provides the opportunity to structure teachers' and peers' timely and informative feedback, but also to clarify and make expectations about the strived-for mastery level(s) of a skill clear in advance (feed-up) to the learner. This helps learners at the start to envisage the targeted mastery level (Berry et al., 2007) and enables them to focus on the aspects of a skill that they didn't master yet very well while practicing.

However, many aspects of complex skills mastery refer to motoric activities, time-consecutive operations and processes that are hardly captured in text (e.g. body posture or use of voice during a presentation) (De Grez et al., 2013; O'Donovan et al. 2004). Furthermore, the context in which a skill is practiced and behavior enacted is important. Contextual conditions and characteristics imply and generate implicit knowledge (tacit knowledge, 'knowing how/'knowing why'), which is interwoven with practical activities, operations and behavior in the physical world (Westera, 2011). Text supposedly also leaves more space for personal interpretation of performance indicators of a complex skill than video. Also, in educational practice it showed that textual analytic rubrics didn't clarify the desired mastery level of a skill sufficiently and concrete enough for pupils, as students often asked questions like "What should I exactly do?" and "To what kind of things should I pay attention to?" (Rusman, 2015). Therefore, text-based analytic rubrics only have a restricted capacity to clarify the targeted mastery level of a skill and to assess shown behaviour, as they don't provide information on visible behavioral aspects of mastering a skill (Berry et al., 2007). This could supposedly lead to incomplete and inconsistent mental models of students of the expected skill performance level.

However, these restrictions could might be overcome with video-enhanced rubrics. A video-enhanced rubric (VER) is the synthesis of video modelling examples and a text-based analytic rubric in a digital formative assessment format (Ackermans et al., 2017, 2019b). Video-enhanced rubrics could foster observational learning from desired behavior of a role model in (good/bad) video modelling examples (De Grez et al., 2013; Rohbanfard & Proteau, 2013; Van Gog

et al., 2014). They can also capture implicit contextual knowledge, as they show motoric, temporal and contextual information of a skill, which cannot be expressed in words (Ackermans et al., 2017; Westera, 2011). Van Gog et al. (2014) found an increased performance of task execution when a video-modelling example of an expert was shown and De Grez et al., (2013, 2014) found comparable results when learning presentation skills. Moreover, when teacher trainees compare their own performance with video-modelling examples they 'overrate' their own performance less during self-reflection than without these examples. Additionally, teacher trainees gained an improved insight in their performance compared to the targeted mastery level of a complex skill (Baecher et al., 2013). Therefore, we alternatively proposed to use video-enhanced rubrics within the Viewbrics method as a manner to counterbalance disadvantages of textual rubrics.

10.1.2 Technology-Enhanced Formative Assessment: Process Support for Goal Setting, Practice, Feedback, Reflection and Self-regulation

Formative assessment or 'assessment for learning' aims to support teaching and learning processes by providing developmental feedback to learners (and their teachers) on their understanding or skills during a period of practice and instruction (Black & William, 1998). Formative assessment differs from summative assessment in that it is a continuing process of feedback. In this continuing process information on learners' performances is gathered continuously and mirrored against a set of predefined criteria or good practices. Information is also used to shape improvements and promote an individual's learning, rather than serve as a final formal summary of learners' achievements (Sluijsmans et al., 2013). Providing feedback during formative assessment is also one of the most effective ways to support learning processes (Hattie & Timperley, 2007). Feedback can be specified at different levels (e.g. looking at self-, task-, process-, or self-regulation aspects (Hattie & Timperley, 2007), and by means of different sources, such as self-, peer-, expert- or teacher feedback or via 'built-in' feedback in (technology-enhanced) educational materials (Sluijsmans et al., 2013). The aim is to gather information about (the gap between) the current and desired personal performance goal or mastery level and how this gap can be closed. For example, by carrying out specific learning activities, altering behaviour or (adapted) instruction. To learn new skills, learners first need support to form a clear mental model of the strived-for performance objectives (feed-up). Second, they need concrete, supportive and timely information (Shute, 2008) on their performance in relation to these objectives and instructions or guidelines on how further growth could be achieved by altering learners' thinking or behaviour (feedback). Finally, learners need to reflect on the gained feedback so that they can specify new or adapted objectives and determine where their focus should be when practicing further (feed-forward) (Hattie & Timperley, 2007). The responsibility for learning is shared between both learners as teachers and eventually with their peers (McManus, 2008; Black & Williams,

2009). They determine (jointly) where a learner is going (goals), where (s)he is now (how am I going?) and how a learner can get where (s)he wants (where to go next?) (Hattie & Timperley, 2007), thus forming a natural self-regulative cycle with a Forethought, Performance and Self-reflection phase (Zimmerman, 2008, p. 178). Peer assessment and feedback can play an important role in formative assessment, next to self and expert assessment (Filius, 2019). Both receiving peer feedback as well as providing peer feedback yield improved learning gains compared to only teacher feedback, such as improved presentation skills, critical thinking, self-regulation and reflection skills (Boud, 2001; Vincent-Wayne & Bakewell, 1995; Vincent-Wayne & Bakewell, 1995). Students also self-report that they learned more from providing peer feedback then receiving it (Filius, 2019). Additionally, providing peer feedback in a written form both force and facilitate students to analyze and think critically about a performance and also to phrase and express it in an understandable manner. With written peer feedback students experience extra time to think, reflect and express their feedback, compared to oral and (often) immediate feedback. Peer feedback also offers a practical merit, in that it can facilitate learning and development of students, with a reduction of teachers' time and effort (Candy et al., 1994; Filius, 2019). However, in order to increase the effectiveness of peer feedback, it is important to instruct students in advance in providing (high quality) peer feedback (Nicol, 2010; Shute, 2008).

Furthermore, technology can offer different affordances that potentially facilitate and enhance formative assessment and feedback processes (Norman, 2013; Rusman et al., 2013). It improves *access to practice and assessment by different actors* (e.g. by peers, experts and teachers) anytime, anyplace and anywhere, enabling learners to measure their understanding when and how often they want and allow them *more control of their learning*. Feedback times can be shortened and this can help to change misconceptions rapidly, or feedback may be given from different perspectives, within a group or adapted to a learner. Thus, technology can affect *feedback quality*. Also, technology can *track, trace, store, process and visualize learners' results as well as actions* (Looney & Siemens, 2011), which makes them *visible and available* for various learning purposes, such as individual or group *reflection* or to evaluate and *visualize learners' progress* and growth. Technology can also affect *teacher efficiency*, as teachers can be supported with various tools helping to reduce assessment time and material (e.g. save 'piles of paper' and related work), thus saving time and costs that can be spent otherwise. Additionally, as technology enables rapid updating and combination of (recent) material and display of various formats (e.g. video, audio, annotation etc.), it can also contribute to *more varied and authentic assessment designs* (Rusman et al., 2013).

10.1.3 The Objectives and Outline of the Viewbrics Project

In the Viewbrics project we designed a technology-enhanced formative assessment method with (video-enhanced or text-based) analytic rubrics, to provide both teachers and learners with structured, feasible and convenient process support to formatively assess and provide high quality feedback while practicing skills and to monitor students' skill performance growth. We aimed to fulfill the need for practical, implementable educational models, methods, assessment indicators and instruments, ICT-tools and guidelines to support the acquisition of complex (twenty-first century) skills. We also aimed to make the process of implementing learning activities and assessment practices for skills acquisition more straightforward (Rusman et al., 2014; Thijs et al., 2014). A valid, standardized, cyclic and repeatable technology-enhanced assessment process, in which (video-enhanced) rubrics are 'set' instruments to provide structured, timely, specific and relevant feedback, was desirable from that (practical and straightforward) stance. This could also help to overcome the use of analytic rubrics for summative assessment purposes only and embed formative assessment more regularly in daily educational practice. Additionally, we wanted to introduce a way to make behavior resembling the various mastery levels of a skill more visible as well as structurally support teachers and pupils in the process of providing and using feedback while practicing skills, for which we designed and developed (video-enhanced) rubrics.

Furthermore, we aimed to study effects of structured technology-enhanced process support for formative assessment, peer feedback and the use of (video-enhanced) rubrics for skills acquisition. More specifically, whether technology-enhanced formative assessment process support, peer feedback and video-enhanced rubrics resulted in a more complex ('richer') mental model of a complex skill; improved feedback quality and/or quantity and a significant growth in learners' skills performance.

In this practice-and design based research project (Rusman et al., 2019), an interdisciplinary project team collaborated intensively with various stakeholders (teachers, students, school board, researchers and experts (educational, ICT, interface design)) in order to develop and investigate the Viewbrics method and accompanying digital tool. This was done for three complex (twenty-first century) skills, namely presentation, collaboration and information literacy skills. The project had two phases:

1. a cyclical design-oriented phase for the development of the (technology-enhanced) formative assessment method, the textual rubrics and the video-enhanced rubrics with stakeholders.
2. a (practice-based) research phase into the effects of implementing the method

In the first phase stakeholders met in a core development team, in order to develop the Viewbrics method and the (video-enhanced, VER) rubrics, both from theoretical as well as practical perspective. The core team met once every two weeks. Theory-informed proposals and prototypes for the development of the method and the video-enhanced rubrics (Ackermans et al., 2017, 2019; Mertler, 2001; Van Strien & Joosten-ten Brinke, 2016) were developed, discussed and adapted in line with the feedback of stakeholders: students, teachers and experts. Questions like "How many performance level descriptions will we use in the rubric?; What are the (dis)advantages of starting with the highest or lowest performance level descriptions at the left side of the rubrics? How can we foster a growth perspective of students on their skills development? What steps should the formative assessment method consist of and what/where could be the added value of technology? What should be the constituent subskills described within the rubrics? What behavior can we show in the video modeling example and how should it connect and relate to the performance level description of a subskill?" were discussed, both from a theoretical (based on scientific literature) as well as a practical stance and jointly decided upon. This resulted in a prototype of the technology-enhanced formative assessment process; three analytic text-based rubrics for presentation, collaboration (see Fig. 10.1) and information literacy skills and the design and development of video-enhanced rubrics in which video modelling examples were combined with textual rubrics in a digital formative assessment format (Ackermans et al., 2017, 2019b).

Once a first working technology-enhanced version of the Viewbrics method was ready, it was evaluated on its usability and usefulness with students and teachers in two secondary schools (Rusman et al., 2018) and further adapted, developed and evaluated, until stakeholders were satisfied with the Viewbrics method. In the second phase, the effect of using the Viewbrics technology-enhanced formative assessment method with video-enhanced rubrics and textual rubrics was investigated at two secondary, pre-university education schools for 24 weeks and compared with existing educational practice for skills acquisition (as a control group). This research took place within project-based education, with secondary school students and teachers in six classes (two classes with video-enhanced rubrics, two classes with textual rubrics and two classes as a control group).

We expected that video-enhanced rubrics and textual rubrics within the technology-enhanced formative assessment method, compared to the current educational practice, could lead to richer mental models and improved feedback quality for both students and teachers. As a result, we ultimately expected an increased mastery of skills by students. Additionally, we expected that video-enhanced rubrics compared to textual rubrics, used within the same technology-enhanced formative assessment method, would lead to richer mental models, improved
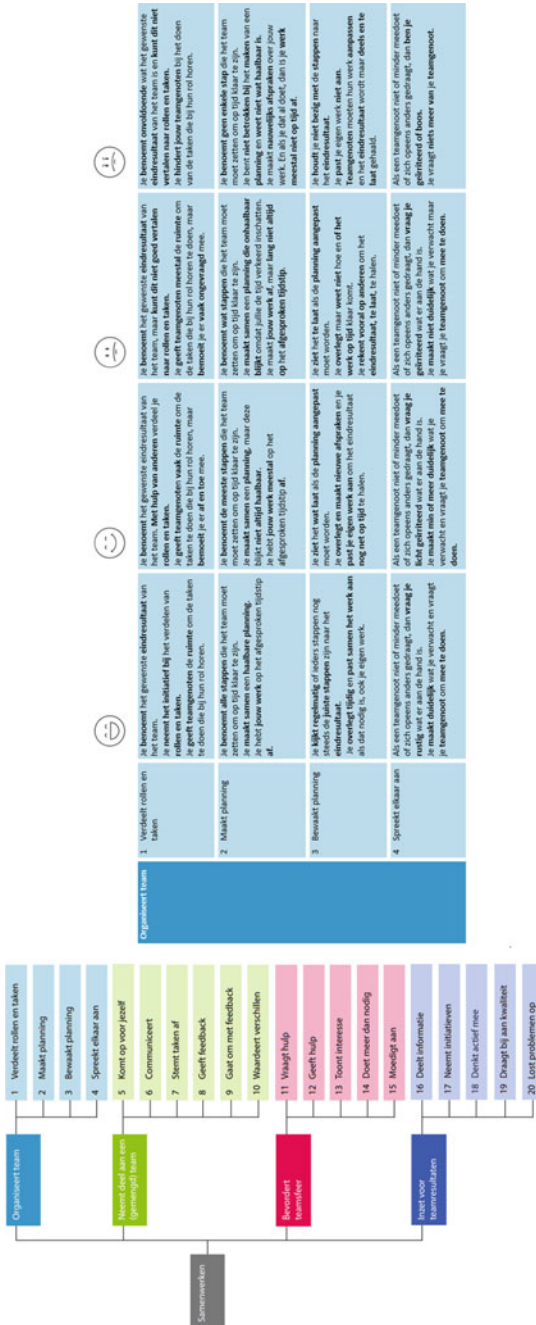
Fig. 10.1 Example of skill decomposition and developed rubric for 'collaborating' skill

feedback quality, and improved skill performance of students. This led to the following twofold research question, which was investigated for three cross-curricular complex skills (presentation, collaboration and information literacy skills):

1. Do rubrics, applied within a (technology-enhanced) formative assessment method, improve (i) the mental model of (ii) the feedback on, and (iii) the performance of a (cross-curricular) complex skill among secondary school pupils when compared to existing educational practice?
2. Do video-enhanced rubrics, applied within a (technology-enhanced) formative assessment method, improve (i) the mental model of (ii) the feedback on, and (iii) the performance of a (cross-curricular) complex skill among pupils in secondary education when compared to textual rubrics?

10.1.4 The Designed Intervention: The Viewbrics Technology-Enhanced Formative Assessment Method

In this section the Viewbrics technology-enhanced formative assessment method is described from the student-learner perspective. The overall formative assessment process supported by the Viewbrics method is visualized in Fig. 10.2 and consists of five main steps, that are described below and illustrated with main interfaces.

Step 1—Watch (video-enhanced) rubrics: Students look either at video-enhanced rubrics (VER) with video-modeling examples and information processing support (by means of a questioning mechanism (Ackermans et al., 2017, 2019b)) or text-based analytic rubrics in the digital tool, in order to form a mental model of a complex generic skill and the strived-for mastery level. This is done to facilitate mental model creation and goal-setting of learners. In the VER implementation of the Viewbrics-method, learners first watch the complete video-modelling example (holistic), then they process the video modeling example by

Fig. 10.2 Formative assessment process in the [Viewbrics] method from a student perspective
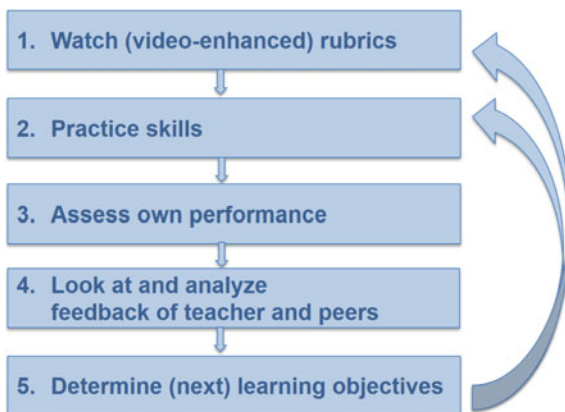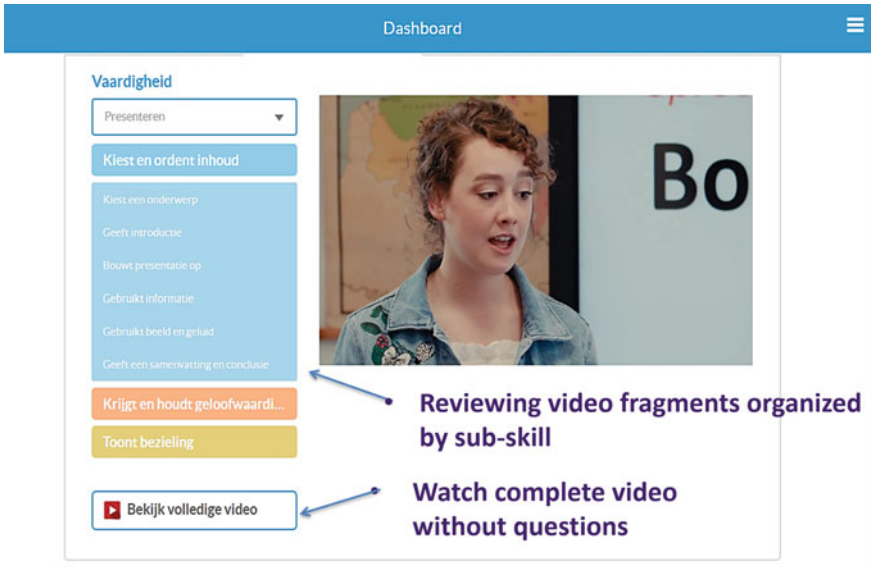
Fig. 10.3 Reviewing video fragments of modeling examples by sub-skill in rubric

means of information processing questions, a modeling example of the highest mastery level of a constituent subskills and color codes which allow learners to link scenes in the video to the related constituent sub-skill in a rubric (Ackermans, 2019; Ackermans et al., 2017, 2019b; Rusman et al., 2019, p. 20) and then they watch fragments of the video-modelling examples, associated with and starting from a subskill (Fig. 10.3) and review the complete video. In the text-based rubric setting, students click through the skill-hierarchy and constituent subskills, and can read through the performance level descriptions related to each subskill.

Step 2—'Practicing a skill': Students go 'into the real world' in order to practice a skill in the educational scenario a teacher provided them with and with the impression of skilled behaviour they formed by looking at the (VER) rubric. In the Viewbrics project this was done in the context of project-based education. Peers and teacher provide feedback on the 'live' performance of a student in class through the use of digital devices (e.g. tablet, laptop), however students only received an overview of this feedback after they did a self-assessment of their performance. Additionally, students provide peer-feedback to the performances of their colleagues, in addition to the teacher (Rusman et al., 2019, p. 21).

Step 3—'Self-assessment': Based on their own experience with practicing skills, their perception of their own performance and the built-in support in the Viewbrics method [(video-enhanced) rubrics, analysis/comparison of performance through peer assessment and technology-enhanced process support] students self-assess their performance by means of the rubrics in the
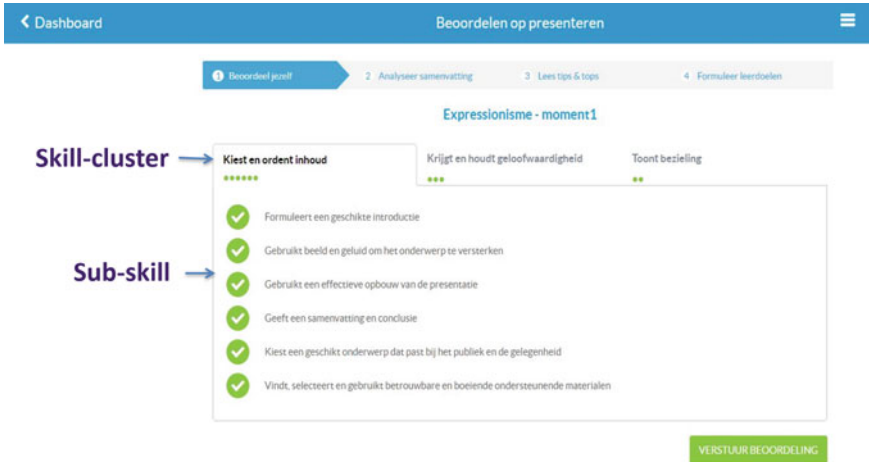
Fig. 10.4 Self-assessment by means of reflection on subskills within a skill-cluster

digital tool (Rusman et al., 2019, p. 22 & 23). The self-assessment is designed comparable to the peer-assessment process, only the person and performance setting vary. Rubrics are organized in skills clusters and sub-skills (Fig. 10.4). Each sub-skill is described in a rubric with four performance level descriptors (Fig. 10.5). Only after completing the self-assessment, students can take a look at the 360-degree feedback of peers and the teacher (who assess students' performances while practicing by scoring the rubrics on a digital device and providing additional tips and tops per skills' cluster). This 360-degree feedback consists of a visualization and a summary of all tips and tops given by peers and teachers.
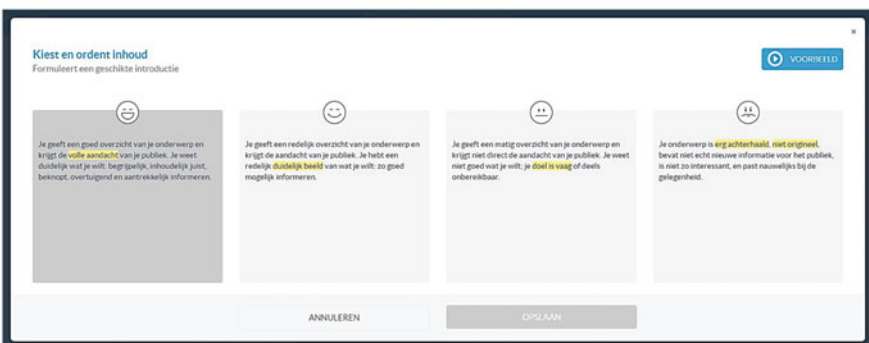
Fig. 10.5 Scoring a rubric with four mastery level descriptions per sub-skill

Fig. 10.6 Skill performance feedback wheel representing students' performance scores

Step 4—'Review and analysis of feedback': The feedback provided by peers and teacher is visualized in a 'skill performance wheel' representing students' performance score on subskills of a complex skill (Fig. 10.6) (Rusman et al., 2019, p. 23 & 24). Each 'spoke' of the wheel represents a constituent subskill of a complex skill and each 'level' on a spoke aligns with a rubric performance level description of this subskill. This visualization allows students (and teachers) to see at a glance on what skills they may still improve and what skills they performed well on, to direct their further and future practice. Performance growth or shrinkage between assessment moments through time are visualized in performance level color highlights (red for performance reduction, green for growth in performance, blue for stable performance) (Fig. 10.7) and the top three skills that went either well or less well during practice are presented below the wheel. Additionally, all provided textual tips and tops are summarized in a feedback report. Students analyze this information and determine what went well and what subskills may still need improvement.

Step 5—'Determine (next) learning objectives': Students describe their learning objectives in the digital tool based on their analysis of self-, peer-and teacher feedback in both the skills performance wheel and the tip/top summary report,
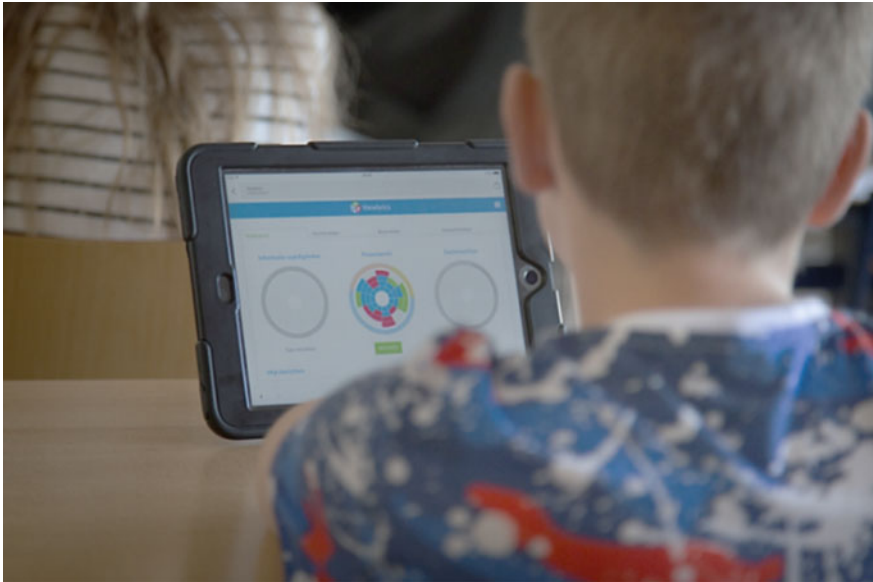
Fig. 10.7   Complex skill growth visualization on dashboard

to determine where to focus on during their next practice session (Fig. 10.8). This information becomes part of their formative assessment report of one specific assessment moment (M1) in time, to be used and referred to for future practice and which can be compared to a latter practice session and performance.

Fig. 10.8   Description of skills' learning objectives for next skills practice session

10.2   Method

To determine the effect of using the Viewbrics technology-enhanced formative assessment method with video-enhanced or textual rubrics on the mental models, (perceived) feedback and skills performance of students, two secondary pre-university education schools used the method for 24 weeks (Ackermans, 2019; Rusman et al., 2019). This study took place within the context of project-based education, with students and teachers in six low-secondary school classes (two classes using video-enhanced rubrics, two classes using textual rubrics and two classes as a control group), to compare with existing educational practice for skills acquisition. A mixed method approach was chosen, in which both quantitative and qualitative data (interviews) were combined, using and combining results from various research instruments, such as concept maps (as a representation of a mental model), rubric scores, written tips and tops, questionnaires and (focus group) interviews. A time-series approach (Field, 2009) for data collection was adopted for detecting differences in the measurement of mental models and skill performance of students. Data were analyzed by means of a test for the practical equivalence of the development models of both experimental and control groups (Ackermans, 2019; Ackermans et al., 2019a, 2019b; Kruschke, 2018; Rusman et al., 2019).

10.2.1 Sample

This study was carried out in an ecological manner and therefore used a convenience sample. Each participating school had one class using video-enhanced rubrics, one using textual rubrics within the technology-enhanced formative assessment method and one control group (skills acquisition education as usual). Participating students were between 12 and 14 years old. In total 153 students and four teachers participated.

10.2.2 Instruments

The change in mental models of the three cross-curricular complex skills was measured via a quantification of the 'richness' of concept maps. A concept or mind map is an external graphic representation of a mental model, derived from the learner's self-generated concepts (Ackermans et al., 2019a; Dhindsa et al., 2011). A rich mental model is rich in concepts (multitude of concepts), has a linear structure, contains hierarchies and a multitude of complex relationships (Besterfield-Sacre et al., 2004; Buzan, 2003; Novak & Gowin, 1985). We used the number of concepts in the concept map as an indicator for the width of the mental model, determined the depth of the mental model by looking at the structure of the concepts and the number of hierarchies and determined the strength of a mental model by counting the number of explained and unexplained relationships between concepts and different segments of the concept map (Ackermans et al.,

2019a; Besterfield-Sacre et al., 2004). These indicators were part of the scoring instrument that we used for mental model richness (Evrekli et al., 2010; Van Beek-Sweep, 2018). The quality of the feedback was determined with a self-developed instrument. This instrument performs a quantitative analysis of (overlap in) word use between the feedback given (in tips and tops) and the text of the rubrics (Ackermans et al., 2021b; Hirschberg & Manning, 2015). Additionally, interviews were carried out with students. The mastery of a skill was determined via an average rubric score (self-, peer-, expert assessment) of a student's performance on this skill (Ackermans et al., 2021a).

10.3   Results

The specific data and results were presented in the Dutch end report of the Viewbrics research project (Rusman et al., 2019) and in a PhD thesis (Ackermans, 2019). We here report and summarize the overall obtained research results. When using the technology-enhanced formative assessment method for the acquisition of cross-curricular complex skills for students in lower secondary education, we obtained the following results (Ackermans, 2019; Rusman et al., 2019):

- Students in both experimental groups performed significantly better in the three cross-curricular complex skills compared to the control group. This effect of the structured Viewbrics technology-enhanced formative assessment method with peer feedback is therefore independent of the modality of the rubrics (textual or video-enhanced) (Ackermans et al., 2021a).
- Students in the video-enhanced rubric settings developed a significantly richer mental model of collaboration and information skills compared to the control group (Ackermans et al., 2021b). This effect of the technology-enhanced formative assessment method is therefore dependent on the modality of the rubrics (video-enhanced). There was no significant difference in the mental model for presentation between the experimental and control groups. Possibly this could be due to the fact that the starting mastery level of students for this skill was initially already higher, so that less "growth" in mental models could be achieved.
- Compared to textual rubrics, applied within the technology-enhanced formative assessment method, the video-enhanced rubrics did not lead to a significant improvement of mental models and performance of collaboration, information literacy and presentation skills (Ackermans et al., 2019a).
- The video-enhanced rubrics, applied within the technology-enhanced formative assessment method, resulted in significantly higher feedback quantity of tips and tops, compared to textual rubrics (Ackermans et al., 2021b). However, feedback quality and consistency of the remarks within the tips and tops were not significantly improved.

10.4   Discussion

Based on various design principles, derived from educational theory on formative assessment, skills acquisition and (peer) feedback, we expected that the Viewbrics (technology-enhanced) formative assessment method would improve (i) the mental model of, (ii) the feedback on, and (iii) the performance on a (cross-curricular) complex skill among secondary school students when compared to existing educational practice. We also looked whether the format of the rubrics used within the method (video-enhanced or text-based) would affect learning outcomes and feedback. Looking at the effectiveness of the Viewbrics technology-enhanced formative assessment method, combining self-, peer- and expert assessment with analytic rubrics for the acquisition of complex generic skills, this study yielded affirmative research results. Based on previous studies on supporting formative assessment with written (self-, peer- and expert) feedback, we expected that the Viewbrics method would support students' skills performance and growth, which it indeed did. This effect was independent of the rubric format. Furthermore, students in the video-enhanced rubric group developed richer mental models compared to existing educational practice, however this effect was insignificant compared to use of the Viewbrics method with text-based rubrics. It seems that mainly the use of the Viewbrics technology-enhanced formative assessment method with (self-, peer- and expert-) feedback by means of rubrics, independent of the format, supported students' skills acquisition. Furthermore, feedback quality and consistency were also independent of rubric format (video-enhanced or text-based), although feedback quantity increased in the video-enhanced setting.

This study has a number of limitations: first, we have implemented the technology-enhanced formative assessment method at a limited number of secondary schools, with a limited number of students and teachers. This may have consequences for the applicability and the generalization of measured effects in other educational settings. Additionally, we had a limited time-frame for implementation (24 weeks, 16 effective lesson weeks) of the method. Perhaps if the method had been used for a longer period, with more (regular) practice moments in multiple classes, this study would have yielded different results. A final limitation is that the video modelling examples of the video-enhanced rubrics were developed only for the highest skill performance level. Perhaps several video modeling examples for different skill levels or multiple examples for one performance level description would have had a different effect. Furthermore, the development of video-enhanced rubrics is time-and cost intensive, which has to be considered. However, looking at previous studies, one might expect that a video-enhanced rubric, combining video modeling examples with a text-based analytic rubric, can have an added value for learning skills, compared to a text-based rubric only (Rohbanfard & Proteau, 2013; Van Gog et al., 2014). Therefore, it is still worthwhile to explore effects of alternative implementations on students' complex skills acquisition in future research.

Although there is research available on (technology-enhanced) formative assessment, the use of rubrics, modelling examples and the use of multimedia for

learning respectively, research on the combination of these concepts to learn complex skills and design specific process support is rare. This study contributed both by the design of video-enhanced rubrics, as by exploring its effects. Moreover, Dutch secondary education is in the process of a transformation, where generic complex skills receive more emphasis and are integrated with learning and applying domain-specific knowledge. The Viewbrics technology-enhanced formative assessment method could be(come) one of the instruments providing teachers with structure to deal with this change in their daily educational practice.

10.4.1 Implications for Practice

This project yielded, in addition to jointly (with stakeholders) developed scientific and practical knowledge about the use of video-enhanced rubrics with video (modelling) examples within a technology-enhanced formative assessment method for the development of skills, a technology-enhanced formative assessment method that has proven to be effective in educational practice in secondary schools, supported with the digital Viewbrics tool. This digital formative assessment tool, with standardized and structured 360-degree feedback and reflection process support, was evaluated (by stakeholders) as effective, usable and user-friendly. It can save time, but also paper, when using rubrics in formative assessments. Moreover, ecologically validated (textual and video-enhanced) rubrics and video-modeling examples were developed for three skills (collaboration, presentation, information literacy skills), which are reusable for other secondary schools. Instruction and workshop material, manuals and various information videos were also developed.

10.5   Conclusion

Based on this study, we can conclude that the structured Viewbrics technology-enhanced formative assessment method with (self-, peer- and expert-) feedback supported via analytic rubrics led to richer mental models and increased skill performance, independent of a video-enhanced or textual rubric format. Moreover, video-enhanced rubrics led to more feedback quantity (tips/tops), however feedback quality (concreteness/consistency) was not improved. In this study, it seems that the technology-enhanced structured 'step-by-step' process support for formative assessment and feedback with rubrics caused the mayor impact on skills acquisition of students, not the format of the rubrics. However, compared to the control group, video-enhanced rubrics did make a difference in the mental model formation (richness of model) for two skills, probably dependent on the initial performance level before practice was started.

Therefore, further and future research is needed to determine whether alternative formats would alter the effectiveness of video-enhanced rubrics within the technology-enhanced formative assessment method (e.g. with video-modeling examples available for more than one performance level description within a

rubric, or alternative examples at each subskill), compared to textual rubrics. Moreover, further research is needed to determine whether and how this technology-enhanced formative assessment method impacts students' skills acquisition at different educational levels and contexts, and for various types of skills. Design-based research is needed to see whether theory-and practice informed adaptations to the method are necessary, to make learning skills even more effective, efficient (e.g. impact on teachers' guidance and support time) and attractive in various educational practices.

References

Ackermans, K., Rusman, E., Brand-Gruwel, S., & Specht, M. (2017). Designing a video enhanced rubric to foster the development of complex skills. In E. Ras (Ed.), *Proceedings of the 20th International Conference on Technology Enhanced Assessment (TEA)*, Revised Selected Papers. October 5–6, Barcelona, Spain.

Ackermans, K., Rusman, E., Nadolski, R., Specht, M., & Brand-Gruwel, S. (2019a). Video-or text-based rubrics: What is most effective for mental model growth of complex skills within formative assessment in secondary schools? *Computers in Human Behavior, 101*, 248–258. https://doi.org/10.1016/j.chb.2019.07.011

Ackermans, K., Rusman, E., Brand-Gruwel, S., & Specht, M. (2019b). Solving instructional design dilemma's to develop a video enhanced rubric with modeling examples to support mental model development of complex skills: The Viewbrics-project use case. *Educational Technology Research and Development (ETRD), 67*(4), 983–1002. https://doi.org/10.1007/s11423-019-09668-1

Ackermans, K., Rusman, E., Nadolski, R. J., Brand-Gruwel, S., & Specht, M. M. (2021a). Video-enhanced or textual rubrics: Does the Viewbrics' formative assessment methodology support the mastery of complex (21st century) skills? *Journal of Computer Assisted Learning, 37*(3), 810–824. https://doi.org/10.1111/jcal.12525

Ackermans, K., Rusman, E., Nadolski, R. J., Specht, M. M., & Brand-Gruwel, S. (2021b). Feedback is a gift: Do Video-enhanced rubrics result in providing better peer feedback than textual rubrics? *Practical Assessment, Research & Evaluation, 26*(1), [17]. https://scholarworks.umass.edu/pare/vol26/iss1/17/

Ackermans, K. (2019). *Designing video-enhanced rubrics to master complex skills* [thesis]. Open Universiteit. https://research.ou.nl/en/publications/designing-video-enhanced-rubrics-to-master-complex-skills

Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation, 10*(3), 1–11.

Arter, J. A., & Chappuis, J. (2006). *Creating and recognizing quality rubrics*. Pearson.

Baecher, L., Kung, S. C., Jewkes, A. M., & Rosalia, C. (2013). The role of video for self-evaluation in early field experiences. *Teaching and Teacher Education, 36*, 189–197.

Berry, O., Price, M., Rust, C., Donovan, B.O., Price, M., & Rust, C. (2007). Teaching in higher education know what i mean? Enhancing student understanding of assessment standards and criteria. *Teaching in Higher Education* 37–41.

Besterfield-Sacre, M., Gerchak, J., Lyons, M. R., Shuman, L. J., & Wolfe, H. (2004). Scoring concept maps: An integrated rubric for assessing engineering education. *Journal of Engineering Education.* https://doi.org/10.1002/j.2168-9830.2004.tb00795.x

Black, P. J., & Wiliam, D. (1998). Inside the Black Box: Raising standards through classroom assessment. *Phi Delta Kappa, 80*, 139–48 (1998)

Black, P. J., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly the Journal of Personal Evaluation in Education), 21*(1), 5–31.

Boud, D. (2001). Peer learning and assessment. In D. Boud, R. Cohen, & J. Sampson (Eds.), *Peer learning in higher education* (pp. 67–84). Kogan Page Limited.

Bowman, K. (2010). *Background paper for the AQF council on generic skills.* Canberra, Australia.

Buzan, T. (2003). The mind map book. In *The mind map book how to use radiant thinking to maximize your brains untapped potential.* Article-id: 960994.

Candy, P., Crebert, G., & O'Leary, J. (1994). Developing lifelong learners through undergraduate education. In *NBEET, Commissioned Report,* (Vol. 28). Australian Government Publishing Service

De Grez, L., Valcke, M., & Roozen, I. (2013). The differential impact of observational learning and practice-based learning on the development of oral presentation skills in higher education. *Higher Education Research and Development, 33*(2), 256–271. https://doi.org/10.1080/07294360.2013.832155

De Grez, L., Van Steendam, E. & Raedts, M. (2014). If they can, I Can! Enhancing Learning through (Peer) Modeling in Instructional Video. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2014* (pp. 741–745). Association for the Advancement of Computing in Education (AACE).

Dhindsa, H. S., Makarimi-Kasim, & Anderson, O. R. (2011). Constructivist-visual mind map teaching approach and the quality of students' cognitive structures. *Journal of Science Education and Technology.* https://doi.org/10.1007/s10956-010-9245-4

Evrekli, E., Inel, D., & Balim, A. G. (2010). Development of a scoring system to assess mind maps. *In Procedia—Social and Behavioral Sciences.* https://doi.org/10.1016/j.sbspro.2010.03.331

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). SAGE Publications.

Filius, R. M. (2019). *Peer feedback to promote deep learning in online education: Unravelling the process* [thesis]. Universiteit Utrecht. https://doi.org/10.13140/RG.2.2.28814.82247

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science.* https://doi.org/10.1126/science.aaa8685

Kirschner, P., & Merriënboer, J. Van (2008). Ten steps to complex learning a new approach to instruction and instructional design. In T. L. Good (Ed.), *21st Century education: A reference handbook* (pp. 244–253). Sage.

Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science.* https://doi.org/10.1177/2515245918771304

Looney, J., & Siemens, G. (2011). Assessment competency: Knowing what you know and learning analytics. It is time for a breakthrough. *Promethean Thinking Deeper Research Paper*, *3.*

McManus, S. (2008). *Attributes of effective formative assessment.* CCSSO.

Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment Research Evaluation, 7*(25), 1–10. Retrieved from http://www.pareonline.net/getvn.asp?v=7&n=25

Nadolski, R. J., Hummel, H. G. K., Rusman, E., & Ackermans, K. (2021). Rubric formats for the formative assessment of oral presentation skills acquisition in secondary education. *Educational Technology, Research and Development, 69*(5), 2663–2682. https://doi.org/10.1007/s11423-021-10030-7

Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment and Evaluation in Higher Education, 35*(5), 501–517.

Norman, D. (2013). *The design of everyday things.* Ingram Publisher Services US.

Novak, J. D., & Gowin, D. B. (1985). Chapter 2: Concept mapping for meaningful learning. *Learning How to Learn.*

O'Donovan, B., Price, M., & Rust, C. (2004). Know what I mean? Enhancing students understanding of assessment standards and criteria. *Teaching in Higher Education, 9*(3), 325–335.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review, 9*, 129–144. https://doi.org/10.1016/j.edurev.2013.01.002

Rohbanfard, H., & Proteau, L. (2013). Live vs.video presentation techniques in the observational learning of motor skills. *Trends in Neuroscience and Education, 2*, 27–32. https://doi.org/10.1016/j.tine.2012.11.001

Rusman, E., Nadolski, R., & Ackermans, K. (2019). *Viewbrics, 'spiegel' je vaardig: Vakoverstijgende vaardigheden aanleren in het voortgezet onderwijs via een (online) formatieve evaluatie-methode met (video-verrijkte) rubrieken.* Open Universiteit.

Rusman, E., & Dirkx, K. (2017). Developing rubrics to assess complex (Generic) Skills in the classroom: How to distinguish skills' mastery levels? *Practical Assessment, Research & Evaluation (PARE), 22*(12). Available online: http://pareonline.net/getvn.asp?v=22&n=12 and http://dspace.ou.nl/handle/1820/8802

Rusman, E., Boon, J., Martínez-Monés, Rodríguez-Triana, & Retalis, S. (2013). Towards the use of new methods for formative e-Assessment of 21st century skills in schools. *Paper in Technology-Enhanced Formative Assessment (TEFA) Workshop by the European Association of Technology-Enhanced Learning (organized by EA-TEL), at 8th European Conference on Technology Enhanced Learning (ECTEL), 17–18 September 2013*, Paphos, Greece.

Rusman, E., Martínez-Monés, A., Boon, J., Rodríguez-Triana, M. J., & Villagrá-Sobrino, S. (2014, 30 June-1 July). Gauging teachers' needs with regard to technology-enhanced formative assessment (TEFA) of 21st century skills in the classroom. In M. Kalz, & E. Ras (Eds.), *Computer assisted assessment. Research into e-Assessment. Proceedings of the international computer assisted assessment (CAA) in conference communications in computer and information science proceedings 439* (pp. 1–14). Springer International Publishing. https://doi.org/10.1007/978-3-319-08657-6_1

Rusman, E., Firssova, O., Janssen, T. & Specht, M. (2015). An integrated approach to inquiry based science learning in a secondary school: Designing a colony on Mars. In *Proceedings of 9th International Technology, Education and Development Conference, 2–4 March, 2015* (pp. 4483–4492). Madrid, Spain. http://dspace.ou.nl/handle/1820/6135

Rusman, E., Nadolski, R., & Ackermans, K. (2018). Students' and Teachers' perceptions of the usability and usefulness of the first viewbrics-prototype: A methodology and online tool to formatively assess complex generic skills with video-enhanced rubrics (VER) in dutch secondary education. In *Paper at the 21th International Conference on Technology Enhanced Assessment (TEA)*, December 10–11, 2018, Amsterdam, The Netherlands.

Schildkamp, K., Heitink, M., Kleij, F. Van Der, Hoogland, I., Dijkstra, A., & Kippers, W. (2014). *Voorwaarden voor effectieve toetsing—Een praktische review.* Universiteit Twente.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.

Sluijsmans, D. M. A., Joosten-ten Brinke, D., & Van der Vleuten, C. P. M. (2013). *Toetsen met leerwaarde. Een reviewstudie naar effectieve kenmerken van formatief toetsen.* NWO.

Thijs, A., Fisser, P., & Hoeven, M. van der (2014). *21e eeuwse vaardigheden in het curriculum van het funderend onderwijs.* SLO.

Van Beek-Sweep, E. (2018). *Ontwikkeling en validatie van een analyse-instrument voor het beoordelen van de kwaliteit van mindmaps binnen het Viewbrics project* (Master's thesis). Open University of the Netherlands. Retrieved from http://dspace.ou.nl/handle/1820/9932 .

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review, 22*(2), 155–174. https://doi.org/10.1007/s10648-010-9134-7

Van Gog, T., Verveer, I., & Verveer, L. (2014). Learning from video modeling examples: Effects of seeing the human model's face. *Computers and Education, 72*, 323–327.

Van Merriënboer, J. J. G., & Kirschner, P. A. (2017). *Ten steps to complex learning: A systematic approach to four-component instructional design.* Routledge.

Van Strien, J., & Joosten-ten Brinke, D. (2016). Het beoordelen van de kwaliteit van rubrics. *Examens, 1*, 33–37.

Vincent-Wayne, M., & Bakewell, C. (1995). Learning without doing: Enhancing oral presentation skills through peer review. *Management Learning, 26*(3), 353–366. https://doi.org/10.1177/1350507695263005

Voogt, J., & Pareja-Roblin, N. P. (2012). A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies. *Journal of Curriculum Studies, 44*, 299–321.

Westera, W. (2011). On the changing nature of learning context: Anticipating the virtual extensions of the world. *Educational Technology & Society, 14*(2), 201–212.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166–183.

Part IV
Empirical Contributions on Peer Learning

PeerTeach: Teaching Learners to Do Learner-Centered Teaching

11

Soren Rosier

11.1 Introduction: Prior Studies, Research Questions, and Significance

The potential of peer tutoring is boundless if peer tutors have the content mastery and tutoring skills to even remotely resemble effective adult tutors. High-dosage tutoring with trained adult tutors is consistently identified as the most productive learning intervention, including among students with low socio-economic status (Dietrichson et al., 2017; Fryer, 2017). Unfortunately, to date, there is little evidence that K-12 students can be quickly trained to teach well (Berghmans et al., 2013).

Studies consistently find that tutors tend to do much more explaining than tutees (King, 1997), place minimal demand on tutees when questioning (Graesser et al., 1995), and rarely stimulate deep-level reasoning or monitor the understanding of tutees (Graesser et al., 1995; Roscoe & Chi, 2007). In short, tutors tend to adopt stereotypical, didactic teaching practices, cutting off opportunities for tutees to actively engage with ideas, sometimes severely hampering their learning. Drawing from in-depth observations of peer helping in middle school classrooms, Webb and Mastergeorge (2003) found that receiving highly didactic help actually predicted poorer content understanding than being left alone to struggle. I have come to label these the *common sins of the Default Didact.* Thus, while recent meta-analyses have found that peer tutoring does significantly increase learning for both tutors and tutees (Bowman-Perrott et al., 2013; Kobayashi, 2019; Leung, 2015), the efficacy

S. Rosier (✉)
Stanford University, 485 Lasuen Mall, Stanford, CA 94306, USA
e-mail: soren@peerteach.org

239

of this learning arrangement is limited by our ability to effectively train peer tutors (Topping et al., 2017).

Few prior studies have attempted to train students to overcome these *common sins of the Default Didact*, and with minimal success. King's (1998) model, ASK to THINK—TEL WHY, is an example of a program that trains students to ask questions. This is a reciprocal model where students take turns as the "questioner" or "explainer" following a whole-class lesson. Questioners ask a series of five types of questions using a card with question prompts. Emblematic of this under-researched area, the one experimental study of this model was underpowered, with three groups of just ten dyads. It found suggestive evidence that students using this structured inquiry model improved their ability to make inferences based on class content, but they did not comprehend class content better.

In this article, we define learner-centered peer tutoring similarly to learner-centered teaching, which emphasizes learners actively participating and constructing their own knowledge, as opposed to passive knowledge transmission (Yeh & Swinehart, 2017). Berghmans et al. (2013) attempted to train advanced college math students to adopt learner-centered peer tutoring strategies. Their training lasted 90 minutes, incorporating an overview on facilitative strategies (mainly questioning and hinting) and opportunity for tutoring roleplay with feedback. They then analyzed the instructional moves used by tutors in an introductory math class and interviewed them to better understand the rationales for their decisions. They rigorously evaluated the impact of their training and found that it did not meaningfully shift the behaviors of peer tutors. In line with past findings and despite the preparation to be more facilitative, tutors inevitably inclined toward directive strategies and "knowledge-telling," and their questioning was "low level and shallow" (p. 717). The authors concluded that novice tutors require extensive training on deep-level questioning, working with tutees of varying levels, and reshaping beliefs about learning.

To address this persistent challenge, I designed a study to test the efficacy of two different interactive online training approaches to increase tutors' use of learner-centered teaching behaviors and promote tutee learning. One approach was prescriptive, telling subjects the exact learner-centered pedagogical behaviors to use then prompting practice in identifying and executing them; the other approach assumed that students inherently possess productive pedagogical notions that must be strategically unearthed and committed to in writing. This comparison intentionally mirrored the classic tension between direct instruction and constructivist approaches to learning new skills. Specifically, this study asked:

1. can short, interactive, online modules prescribing key learner-centered pedagogical strategies shift middle schoolers' tutoring behaviors?
2. can short, interactive, online modules embedded with social psychological intervention strategies unearth dormant learner-centered pedagogical inclinations and shift middle schoolers' tutoring behaviors?
3. can increased adoption of learner-centered tutoring behaviors through either intervention approach increase learning for tutees?

Structures for group learning and Peer Assisted Learning (PAL), which includes peer tutoring, have been studied extensively by numerous researchers, perhaps most prominently by Slavin (2006), who co-developed three common structures: Student Teams-Activity Division, Teams-Games-Tournaments, and Cooperative Integrated Reading and Composition. Despite myriad structures and ample scholarship on their implementation and efficacy, there are few evidence-based models for training K-12 students on *how* to effectively communicate during group learning. Training for peer tutoring—the most obvious and common form of PAL (Topping & Ehly, 2001) where one student actively supports the academic learning of a peer—should be informed by the mass of accumulated knowledge on teacher professional development, but these connections are rarely made. This research project attempted to bridge this gap by transposing the framework of learner-centered teaching onto peer tutoring, and testing the viability of effective training through a web application.

The results of these studies provide strong evidence for the prevalence of the *Default Didact* and the realistic possibility of tutors becoming what I call *Emergent Elicitors*. The Default Didact, though often well-meaning, treats teaching and helping opportunities as opportunities to lecture and demonstrate competence, embodying and mirroring years of being *spoken at* by teachers. As Lortie postulated about novice teachers, this default didact too is a product of the "apprenticeship-of-observation" (1975, p. 67). These studies suggest, however, that this default state is not as sticky for peer tutors as its prevalence among adult teachers might imply.

11.2 Prescriptive Intervention Design to Promote Three Learner-Centered Tutoring Strategies

This study aimed to discover ways to quickly train students to be learner-centered tutors capable of eliciting, probing, and guiding the thinking of peers in much the same way that effective teachers do. The hope was that, after just 40 minutes interacting with either PeerTeach training—a short enough duration to fit within one class period—students would be able to more effectively teach their peers. While the goal of both trainings was to promote learner-centered tutoring, their structures were distinct, testing the comparative affordances of a prescriptive training approach versus a more constructivist one. The Talk Moves training provides students with proven teaching strategies then offers an online environment in which to practice identifying and using them.

Talk moves (at times called "talk tools" or "accountable talk") are the result of three decades of research aimed at identifying the speaking choices of teachers who are skillful at orchestrating equitable and productive classroom discourse (Godfrey & O'Connor, 1995; O'Connor, 2001; O'Connor & Michaels, 1993, 2015). Among the teacher professional development efforts to increase and improve teacher questioning, this approach is among the most specific, practical, and easy to grasp.

From the teacher talk moves described in this literature, a subset of moves were identified that are ideal for peer tutoring as they are conceptually simple, broadly applicable, and intended for one-on-one interactions. These include (1) eliciting questions that encourage students to express their ideas (e.g., "Say more about that"), (2) probing questions that dig into why students think what they think (e.g., "Why do you think that?), and (3) revoicing moves where tutors state what they think the learner is saying (e.g., "I hear you saying _____").

There are two main ways that these three talk moves promote learning. First, *eliciting* and *probing* moves encourage tutees to talk, which forces them to make sense of their thoughts in order to verbalize them. It is common for this alone to help learners work through ideas and develop solutions on their own (King, 1998; Webb & Mastergeorge, 2003). At minimum, eliciting and probing push students to take stock of what they do or do not know at any given moment and make them active participants in knowledge creation. Second, all three talk moves enable tutors to better understand their peers, helping them to identify misconceptions, gaps in knowledge, and errors in reasoning, preparing them to scaffold learning more effectively.

In their study of the Talk Science intervention, Michaels and O'Connor (2015) found that their training quadrupled the frequency that nine teachers used language that video-coders perceived to be "helping students deepen their reasoning" (p. 343). This success in uptake of moves is likely a product of talk moves being "easy to remember and easy to pull out with a bit of practice" (p. 336), making them practical and realistic tutoring techniques for children. Thus, it stands to reason that preparing children to use eliciting, probing, and revoicing talk moves could be an effective way to shift students from what is typical didactic tutoring to more elicitive strategies that promote better dialogue and deeper learning.

11.2.1 Design

The first PeerTeach intervention focuses on *Talk Moves* and uses Sherin and Van Es's (2005) video-based *noticing framework* as a vehicle for promoting their uptake. That framework asserts that those who teach must attend to important teaching moments, relate them to useful pedagogical frameworks, and act based on pedagogically sound reasoning. PeerTeach creates such experiences when students watch animated tutoring interactions and practice noticing and tagging effective talk moves (see Fig. 11.1 for an example of this type of PeerTeach level). The theory driving this intervention is that if students are trained to notice and identify effective talk moves, they might internalize and use them in real-world tutoring interactions.

Figure 11.1 shows the intersection of a curated set of talk moves with the first two elements of Sherin's (2005) Noticing Framework for professional development: attending to important teaching moments and relating them to useful pedagogical frameworks. To accomplish the third and final element of that framework—acting based on pedagogically sound reasoning—PeerTeach has students
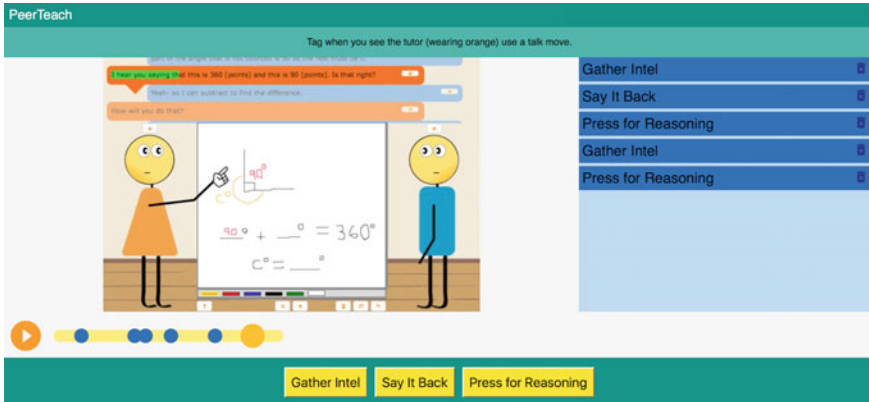
Fig. 11.1 Noticing practice level. *Note.* Students tag the video each time the cartoon tutor uses one of the focal talk moves

practice making teaching decisions. Within the application, students engage in virtual tutoring sessions where they practice selecting the most strategic utterance (of three) to propel a virtual student forward. After selecting an utterance, students receive two forms of feedback: (1) the virtual learner responds verbally, revealing the impact of the selected utterance, and (2) the learn-o-meter, an indicator of the virtual student's thinking, goes up or down. See Fig. 11.2 for an example of this type of level.

Great tutoring, like great teaching, involves a complicated set of processes. While some peer tutoring models restrict tutors to solely asking questions (King, 1998), this training simply encourages their inclusion.

11.3 Constructivist Intervention Design to Unearth Learner-Centered Tutoring Strategies

The first intervention was driven by the theory that students (1) lack useful pedagogical intuitions, (2) should be directly told what constitutes effective teaching, and (3) need practice using those learner-centered techniques. The second intervention was premised on the idea that students intuitively possess productive notions of learner-centered teaching—that students believe, either innately or through experience as learners, that learning happens best when the learner is engaged, actively verbalizing thoughts, and in a dialogic back-and-forth with a responsive, question-asking guide. This intervention gives students mild priming to make salient their existing conceptions of learner-centered teaching, then prompts them to describe the helper they want to be in a letter to themselves. It is modeled after "wise" interventions from the social psychological literature, in particular, the Saying is Believing intervention strategy (Aronson et al., 2002). This particular intervention technique has proven very effective in prompting psychological
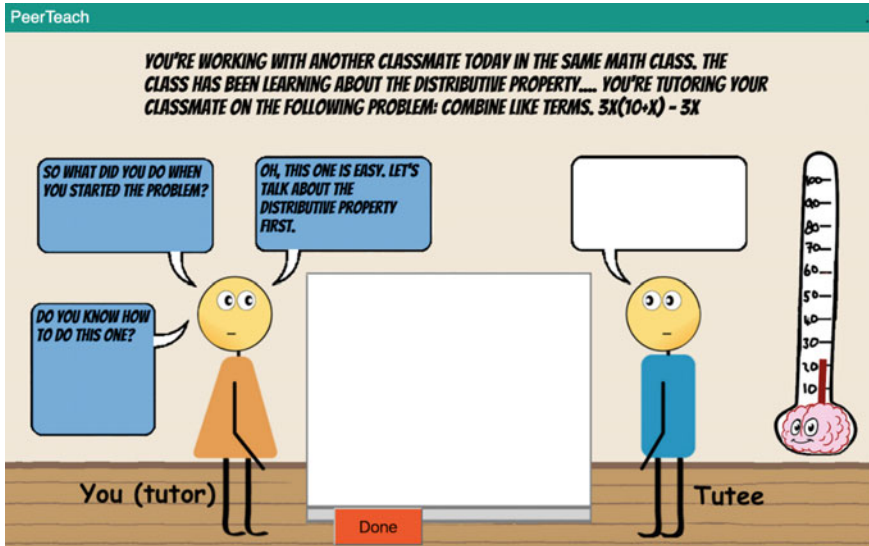
Fig. 11.2 Practice level for choosing evidence-based teaching moves. *Note.* Students practice making strategic teaching decisions. The Learn-o-meter ticks up when the virtual tutee is learning

shifts in other areas: to believe intelligence is malleable (Aronson et al., 2002) and to believe they belong in college (Walton & Cohen, 2011) to name two of many.

Aronson argues that people want to be consistent. If they are prompted to write that learner-centered teaching behaviors are key to good tutoring, they can only maintain consistency and avoid feeling hypocritical if they tutor accordingly. Thus, this intervention works by priming subjects to write down that they believe good tutoring is about asking questions, understanding the other person, and encouraging that person to do the thinking work. In this way, the Wise intervention approach more closely resembles discovery learning, which assumes and calls forth prior knowledge as a central component of learning.

11.3.1 Design

Through the PeerTeach web application, students who engage with this intervention take notes while watching a series of videos. The first two videos (each approximately one minute long) show a compilation of interview clips where experienced peer-tutors discuss the lessons they have learned (shown in Fig. 11.3). These clips are curated to reinforce specific messages: tutees need to be actively problem solving and tutors need to be asking questions and probing the other students' thinking. Those brief videos are followed by videos of example tutoring sessions, illustrated by Fig. 11.4. While they are not marked "good" and "bad," it is clear from extensive user testing that students intuitively pick up on one tutor
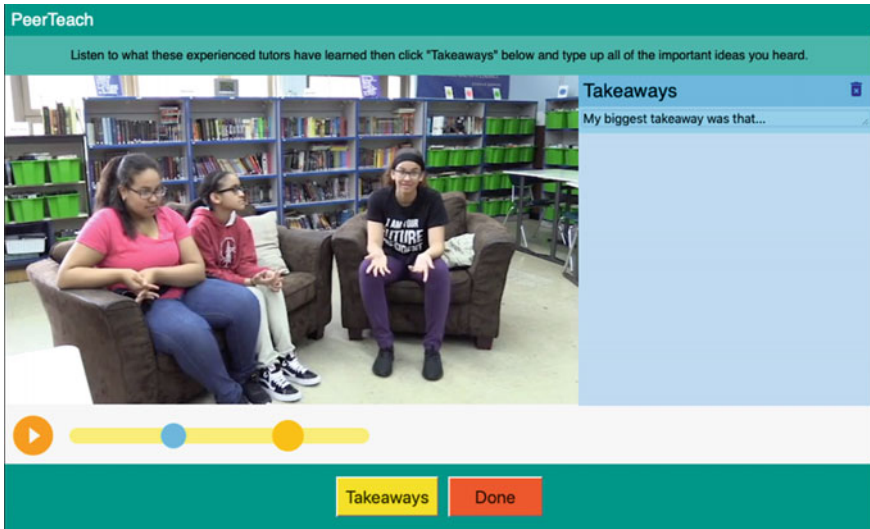
Fig. 11.3 Priming interviews on PeerTeach. *Note.* Peer tutors discuss lessons learned, focusing on learner-centered strategies

dominating the conversation and explaining too much while a different tutor asks questions that help the other student think through a problem. Pilot testing showed that students make this discovery themselves; past research has shown that learning can be longer lasting when students make discoveries themselves, even through computer simulations (De Jong & Van Joolingen, 1998). After watching videos and taking notes, students write a letter to themselves about the kind of helper they want to be, tacitly committing to enacting those behaviors in the real world.

11.4 Methods

These studies took place in a Northern California middle school in partnership with one sixth and one seventh grade teacher. They were conducted with 198 sixth and seventh graders in regular, non-advanced math classes. The students were 53% Latino and 42% White at a school where 33% of students are eligible for free or reduced-price lunch.

11.4.1 Round One Implementation Sequence

In both rounds of data collection, which were separated by five months, students first engaged in training to become effective helpers, then employed their new skills in real teaching interactions with peers. Students in each of seven class-rooms were randomly assigned to one of three conditions: the wise psychological
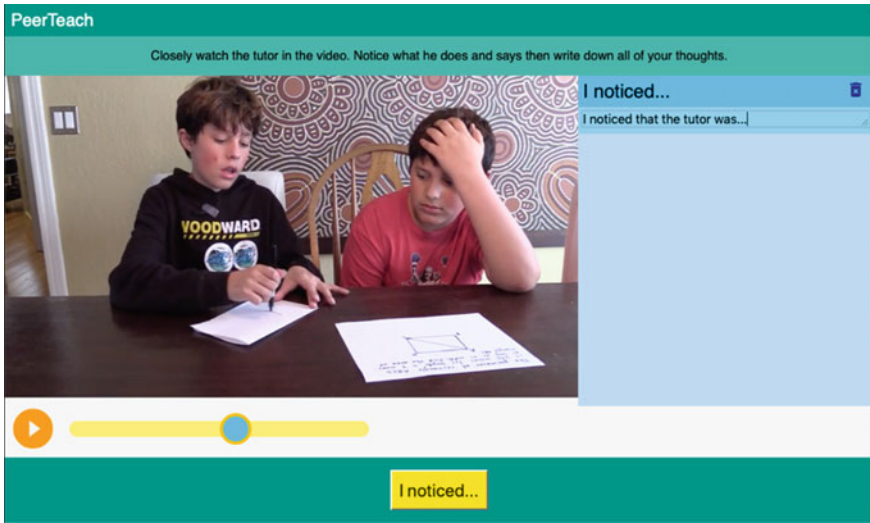
Fig. 11.4 Contrasting cases videos on PeerTeach. *Note.* Students watch contrasting tutoring videos, identifying the problematic nature of overly didactic teaching and the learning benefits of more elicitive strategies

intervention, the Talk Moves (TM) Training, and the control condition. To minimize classroom effects, randomization occurred within each classroom. Students received the same training in both studies, so round two can be considered a *re-dosing* of treatment.

The first round of data collection was underpowered for detecting learning differences by tutee condition, since only half of the students were tutees. The main aim was to validate that the interventions could successfully shift students' online tutoring inclinations from didactic knowledge-telling to more learner-centered approaches. Significant learning differences following in-person tutoring, by condition, was an aspirational outcome, not an expected one.

11.4.1.1 Day 1—Determining Baseline Content Understanding and Tutoring Inclinations

To measure students' a priori inclinations toward didactic helping versus elicitive helping, all students in this study—those in both treatment conditions, along with the control students—began their intervention experience making teaching decisions in an online game. On this level, each student individually controlled a virtual peer tutor helping a virtual cartoon learner. For each of four scenarios, students were presented with three speech options: one learner-centered teaching move and two didactic (or overly directive) options that shut down opportunities for the virtual learner to think. Many of these overly directive speech options were cloaked in questions (e.g., "Would you like me to show you how to solve this?") so that students could not "game" the system by just picking questions.

All students were taught a lesson on ratios then given an assessment to determine how well they learned the content. The top half of student performers were designated as tutors. To increase the likelihood that tutors in each condition would have similar tutoring ability at baseline, tutors were ranked by score on their baseline tutoring decision-making then sorted into conditions through blocked sampling (i.e., the tutors with the top 3 scores were randomly assigned into each condition, then the next three were assigned, etc.). The same blocked sampling strategy was used to assign tutees to conditions. Lastly, tutors and tutees within conditions were paired randomly.

11.4.1.2 Day 2—Training and then Tutoring

Students completed their assigned training silently on laptops sitting at desks that were spaced out in their classrooms. Following the intervention, students played a similar game with 4 new scenarios to reveal any shifts in their online teaching inclinations.

Tutoring pairs were given worksheets with practice problems. Tutors were instructed, "You can do whatever you think is best to help the other student learn." Tutoring occurred for 10 min then all students took a final assessment on ratios the following day. That assessment was scored using an adaptation of the "Representing and Solving the Task" portion of the Mathematics Problem Solving Official Scoring Guide used by the Oregon Department of Education Office of Assessment and Evaluation (2011). See Appendix A. Each of four problems was scored on a rubric of 1–4 to allow us to distinguish between degrees of mathematical understanding. The author and a research assistant scored the assessments, achieving an interrater reliability of 87.5% on 20% of the data.

11.4.1.3 Control Group

The aim in designing the control was to mimic every contextual feature of the intervention experience without actually shifting how students thought about peer tutoring. It was hoped that controls would (1) believe they were being trained as effective helpers, but (2) teach in the natural way they would have without any training. To accomplish this, controls were treated identically by facilitators, partnered with a student in the same group, and completed their training through PeerTeach. In order to avoid changing how they conceptualized peer tutoring, leaving intact their natural inclinations, this *training* focused on the importance of tutors understanding math. A prior survey revealed this belief to be nearly universal among middle school students, making it appropriate for the control "training." Thus, controls spent their training time engaged in solving math problems accessed through PeerTeach as preparation for future peer tutoring.

11.4.2 Round Two Implementation Sequence

The second round of data collection took place five months after the first, with the same group of students. It focused on two main questions: (1) do shifting pedagogical mindsets translate into measurably different teaching behaviors in real-life, particularly more learner-centered moves? and (2) do these shifts in tutoring style produce more learning for tutees? Fig. 11.5 illustrates the study design.

11.4.2.1 Day 1—Sorting by Condition and Training

Students completed the same assigned training as before through the PeerTeach website sitting next to a new randomly assigned partner in the same experimental group. The three experimental groups were clustered together with an assigned facilitator (one of two researchers or the teacher) facing away from the middle of the classroom to maintain the facade that all students were engaged in the same training. By and large, students only paid attention to their own training, minimizing the cross-pollination of ideas between treatment conditions. Only one student appeared to notice that each cluster was advancing through a different training.

While Round One showed promising training results without interaction between participants, past studies on the learning benefits of collaboration suggested that these interventions might be even more powerful if children could talk through their thinking with one another. As just one of many examples, Bamiro (2015) demonstrated that teachers could produce significant learning gains
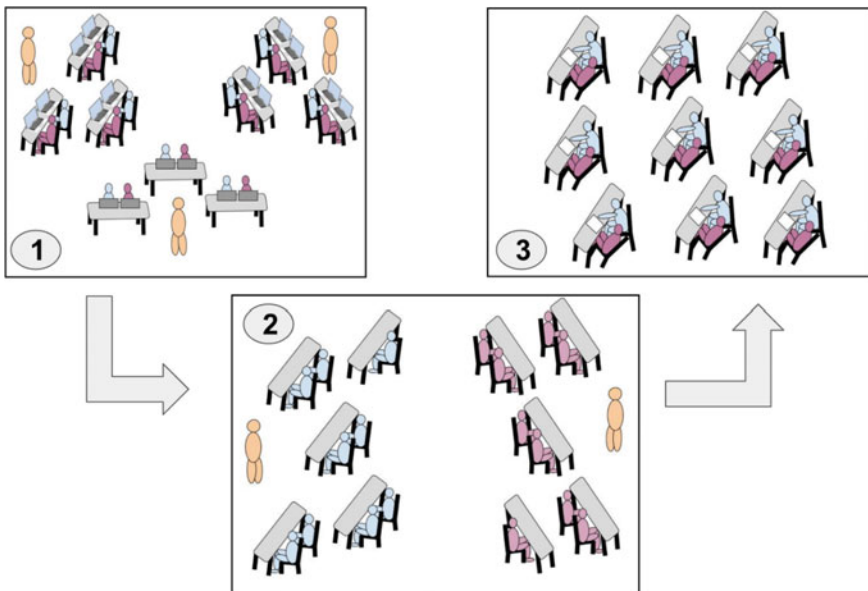

Fig. 11.5 Implementation flow of round 2

in chemistry classrooms simply by adding in think-pair-shares. As such, facilitators in Round two encouraged partner pairs to discuss the training ideas to better understand them.

The PeerTeach interventions were administered consistently, largely because students' experiences were facilitated by a computer program. To ensure that facilitators acted predictably, we collaborated to develop a facilitation script that included what we would say before students opened their laptops, along with three acceptable prompts to encourage collaboration between partners. To account for slight differences that could emerge from the presence of one facilitator instead of another, facilitators rotated between experimental groups each class period.

11.4.2.2 Day 2—Learning Different Math Content

Each class was split in half to learn different content, either comparing means and medians (taught by the researcher) or comparing rates (taught by the teacher). Partner pairs from Day 1 were split and randomly assigned to these different content groups. These topics were selected through negotiation with the two teachers. These topics—ideal for peer tutoring because they are conceptually rich with multiple solution paths—were on the pacing guide for the 6th grade teacher and were deemed important, challenging, and worth re-teaching by the 7th grade teacher. In this way, the study was built into the fabric of a legitimate learning sequence, aiming to both answer important research questions and serve the learners within the context of their classrooms. Following the Day 2 lessons, quizzes were administered to enable later examination of the relationship between tutors' content knowledge and how well their tutees learn.

11.4.2.3 Day 3—Peer Tutoring and Post Assessing

Students taught partners the content they learned the prior day. After 20 min of peer tutoring, each student wrote a reflection describing the teaching of their partner, then took an assessment to measure their learning. That assessment, like the one used in Study 1, was later scored by the author and a research assistant using an adapted version of a rubric focused on "Mathematics Problem Solving" (Oregon Department of Education, 2011). Again, problems were scored 1—4 and an interrater reliability of 83.5% was achieved on 20% of the data.

11.4.3 Measures

After both rounds of data collection, the three conditions were compared on a number of variables: the frequency that students chose elicitive teaching moves in online scenarios, tutees' assessment scores, and in Round Two, also the frequency of tutees describing particular tutoring behaviors in real life. To account for classroom differences, linear mixed-effects models were implemented from the lme4 package (Bates et al., 2015) in the statistical software R (Version 3.0.3. R Development Core Team. 2008). The primary comparisons were treated as fixed

effects while the classroom was treated as a random effect. Each dependent variable was regressed using orthogonal contrasts to test two comparisons: whether treatment conditions combined (coded as $+1/3$ each) produced more effective outcomes than the control condition (coded as $-2/3$) and whether one treatment was more effective than the other (coded as $-1/2$ and $+1/2$). Only one outlier was excluded.

One key difference between Round One and Round Two was that tutor and tutee sample sizes were doubled in Round Two because all students served as tutors, not just the top half of performers on the pre-assessment. To determine appropriate sample sizes, the most reliable method is to identify prior studies with near-identical measures to make a priori power estimates. Unfortunately, no substantive body of research exists measuring learning impacts of training K-12 peer tutors. Instead, past studies measuring the learning impacts of teacher professional development and teacher questioning were selected as the nearest analogue. Hattie (2012, p. 252) estimates the effect size of teacher questioning on student learning to be 0.48 and the effect size of teacher professional development to be 0.51. With an effect size of approximately 0.5, alpha of 0.05, and a power score of 0.80, samples should have 50 participants to perform a well-powered one-sided t-test. For this study, after removing students who were absent during any day of the study, the three samples had, on average, 52 students each. Thus, if the effects on student learning resembled prior success levels training adult teachers, this study was adequately powered to detect statistical differences.

11.4.3.1 Qualitative Measures of Tutoring Behavior

To gauge differences in tutoring behaviors post-intervention, an open-ended survey was administered immediately after peer tutoring occurred. It asked, "What was the most helpful thing your classmate did or said when teaching you? Give as much detail as you can." The author and a research assistant applied emergent codes to these responses to unearth patterns in the ways that students taught each other (and what their peers considered their best teaching moves). A codebook was developed with 13 main codes (e.g., "Asked questions" or "Checked work/understanding") and 29 sub-codes (e.g., "Asked probing questions" or "Used yes or no checks for understanding"). Codes were applied to descriptions without names or experimental conditions visible to ensure unbiased coding. The frequency of applied codes is shown in Appendix B.

To ensure accuracy, two procedures were employed, as described by Saldaña (2021, p. 27–28): a check for intercoder reliability and consensus coding on the full corpus of data. After every response was coded by both the author and a research assistant using NVIVO 11 software, a check for reliability revealed 86% overlap in applied codes, which is above the 80% threshold as recommended by Miles and Huberman (1994). Next, to ensure the accuracy of final codes, the 14% of cases with disagreement were discussed until consensus was reached. Combined, these two procedures ensured that the codebook was reliably employed and that final codes were accurate representations of the data.

11.5   Results

11.5.1 Students Default to Didactic Teaching Online, but Shift with Training

Past studies have shown that peer tutors tend to explain more than they should (King, 1997). To measure students' inclinations toward over-explaining versus more learner-centered behaviors, students made decisions in online scenarios before and after their intervention experiences. Unsurprisingly, before receiving the training, students across conditions tended to choose didactic speech options (e.g., "The first thing you need to do is…"). More surprising was the extent to which students avoided trying to elicit the virtual student's thinking. Out of four scenarios, students selected the more elicitive move only 1.04 times, on average, markedly lower than *by chance*. See Fig. 11.6 for an example of one scenario and the frequency with which students selected utterances.

When given a similar scenario-based game post-intervention, as predicted, students in both the wise intervention group (labeled "WISE Training" in plots) and the Talk Moves training ("TM Training") became more elicitive online helpers than controls ($p < 0.001$), as illustrated in Fig. 11.7. This analysis was executed using planned orthogonal contrasts to compare combined treatment groups with controls. Students in the Talk Moves Training chose elicitive moves most often, likely because their training incorporated practice making decisions in similar online scenarios, but their performance was not significantly different from students in the wise intervention condition. Compared to controls, the Cohen's D effect size was 0.95 for the Talk Moves Training and 0.63 for the Wise Training.
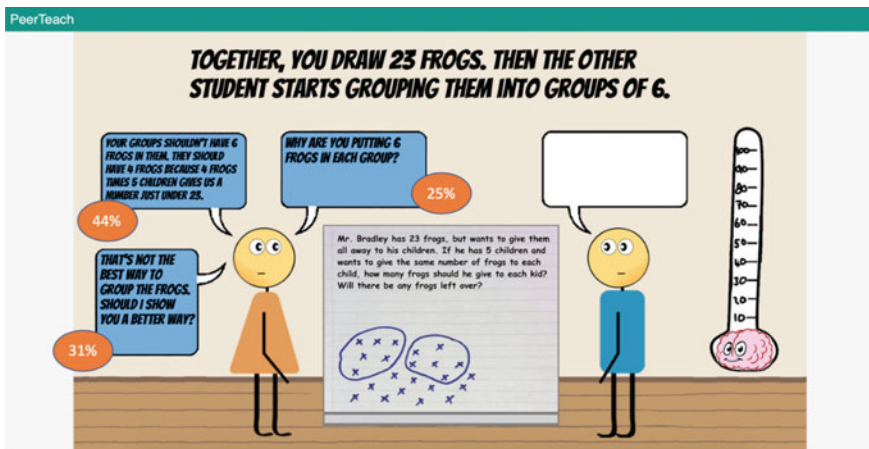

Fig. 11.6 Example teaching scenario with frequency of selected moves (pre-intervention)

Fig. 11.7 Following training, elicitive teaching moves increase in online scenarios. *Note.* Error bars represent 95% confidence intervals
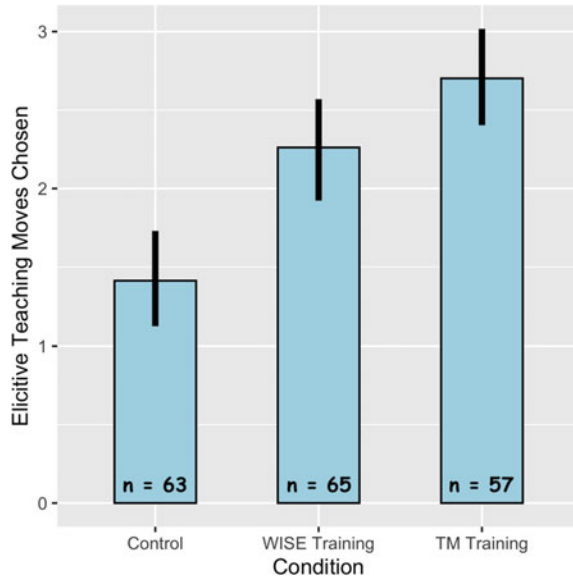
11.5.2 Learning Gains in Round 1 of Data Collection

More elicitive decision-making in online scenarios was not, however, the ultimate goal. This was an intermediate measure. The true test of the efficacy of these training approaches was how well students' training experiences translated into effective real-world teaching. As stated prior, it seemed unlikely that differential learning effects would be detected with such underpowered samples (since only half of students were tutors) and relatively short, 10-min tutoring experiences. Despite those constraints, tutees in treatment conditions did appear to learn more than controls. Using planned orthogonal contrasts to compare student groups, we find that tutees taught by tutors in treatment conditions did indeed have higher post-assessment scores than controls. To account for possible differences by teacher, a linear mixed effects model was utilized where tutee scores are treated as fixed effects while teacher was treated as a random effect. To confirm that post-assessments were not influenced by differing content mastery between intervention groups, pre-assessments were compared across groups and were not significantly different.

Post-assessment analysis suggests that treatment group tutors (combined) were more effective than controls [$F_{(1, 71)} = 1.91$, $p = 0.009$]. The results were not significantly different between treatment groups [$F_{(1, 48)} = 0.38$, $p = 0.398$]. Table 11.1 summarizes scores by condition. Given that the variance was significantly different between control and treatment conditions, and that the variance of controls is more likely to reflect the true non-treated population variance, Glass's Delta could be a more accurate measure of effect size than the more commonly used Cohen's D, which is also provided (Fritz et al., 2012).

Table 11.1 Treatment tutors produce tutees with higher assessment scores

	M	SD	Cohen's D	Glass's delta	p value
Control	5.09	1.7			
TM training	6.41	2.75	0.74	1.34	0.009
Wise training	7.34	4.05	0.58	0.78	0.029

11.5.3 Round Two: Peer Instructional Behaviors Shift to Make Room for Peers to Think

Before examining tutee learning, let's consider *how* tutors taught. Several key patterns emerged: while behaviors related to explaining were common across groups, tutees in treatment groups described their tutors as asking questions and promoting active learning (i.e., "helping when needed" and "letting them try to solve the problem"). In Table 11.2 below, the percentages represent how often tutees mentioned these teaching moves, along with the other major categories mentioned, as the "most helpful thing" their tutor said or did.

These percentages are likely low estimates of how often these teaching practices occurred, as students were not specifically asked about each teaching practice, but rather given a general prompt to recall the "most helpful thing" the tutor did. That said, even though these data are not precise indicators of how often each of these teaching practices occurred, they do draw striking distinctions between treatment students and controls. While control group tutors were almost never described as asking questions, helping when needed, or letting their tutees try to solve problems, these were common descriptions of treatment group tutors. Here are several illustrative examples of tutee descriptions of treatment group tutors:

Table 11.2 Frequency of "most helpful" teaching moves as recalled by tutee

Most salient codes	Control (n = 49) (%)	WISE training (n = 55) (%)	TM training (n = 51) (%)
Explanation	37	31	41
Asked questions	0	13**	12*
Promoted active learning	4	20*	24**
Scaffolded the problem	10	13	12
Checked my work	12	13	16
Tutor was unhelpful	14	7	8

Note. To determine significant differences between conditions, a linear mixed effects model was used where teaching moves (represented by codes) were treated as fixed effects while teacher was treated as a random effect. $*p < 0.05$, $**p < 0.01$, compared to controls

- "He kept trying to get my thinking and he did that so he could explain the parts of the problem I did not know."
- "She gave me time to think. She also helped me with the problem when I needed it."
- "He asked me very helpful questions."
- "The most helpful thing was when they let me try the problem without trying to quickly correct my mistakes."

11.5.4 Tutoring Improves with Training and Content Mastery

While shifting teaching behaviors is an important intermediary goal, a successful intervention would additionally result in increased learning. Tutee assessment data suggest that both the wise intervention and Talk Moves Training were effective tools for improving peer tutoring quality, particularly when tutors first mastered the content.

Using orthogonal contrasts to compare the effect of tutors' training on tutee assessment scores, we find that being in either treatment group rather than control had a significant effect on tutee scores [$F(1, 152) = 8.65$, $p = 0.004$]. Neither treatment produced significantly different results than the other [$F(1, 105) = 0.07$, $p = 0.79$]. The mean score for tutees taught by control tutors (M = 39.9, SD = 14.5) was far below tutees taught by Wise Intervention (M = 50.3, SD = 21.1) and Talk Moves tutors (M = 49.2, SD = 21.3). The Cohen's D effect size was 0.58 and 0.51 for the Wise Intervention and Talk Moves training, respectively, compared to controls. Using the Glass's Delta formula, which substitutes the control SD for the pooled SD in cases where variance differs significantly by condition, the effect size was 0.72 for the Wise Intervention and 0.65 for the Talk Moves training, compared to controls.

To determine how much variance in tutee scores can be explained by tutors' content knowledge and treatment condition when controlling for each, a multiple regression analysis was conducted. In order to better understand the relationship between tutors' pre-assessment scores (indicating their content knowledge) and tutees' scores after being tutored, both sets of scores were converted into standardized Z scores where their mean is 0 and their standard deviation is 1. As shown in Table 11.3, analysis revealed significant effects for both tutor knowledge (i.e., tutor pre-assessment scores) and tutor training on tutees' scores following tutoring. Treatment condition and tutors' pre-test scores were not significantly associated with one another.

Table 11.3 Treatment and pre-test both have significant independent association with tutee learning

Coefficients	Estimate	Std. Error	t value	p value
Intercept	−0.029	0.084	−0.349	0.727
Standardized treatment versus control	0.501	0.185	2.705	0.008
Standardized pre-test	0.266	0.085	3.138	0.002
Treatment: pre-test	0.091	0.192	0.474	0.636

Note. The dependent variable was tutee assessment scores. Orthogonal contrasts were employed to combine treatment conditions

11.5.5 Combining Data from Both Studies Highlights Need for Mastery and Training

Given the similar data collection designs of Round 1 and Round 2, an even more robust statistical analysis is made possible. By standardizing tutee assessment scores and tutor pre-test scores (i.e., calculating z scores for each value where the mean score is 0 and the SD is 1), regressions were enabled for a combined dataset. Multiple regression with this data, which includes all students who participated in the entirety of either study (n = 204), reveals large training effects and large pre-test effects, both of which occurred independently of the other, as shown in Table 11.4. The Cohen's D effect sizes were 0.65 and 0.62 for the wise and talk moves trainings, respectively, compared to controls. The Glass's Delta effect sizes, which use controls' variance as their basis, were 0.92 and 0.78 for the wise and talk moves trainings, respectively.

To visualize the combined effects of tutors' content knowledge and treatment condition on tutees' post assessment scores, the data was broken down by pre-teaching quiz score bands. About a third of tutors fit in each of three categories: tutors who scored lowest, middling, or highest on the pre-test. After separating all tutors into pre-teaching quiz score bands in Fig. 11.8, we find that: (1) trained tutors are more effective helpers within every content knowledge band, and (2) tutors with strong mastery of the math content before teaching who received the PeerTeach training were much more effective helpers than every other group. This

Table 11.4 Treatment versus control and tutor pre-test are both independently associated with tutee learning

Coefficients	Estimate	Std. Error	t value	p value
(Intercept)	−0.024	0.066	−0.363	0.717
Treatment versus control	0.540	0.145	3.716	0.0003
Standardized pre-test	0.226	0.067	3.391	0.0008
Treatment: pre-test	0.103	0.146	0.703	0.483

Note. The dependent variable was tutee assessment scores. Orthogonal contrasts were employed to combine treatment conditions
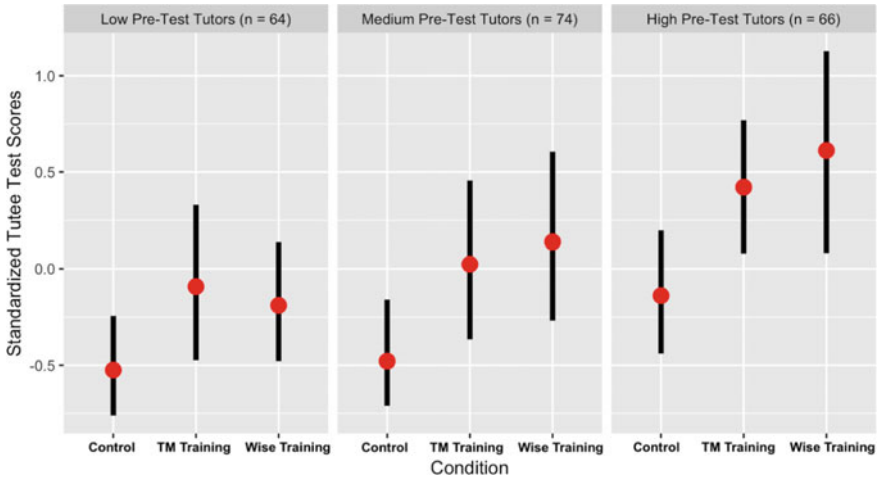
Fig. 11.8 Combining both rounds of data, tutor pre-test scores and condition both predict tutee learning. *Note.* Data was combined from both rounds of data collection by first converting tutee assessment scores and tutor pre-test scores into standardized z-scores. Dots represent means. Lines represent 95% confidence limits for the population mean obtained through nonparametric bootstrapping of the data

suggests that peer tutoring should occur when helping students have both strong content understanding and training on learner-centered teaching practices. Both pieces appear critical.

11.6 Discussion

As Paul and Elder (2019) write, "The history of education is also the history of educational panaceas, the comings and goings of quick fixes for deep-seated educational problems." The human tutor is not a novel innovation of the twenty-first century, but its efficacy is unparalleled by modern "panaceas." Instead of maintaining the churn of new innovations, identifying ways to expand and improve this millennia-old instructional strategy could pay more dividends.

Enlisting students to teach one another is a clear way to expand access to individualized coaching. The limiting factor is students' ability to teach as past studies have repeatedly documented their inclinations toward over-explaining and shallow questioning (Roscoe & Chi, 2007), which generally hinder learning. This investigation offers promising solutions. The two PeerTeach interventions increased the frequency of students using elicitive teaching techniques in both virtual and real-life tutoring scenarios, which translated into significant learning gains for tutees. While content mastery was a strong predictor of tutoring success, the combination of math knowledge with PeerTeach training produced more learning at every level of math proficiency. Given the seeming importance of both mastery and training,

it seems likely that activity structures that do not vet tutor mastery—for instance, ASK to THINK—TEL WHY—will yield less learning.

The results of this study suggest that (1) both prescriptive and constructivist online training modules can successfully shift peer tutoring behaviors, and (2) when those behaviors shift, tutee learning can be greatly amplified. While one might imagine other ways of improving peer tutoring, these specific intervention approaches are promising. Educators aiming to train tutors should consider combining these evidence-based training techniques with their own strengths as trainers and knowledge of their students. When facilitating teaching between children, confirming the tutor's mastery of content and monitoring their use of learner-centered teaching strategies will likely increase tutee learning.

The students of this study were split between two math teachers. One teacher's tutors exhibited learner-centered teaching behaviors at a much higher rate and their tutees performed significantly higher. Consequently, one alternative explanation of the results is that the effect of tutor training relies on how well teachers model the kinds of learner-centered teaching behaviors that are central to the trainings. With only two teachers participating in this study and without systematic measures of their teaching behaviors, this analysis was not possible in this study. Exploring the link between teachers' behaviors and student uptake of training ideas should be a priority in future studies.

The PeerTeach interventions are predicated on the consistent finding that tutors tend to explain too much, ask shallow questions, and fail to open up space for tutees to engage thoughtfully with content. To the degree this study underscored the potential for evidence-based training to cultivate Emergent Elicitors, it also highlighted the pervasiveness of the Default Didact. Before the intervention, students were less likely to select a learner-centered utterance out of three options than if selecting at random. When asked to report the most helpful thing their tutor did or said, tutees never described control tutors asking questions and only once described tutors helping when needed and letting them try to solve the problem. With this in mind, teachers who casually enlist students to help peers should heed this finding and take a more active role when facilitating peer helping. Indeed, as tutoring becomes a more integral feature for a broader swath of students in a Covid-impacted world, it is increasingly critical that non-expert tutors (peers or otherwise) learn to employ learner-centered pedagogy.

These interventions do not, however, advocate for a model of tutoring that is strictly question-based, like King (1998). There is a place for explanation, modeling, and many other non-questioning moves. Peer tutors should put together a toolbox of varied techniques to be applied when the situation is appropriate (MacDonald, 2000). In fact, backend data showed that tutors who selected learner-centered teaching moves 50–75% of the time (not 100%) helped tutees learn the most.

11.7   Limitations

These promising results are accompanied by several caveats. First, students' decisions in four online tutoring scenarios were not identical reflections of how they would behave in real life. They were proxies that suggest where students likely fall on a spectrum between didactic and elicitive endpoints. In order to predict tutoring tendencies based on online behaviors, building a sizable bank of teaching decisions in varied tutoring contexts (e.g., with different types of tutees or problems) could offer a more nuanced and precise indication of students' inclinations. The possibility of writing their own utterances could also lend further measurement precision. While providing added accuracy and nuance, these changes would also carry drawbacks. Drastically increasing the number of scenarios would be much more time-consuming for students and the inclusion of free responses would make data analysis and reporting more challenging. That said, future work should explore both mechanisms as tools for evaluating students' teaching inclinations and tracking progress.

Students' in-person teaching behaviors are also challenging to track. This investigation opted to measure them by asking tutees, "What was the most helpful thing your classmate did or said when teaching you? Give as much detail as you can." While this technique provided useful insights into the behavioral differences by condition, a more precise or in-depth method would utilize video or audio recordings of tutoring interactions. That way, a permanent record could be transcribed and coded by researchers to pinpoint exactly what students did. While video data was collected and analyzed to better understand the interactional mechanics of about ten tutoring pairs, tutee-written records allowed more coverage for this analysis. With more researchers and resources, video-based measurement will hopefully be utilized more extensively in future iterations of this work.

11.8   Conclusion

Emerging from COVID's devastating toll on learning, districts are turning to professional tutoring more than ever before. While there is solid evidence of the powerful impacts of high dosage tutoring (Dietrichson et al., 2017; Fryer, 2017)—often considered one-on-one instruction at least thrice weekly—it is logistically challenging to execute in schools (Allor & McCathren, 2004; Bryant et al., 2011) and expensive; even when scaled efficiently, costs are estimated between $2,500 and $3,800 annually per student (Ander et al., 2016). This study provides reason for optimism, suggesting that *peer* tutoring could be a viable alternative when coupled with the right training or effective assessment and matching systems. After just 40 min with both PeerTeach trainings, middle schoolers became demonstrably more effective tutors, particularly when they first mastered the math content. This finding was repeated in Round One and Round Two of data collection, offering a robust corpus of evidence.

This demonstration, though, is just a signal of how powerful peer tutoring can be when accompanied by research-based training. The next step in this line of research is to measure the impact of sustained peer tutoring that incorporates other elements of teacher professional development that can be applied to student tutors. For instance, as the Measures of Effective Teaching (MET) project evidenced, feedback from learners and instructional expert observers can be powerful tools for promoting teaching improvement (Rothstein & Mathis, 2013). Future studies could also measure students' growth in teaching ability over time as they engage in different forms of training, practice, and reflection, offering more precise insights on how to support development. In situating peer tutoring as a classroom routine, there are also opportunities for identifying useful principles for determining which students should teach what content and when.

For decades, we have known that all children can learn more with individualized support (Bloom, 1984), but we forgo such investments in our children. Fortunately, though, the benefits of tutoring may be within every child's grasp if we can harness the existing talent and ingenuity that abounds in every classroom. If we give students the responsibility of tutoring each other, though, we as educators must take on the responsibility of training children to teach effectively. This study suggests that—so long as students attain sufficient content mastery before tutoring—training them to use more learner-centered teaching strategies is an effective and realistic goal.

Appendix A: Rubric for Study 1 and 2 Post-assessment

Process dimensions	4	3	2	1
Representing and Solving the Task *Use models, pictures, diagrams, and/or symbols to represent and solve the task situation and select an effective strategy to solve the task*	The strategy selected and the representations used are: • Effective • Complete • Accurate • Logically explained	The strategy selected and the representations used are: • Mostly effective • Mostly complete • Correct, but lacking work or poorly explained	The strategy selected and representations used are 1. partially effective 2. underdeveloped	The strategy selected and representations used are: • Minimal, • Not evident • Not useful

Appendix B: Frequency of "Most Helpful" Teaching Moves as Recalled by Tutee

Codes	Control (n = 49)	TM training (n = 51)	Wise training (n = 55)	Total (n = 155)
Unhelpful	**7**	**4**	**4**	**15**
Was distracted	2	1	0	3
Was rude	3	0	0	3
Didn't know how to help	3	2	0	5
Did not talk	0	0	2	2
Gave me the answer	1	1	2	4
Asked questions	**0**	**6**	**7**	**13**
Asked questions (nonspecific)	0	4	6	10
Elicited my thinking	0	1	1	2
Probe my thinking (why)	0	0	0	0
Revoiced my thinking	0	1	0	1
Checked my work/understanding	**5**	**7**	**7**	**21**
Checked for understanding	1	1	3	5
Checked my work	3	1	1	5
Yes/no check for understanding	1	5	3	9
How well they explained	**6**	**8**	**5**	**19**
Explained the best they could	1	3	1	5
Explained well	5	5	4	14
Explained poorly	0	1	0	1
What they explained	**14**	**14**	**12**	**40**
Defined a term	4	3	3	10
Explained (nonspecific)	3	1	1	5
Explained the answer	0	1	1	2
Explained the problem	4	9	8	21
Modeled a problem	4	0	0	4

Codes	Control (n = 49)	TM training (n = 51)	Wise training (n = 55)	Total (n = 155)
Knew the topic well	2	0	0	2
Promoted active learning	**2**	**12**	**11**	**25**
Helped when needed	1	9	7	17
Let me try the problem	1	3	4	8
Scaffolded the problem	**5**	**6**	**7**	**18**
Explained step by step	0	1	3	4
Gave an example	1	1	1	3
Gave a hint	2	1	2	5
Guided through problem	2	3	2	7
Simplified the problem	0	0	1	1
Total (Unique)	48	46	52	146
Other codes				
Kept me focused	1	1	0	2
Moved slowly/patiently	2	1	0	3
Spoke clearly	3	0	4	7
Positive or encouraging attitude	6	1	2	9
Read the problem	2	1	0	3
Answered questions	0	1	2	3

References

Allor, J., & McCathren, R. (2004). The efficacy of an early literacy tutoring program implemented by college students. *Learning Disabilities Research & Practice, 19*(2), 116–129.

Ander, R., Guryan, J., & Ludwig, J. (2016). Improving academic outcomes for disadvantaged students: Scaling up individualized tutorials. The Hamilton Project–Brookings.

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*(2), 113–125.

Bamiro, A. O. (2015). Effects of guided discovery and think-pair-share strategies on secondary school students' achievement in chemistry. *SAGE Open, 5*(1), 1–7.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Berghmans, I., Neckebroeck, F., Dochy, F., & Struyven, K. (2013). A typology of approaches to peer tutoring. Unraveling peer tutors' behavioural strategies. *European Journal of Psychology of Education, 28*(3), 703–723.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16.

Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Greenwood, C., & Parker, R. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School Psychology Review, 42*(1), 39–55.

Bryant, D. P., Bryant, B. R., Roberts, G., Vaughn, S., Pfannenstiel, K. H., Porterfield, J., & Gersten, R. (2011). Early numeracy intervention program for first-grade students with mathematics difficulties. *Exceptional Children, 78*(1), 7–23.

De Jong, T., & Van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*(2), 179–201.

Dietrichson, J., Bøg, M., Filges, T., & Klint Jørgensen, A. M. (2017). Academic interventions for elementary and middle school students with low socioeconomic status: A systematic review and meta-analysis. *Review of Educational Research, 87*(2), 243–282.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141*(1), 2.

Fryer, R. G. Jr. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experiments. In *Handbook of Economic Field Experiments* (Vol. 2, pp. 95–322). North-Holland.

Godfrey, L., & O'Connor, M. C. (1995). The vertical handspan: Nonstandard units, expressions, and symbols in the classroom. *Journal of Mathematical Behavior, 14*(3), 327–345.

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology, 9*(6), 495–522.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning.* Routledge.

King, A. (1997). ASK to THINK-TEL WHY: A model of transactive peer tutoring for scaffolding higher level complex learning. *Educational Psychologist, 32*(4), 221–235.

King, A. (1998). Transactive peer tutoring: Distributing cognition and metacognition. *Educational Psychology Review, 10*, 57–74.

Kobayashi, K. (2019). Learning by preparing-to-teach and teaching: A meta-analysis. *Japanese Psychological Research, 61*(3), 192–203.

Leung, K. C. (2015). Preliminary empirical model of crucial determinants of best practice for peer tutoring on academic achievement. *Journal of Educational Psychology, 107*(2), 558–579.

Lortie, D. (1975). *Schoolteacher: A sociological study.* University of Chicago Press.

MacDonald, R. B. (2000). *The master tutor: A guidebook for more effective tutoring.* Cambridge Stratford Ltd.

Michaels, S., & O'Connor, C. (2015). Conceptualizing talk moves as tools: Professional development approaches for academically productive discussion. In L. Resnick, C. Asterhan, & S. Clarke (Eds.), *Socializing intelligence through talk and dialogue* (pp. 347–362). American Educational Research Association.

Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook. Sage Publications.

O'Connor, M. C. (2001). Can any fraction be turned into a decimal? A case study of a mathematical group discussion. *Educational Studies in Mathematics, 46*(1), 143–185.

O'Connor, M. C., & Michaels, S. (1993). Aligning academic task and participation status through revoicing: Analysis of a classroom discourse strategy. *Anthropology and Education Quarterly, 24*(4), 318–335.

Oregon Department of Education Office of Assessment and Evaluation. (2011). Mathematics problem solving scoring guide. Accessed at https://www.ode.state.or.us/wma/teachlearn/testing/scoring/guides/2011-12/mathpsscoringguide_eng.pdf

Paul, R., & Elder, L. (2019). *A critical thinker's guide to educational fads: How to get beyond educational glitz and glitter*. Rowman & Littlefield.

Roscoe, R. D., & Chi, M. T. H. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research, 77*(4), 534–574.

Rothstein, J., & Mathis, W. J. (2013, January 31). Review of "Have we identified effective teachers?" and "A composite estimator of effective teaching: Culminating findings from the measures of effective teaching project." *National Education Policy Center*. Retrieved from https://nepc.colorado.edu/thinktank/review-MET-final-2013

Saldaña, J. (2021). *The coding manual for qualitative researchers*. Sage.

Sherin, M. G., & Van Es, E. A. (2005). Using video to support teachers' ability to notice classroom interactions. *Journal of Technology and Teacher Education, 13*(3), 475–491.

Slavin, R. E. (2006). *Educational psychology theory and practice* (8th ed.). Library of congress Cataloging in Publication Data.

Topping, K. J., & Ehly, S. W. (2001). Peer assisted learning: A framework for consultation. *Journal of Educational and Psychological Consultation, 12*(2), 113–132.

Topping, K., Buchs, C., Duran, D., & Van Keer, H. (2017). Effective peer learning: From principles to practical implementation. Taylor & Francis.

Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. Science, 331(6023), 1447–1451.

Webb, N. M., & Mastergeorge, A. (2003). Promoting effective helping behavior in peer-directed groups. *International Journal of Educational Research, 39*(1–2), 73–97.

Yeh, E., & Swinehart, N. (2017). A learner-centered approach to technology integration: Online geographical tools in the ESL classroom. In *Handbook of Research on Learner-Centered Pedagogy in Teacher Education and Professional Development* (pp. 1–22). IGI Global.

A Thematic Analysis of Factors Influencing Student's Peer-Feedback Orientation

12

Julia Kasch, Peter van Rosmalen, and Marco Kalz

12.1 Introduction

Providing students with personalized feedback is a challenging task for teachers in (open online) higher education (Carless & Boud, 2018). Courses with high student numbers require scalable teaching practices in order to serve the educational needs of students by providing formative feedback and interaction opportunities (Kasch et al., 2017). In an earlier study we identified (online) lectures, students' self-assessment, peer-assessment and peer-feedback as scalable teaching practices (Kasch et al., 2021a, 2021b). Peer-feedback has a formative function and takes place between two (or more) students. It includes providing and receiving feedback with the goal of supporting the peer in his/her learning process (Topping, 2009). Due to innovation funding on peer-feedback, peer-feedback is more and more explored, implemented and analysed by Dutch universities and higher education institutes (SURF, 2020).

J. Kasch (✉)
Copernicus Institute of Sustainable Development, Utrecht University, Utrecht, Netherlands
e-mail: j.kasch@uu.nl

P. van Rosmalen
Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

M. Kalz
Digital Education/Mediendidaktik, Heidelberg University of Education, Heidelberg, Germany

265

Peer-feedback is a learning method in which students actively engage in so called 'assessment as learning' activities either in a face-to-face or online-context. Building on previous definitions of feedback, Carless and Boud (2018) define feedback as a "process through which learners make sense of information from various sources and use it to enhance their work or learning strategies". We refer to peer-feedback when students provide and receive formative feedback in the context of a learning activity (Huisman et al., 2019). During peer-feedback, both the provider as well as receiver learn with and from each other (Esterhazy & Damsa, 2019). Literature supports that students value both receiving as well as providing peer-feedback (Palmer & Major, 2008; Saito & Fujita, 2004) however, there are also studies reporting mixed results about students' perceptions (Liu & Carless, 2006; McConlogue, 2015; Nicol et al., 2014; Wen & Tsai, 2006). Regardless of the perceived value, providing and receiving feedback requires student engagement and openness and is a valuable workplace competence (Boud & Molloy, 2013; Carless & Boud, 2018; Huisman et al., 2019). It is influenced by students' previous peer-feedback experiences. Mulder et al. (2014) point out that students' beliefs change over time and that the perceived value of peer-feedback decreases after having participated in a peer-feedback activity. Some state that peer-feedback responses and beliefs can be seen as an outcome of a peer-feedback process, meaning that negative experiences have led to negative beliefs and vice versa (Price et al., 2011; van Gennip et al., 2009). Therefore, it is vital to create positive and valuable peer-feedback experiences early on.

Given the educational benefits of peer-feedback and the need to support positive peer-learning experiences, this chapter focuses on personal factors that influence students' openness to provide and receive peer-feedback (i.e. peer-feedback orientation). As teachers we can support students and increase peer-learning by being aware of personal factors that influence students' peer-feedback thoughts and behaviour. But currently, there is a research gap regarding personal factors influencing students' peer-feedback behaviour and a better understanding of individual differences (in higher education) of peer-feedback perception is missing (Dawson et al., 2019; Mulliner & Tucker, 2017; Srichanyachon, 2012; Strijbos et al., 2021; Taghizadeh et al., 2022). Overall research about beliefs and perceptions of feedback mainly focused on the feedback receiver (Alqassab et al., 2019) which is why we know less about the feedback provider (Winstone et al., 2017). Regarding students' peer-feedback beliefs, Huisman et al. (2019) developed a 'Beliefs about Peer-Feedback Questionnaire' (BFPQ). They argue that student's beliefs relate to the following four themes: (1) valuation of peer-feedback as an instructional method, (2) confidence in own peer-feedback quality, (3) confidence in quality of received peer-feedback and (4) valuation of peer-feedback as an important skill.

Outside educational settings, in the work field and performance management, we see more studies focusing on personal factors influencing feedback processes between employee and employer. In this context, the concept of 'Feedback Orientation' (London & Smither, 2002) was proposed which describes an "individuals' overall receptivity to feedback, including comfort with feedback, tendency to seek feedback and process it mindfully, and the likelihood of acting on the feedback

Table 12.1 Interview structure and questions

Demographic and background information	Background relates to occupation and peer-feedback experience
1. Introduction and open brainstorm	Q1. Try to give as many as possible factors in which students differ with regard to how they perceive giving, receiving and using peer-feedback
2. Interviewer shows the 4 factors of FOS	Q2. How would you understand the factors shown in relation to giving, receiving and using peer-feedback? Q3. To what degree are each of these factors relevant? Q4. Try to link the elements you mentioned in Q1 to the 4 factors shown
3. Why is it relevant?	Q5. Try to explain why these factors (the ones you mentioned) influence students' openness to peer-feedback. Do you link them to students in general or to a specific type of students?
4. Round up	Q6. Do you believe that the factors discussed sufficiently cover students' openness with regard to peer-feedback? If so, why? Would you like to add anything?

to guide behaviour change and performance improvement" (London & Smither, 2002, p. 81). Linderbaum and Levy (2010) elaborated on their work and developed a 'Feedback Orientation Scale' (FOS) which is used to investigate employees feedback orientation (openness towards feedback). Their work is focused on work-related feedback and performance appraisal in the job context. Nonetheless, the maturity of the work and the similarity to peer-feedback has motivated us to build in the authors' work. Focusing on the feedback receiver (employee), their scale (FOS) consists of four feedback orientation dimensions: utility, accountability, social awareness and self-efficacy (see Table 12.1 right column).

The feedback orientation concept and its translation into four dimensions (FOS) inspired us to use and transfer it to a higher education peer-learning setting. We expect that the four dimensions of the FOS are relevant in a higher education peer-feedback context. Various aspects of these dimensions have been mentioned in earlier feedback related studies (Alqassab et al., 2019; Boud & Molloy, 2013; Carless & Boud, 2018; Hulleman et al., 2008; Latifi et al., 2020, 2021; Patchan & Schunn, 2015). However, scales related to FOS such as the 'Feedback Environment Scale' (Steelman & Snell, 2004) or the 'Instructional Feedback Orientation Scale' (IFOS) (King et al., 2009) suggest that the context in which feedback orientation is studied influences the factors that can be attributed to it. Given the context of this study, we expect a different interpretation of the dimensions. Therefore, the goal of this study is to investigate if and how the four feedback orientation dimensions (utility, accountability, social awareness and self-efficacy) fit in the context of (higher) education and peer-feedback and if additional dimensions are needed to describe students' peer-feedback orientation.

Accordingly, the following research questions were investigated:

RQ1: Which personal factors are playing a role in students' peer-feedback orientation (i.e. openness to provide and receive peer-feedback) according to higher education students, teachers and researchers?
RQ1a: How can these elements be mapped by the existing feedback orientation dimensions (utility, accountability, social awareness, self-efficacy)?
RQ1b: How are utility, accountability, social awareness and self-efficacy interpreted in the context of peer-feedback in higher education?
RQ1c: Are additional dimensions needed to map elements that play a role in students' peer-feedback orientation?

12.2 Research Design and Method

This study is phase 1 of a 2-step-study design (exploratory sequential mixed methods). In a sequential exploratory mixed methods design, first, qualitative data is collected and analysed, followed by quantitative data collection and analysis. Data collection and analyses can take place separately, concurrently or sequentially (Creswell et al., 2011). In this study, data is collected sequentially which means that during the qualitative phase, interview data was collected and analysed to find elements which were used for the development of a quantitative instrument ('Peer-Feedback Orientation Scale'). This chapter (Fig. 12.1) covers the qualitative data collection, analyses and results while the quantitative part (exploratory factor analysis) is presented in a separate paper (Kasch et al., 2021a, 2021b).

12.3 Qualitative Data Collection and Analysis

Semi-structured interviews were held individually and face-to-face with each participant. An interview protocol was developed and tested beforehand which included a demographics- and a content section. In the demographics section the occupation and peer-feedback experience of the participants were asked. The content section (Table 12.1) started with an open think-aloud phase in which participants were asked to list and explain personal elements that influence their peer-feedback orientation (i.e. openness to provide and receive peer-feedback). Next, participants were presented with the four feedback-orientation-dimensions by Linderbaum and Levy (2010). Without further explanation of their meaning, the participants had to describe and interpret each dimension in the context of peer-feedback. Additionally, we asked them to explain the relevance of the dimensions regarding peer-feedback orientation. Lastly, participants had to assign their previously listed elements to the four dimensions (utility, accountability, social awareness and self-efficacy) and were allowed to add new dimensions if needed. An interview took on average 1 h and was tape-recorded with the permission of the
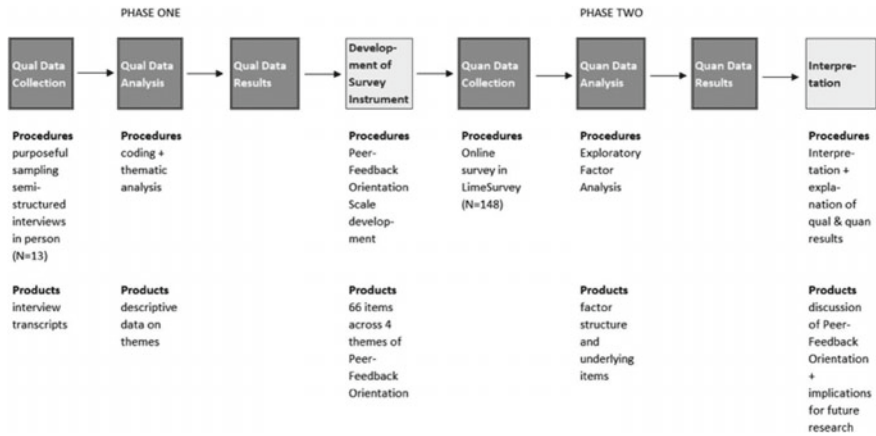
Fig. 12.1 Sequential Exploratory Design applied for this study adapted from Berman (2017). *Note* Adapted from "An exploratory sequential mixed methods approach to understanding researchers' data management practices at UVM: Integrated findings to develop research data services." E. A. Berman, 2017, *Journal of eScience Librarianship, 6*, p. 6 (https://doi.org/10.7191/jeslib.2017. 1098). In "The factor structure of the peer-feedback orientation scale (PFOS): toward a measure for assessing student's peer-feedback dispositions." J. Kasch, P. van Rosmalen, M. Henderikx and M. Kalz, 2021, *Assessment & Evaluation in Higher Education, 47,* p. 5 (https://doi.org/10.1080/ 02602938.2021.1893650)

participant. This study was approved by the ethical commission of our university and participation to the study was based on informed consent.

12.3.1 Participants

A sample (N = 13) of researchers, teachers and students from Dutch universities and higher education institutes participated in the semi-structured interviews. Using a purposeful sampling strategy enabled us to yield perspectives from individuals involved in a peer-feedback process (researchers, teachers and students). We approached teachers from seven research projects who had received a grant from the Dutch Ministry of Education to conduct peer-feedback related practice, four researchers on peer-feedback related research and four students with peer-feedback experience. A gift voucher was given for participation. The 13 semi-structured interviews (nine female and four male) were held within five universities and four universities of applied sciences. The data from five teachers (Amsterdam University, Delft University, Wageningen University, Saxion and HAN University of Applied Sciences), four university researchers and four students (Maastricht University, Open University of the Netherlands and Fontys University of Applied Sciences, Zuyd University of Applied Sciences) were included.

12.4   Data Analysis

The qualitative data analysis comprised multiple steps:

Transcription of interviews: The tape-recorded interviews were transcribed to prepare them for qualitative analysis by using GOM player (https://www.gom lab.com/). The interview transcripts were entered into N-Vivo 12 Pro for coding (https://www.qsrinternational.com/nvivo/nvivo-products/nvivo-12-pro).

Data coding: The transcripts were then coded using an 'In-Vivo' coding method (Saldana 2016). The 'In-Vivo' coding method is recommended for studies with the goal to develop new theory about a phenomenon. It is also suitable for novices, since the actual words, phrases and/or sentences of the interviewee are used as codes (Saldana 2016).

Construction of (sub-)themes: The four dimensions of FOS (Linderbaum & Levy, 2010) were guiding during the interviews and the analysis process. However, due to the shift from feedback in a workplace to peer-feedback in an educational setting, this study revisited the interpretation and number of dimensions that play a role in students' openness, within the perspective of both receiver and provider. The construction of (sub-)themes was done by the first two authors together. The result was presented to and discussed with the third author to produce a final version.

12.5   Findings

Research Question 1: Which personal factors are playing a role in students' peer-feedback orientation (i.e. openness to provide and receive peer-feedback) according to higher education students, teachers and researchers?

As mentioned previously, the FOS (Linderbaum & Levy, 2010) and its four dimensions were used as basis for the investigation of students' peer-feedback orientation. To get insight into the underlying personal factors that could play a role in students' peer-feedback orientation (RQ1) an open think-aloud interview took place. The findings of this phase show that various personal factors can influence students' peer-feedback orientation (see Appendix A for a translated list). All participants reported that the bond students have with their peer and the general atmosphere in the group has an influential factor for their orientation. Whilst a positive atmosphere in the group was seen as beneficial for the peer-feedback process, mixed responses were given about the influence of having a positive bond with their peers:

> *If you like somebody you don't want to run them into the ground and if you don't like somebody at all then maybe you are more inclined to do so.*

Students' confidence about their skills and knowledge were also seen as influential personal factors. The less confident, the more a student can struggle to provide as well as receive feedback. Another element highlighted was the idea of mutuality.

Peer-feedback is seen as a give-and-take process and students reported to feel more open if they have the feeling that the other person is putting effort into the provided feedback. However, mutuality seemed to be threatened by other factors such as the hierarchy between students. It was reported that students are more open to receive feedback from a knowledgeable peer than from a less knowledgeable one:

> *I have groups of seven students and there are good and bad students in them and they all know each other. They know who the good ones are and they know who the bad ones are. And the good ones think, yes, the bad ones don't matter to me, I'm not going to put any energy into them.*

> *If you think that your peer is not as knowledgeable, you are less likely to accept his feedback.*

Additionally, students' prior experience with peer-feedback was highlighted as a factor that can influence students' orientation. Uncertainty about the procedure and unfamiliarity with the aim of peer-feedback were seen as elements that could negatively influence openness. Students' feedback needs and readiness to provide and receive peer-feedback were also seen as relevant elements as well as the type of feedback (formative vs. summative) and the moment in which students provide and receive it. It was stated that students are more open to receive formative feedback compared to summative feedback because they are still able to use it for improvement.

> If you just started with the task and are not quite ready, receiving feedback can be too much.

> The receptivity for feedback will be positively influenced if you know what to expect and if you know that the feedback will be valuable for you.

By revisiting the meaning of the FOS dimensions (utility, accountability, social awareness, self-efficacy), we found first of all, that participants were able to map their generated elements by the FOS dimensions (RQ1a) and secondly, that the dimensions were perceived as relevant in the context of students' peer-feedback orientation.

Research Question 1b: How are utility, accountability, social awareness and self-efficacy interpreted in the context of peer-feedback in higher education?

Next, participants were presented with the four feedback orientation dimensions by Linderbaum and Levy (2010). Without further explanation of their meaning, the participants had to describe and interpret each dimension in the context of peer-feedback.

We found that the participants interpreted the FOS dimensions in a different way compared to Linderbaum and Levy (2010). Table 12.2 (right column) shows the different ways in which the FOS dimensions were interpreted when discussed in a peer-feedback setting versus a work-related setting (Linderbaum & Levy, 2010).

Transcribing and coding the responses regarding the meaning of the FOS dimensions resulted in a total of 562 codes. Two researchers clustered the 562 codes to meaningful subthemes within each feedback orientation dimension using

Table 12.2 Dimensions of the 'Feedback Orientation Scale' by Linderbaum and Levy (2010) and the 'Peer-Feedback Orientation Scale' by interviewees (N = 13)

	Dimensions Peer-Feedback Orientation Scale based on semi-structured interview data	Dimensions Feedback Orientation Scale (Linderbaum & Levy, 2010)
Utility	The personal added value a student perceives for their learning process by engaging in peer-feedback	An individual's tendency to believe that feedback is instrumental in achieving goals or obtaining desired outcomes at work
Accountability	A students' sense of responsibility for their own learning process and that of a fellow peer	An individual's tendency to feel a sense of obligation to act on feedback
Social awareness	A student's social connection with the group and/or peer and seeing peer-feedback as a social process	An individual's tendency to use feedback to be aware of other's views of oneself and to be sensitive to these views
Self-efficacy	A student's confidence in their knowledge and skills to provide valuable feedback	An individual's tendency to have confidence in dealing with feedback situations and feedback

principles of thematic analysis (Braun & Clarke, 2006; Maguire & Delahunt, 2017). This resulted in 15 subthemes (see Tables 12.2, 12.3 and 12.4). For a more detailed overview of the themes, subthemes and main corresponding codes see Appendix B, C and D.

The subthemes helped to get a better understanding of how the four dimensions were interpreted in the higher-education peer-feedback context (research question 1b). Additionally, the subthemes were needed for the item writing process for the 'Peer-Feedback Orientation Scale' in the quantitative part of this study (Kasch et al., 2021a, 2021b).

12.5.1 Utility

Utility plays an important role for students because they expect to improve from the feedback they receive. For them, utility mainly has to do with receiving new information, new perspectives and the way and the moment they receive the feedback. Formative feedback on draft versions is experienced as more useful than summative feedback on a finished piece where it is no longer possible to use the feedback. Extended feedback containing explanations, comments and discussions is experienced as clear and valuable. Additionally, classroom discussions ensure that students can learn from each other's cases. It was indicated that students take peer feedback seriously and expect their peers to take it seriously, too. The reciprocity of peer-feedback was mentioned by several participants as well as the need to provide and receive useful feedback in a constructive way. The knowledge level

Table 12.3 Themes and subthemes resulting from thematic analyses

Themes	Utility	Accountability	Social Awareness	Self-efficacy
Sub-themes	– Teacher role – Learning with feedback – Feedback is tailor-made – Creating meaning – Feedback moment	– Influence of the process on your accountability – Things you hold the other accountable for – Things you hold yourself accountable for	– On a group level – Behaviour that contributes positively to social awareness – Behaviour impairing social awareness	– Your role as a giver – Your role as a receiver – Self-efficacy for giver and receiver – Context prerequisites for self-efficacy

Table 12.4 Participant quotes about the peer-feedback orientation themes

Themes	Typical Quotes (translated from Dutch)
Utility	*"This is whether you think the feedback is useful for you."* [T4] *"The perceived usefulness influences the openness of a student. The following aspects play a role here: The quantity of the feedback: at various moments but not too much feedback. Moment of learning. Useful at the start but there must already be a basis."* [R1] *"Utility is really just about the product."* [S1]
Accountability	*"Are you accountable, do you feel accountable. So it's about, yes, when you provide feedback that you think 'I am responsible for what I have written down'."* [T4] *"I translate it as a kind of approachability."* [R4] *"That people can count on you."* [S3]
Social Awareness	*"Social awareness, then I immediately think of whether you have sensitivity, the social sensitivity."* [T3] *"I would translate it as the social context in which the peer-feedback takes place."* [R4] *"Yes, here the group feeling and the hierarchy within the group play a role."* [S4]
Self-efficacy	*"Self-efficacy is important for openness because if you feel that you cannot add valuable things you will be more reluctant to give feedback…"* [T3] *"I can imagine that if people have that, they are more likely to be active in peer-feedback."* [R3] *"Self-efficacy, like accountability, is about the product and yourself..that you want to give good feedback which not necessarily has to be about the product."* [S1]

T = Teacher; R = Researcher; S = Student

of the student and of peers can also play a role. Insecurity about their knowledge, can result in less openness to provide feedback. The same applies for the timing of feedback and students´ readiness to receive. For example, students who are working on the structure of a piece will perceive feedback on the completeness of content less useful since it does not match their current phase and needs. Additionally, the role of the instructor can influence how students view and deal with feedback. By assessing peer-feedback, giving feedback themselves, or simply checking on the feedback process can influence students' feedback perceptions and behaviour.

12.5.2 Accountability

Accountability was described as the sense of responsibility students have regarding their own learning process and that of someone else. Mutual commitment of both parties is important here. Familiarity, friendship and the setting (online or face-to-face) can influence the way students provide and perceive the received feedback. It was also mentioned that there is a difference between 'good' and 'weak' students

and it was claimed that good students take it more seriously. All in all, peer-feedback was described as an unselfish process in which you, as a student, have the goal of being able to help someone else with your feedback.

12.5.3 Social Awareness

All interviewees agreed that peer feedback is a social process. It takes place in a social context between one or more students and is therefore influenced by a number of (social) elements such as the group feeling, the bond with the group, the position in the group/hierarchy in terms of knowledge but also ranking/popularity. If students feel that the other person is empathetic, yet able to give feedback in an objective way, their openness to receive peer-feedback increases. Being aware of the fact that different perspectives are valid and that in some cases there is no one correct answer, is something students have yet to learn. The instructor should have an advisory role in this regard and lead discussions about different perspectives, which can increase students' sense of safety. Feeling safe in the way that it is OK to not know 'the' answer, to make mistakes, that there is room for discussions and for different perspectives was reported as important in peer-feedback. However, tactical play, favouritism, not being able to get along with each other, are social aspects that can stand in the way of students' openness.

12.5.4 Self-efficacy

Participants who were familiar with the term described it as faith in your own abilities. Those who did not know it could identify with this description. Whether students believe in their ability/knowledge or not influences their openness to provide feedback. Participants reported that previous experiences with peer-feedback can influence self-efficacy. Additionally, individual elements such as a student's self-image and self-confidence were also contributed to effect self-efficacy. The peer-feedback context and function (online vs. offline; formative vs. summative) can influence the degree to which students feel safe and thus influences their self-efficacy. To strengthen students' self-efficacy, instructors need to provide clear expectations and instructions around the peer-feedback process, examples and transparency.

Research Question 1c: Are additional dimensions needed to map elements that play a role in students' peer-feedback orientation?

Lastly, participants had to assign their previously listed elements to the four dimensions (utility, accountability, social awareness and self-efficacy) and were allowed to add new dimensions if needed.

A small number of participants proposed additional dimensions that could be considered when investigating students' peer-feedback orientation. These were 'psychological safety' (n = 1), 'personality traits' (n = 3) and 'socioeconomic

status' (n = 1). Psychological safety was described as an overarching basic require-
ment for peer-feedback to be effective. Students need to feel safe in a sense that
they know that there is nothing at stake and that others have to follow a code
of conduct. A few participants mentioned that personality traits such as being an
introvert or extrovert can play a role in students' openness towards providing and
receiving feedback.

12.6 Discussion and Conclusion

12.6.1 Discussion

An exploratory sequential mixed methods design was used to explore elements that
influence students' peer-feedback orientation and to investigate whether existing
feedback orientation dimensions fit to the higher education peer-feedback context.
The findings confirm our expectations, that the four feedback orientation dimen-
sions identified by Linderbaum and Levy (2010) (utility, accountability, social
awareness, self-efficacy) are seen as relevant in a peer-feedback context. Addi-
tionally, the findings confirm that the four feedback orientation dimensions have
another, more broader meaning when applied in a peer-feedback context and that
both receiving as well as providing feedback play a role in peer-feedback ori-
entation. The wide range of elements reported by the participants suggests that
student's peer-feedback orientation is influenced by diverse elements such as stu-
dents´ beliefs about what makes peer-feedback useful and fair. The findings also
show that peer-feedback is a complex process and to cover all student elements
that underlie students' peer-feedback orientation is a difficult task.

Related research on students' peer-feedback perceptions and beliefs, state that
student engagement increases if the value of feedback is clear (Moore & Teather,
2013). The findings that students value personal, specific, objective and construc-
tive feedback are also in line with the literature (Dawson et al., 2019; Li & De
Luca, 2014). Being confident in their own peer-feedback quality and in the quality
of the received peer-feedback was also found by Huisman et al. (2019).

Formative feedback was seen as more valuable for students as opposed to
summative feedback since students still have the chance to use the formative feed-
back to improve their current work. The importance for students to receive timely
feedback is shared with previous research on student perceptions (Carless, 2017;
Dawson et al., 2019; Pearce et al., 2010). Being able to use feedback in order to
improve, supports previous research by Price et al. (2010) who state that feed-
back on drafts is perceived as more helpful and valuable than feedback on an
end product. During the interviews, it was also stated that discussing the received
peer-feedback is valued by students and that it can increase their openness to
receive and use it. Especially when it comes to written peer-feedback, miscom-
munication and difficulties with interpreting comments can result in students not
using it, which was also reported in other studies (Carless, 2017; Price et al.,
2010; Schillings et al., 2021). These barriers can be resolved through discussion

and reflection. Additionally, dialogues about feedback and discussing examples increases students' perceived value of feedback (Price et al., 2010).

Utility was described as the added value of feedback in order to improve and to reaching goals, which is consistent with the study by Linderbaum and Levy (2010). In the workplace context, it was defined by variables regarding work success, skills development, performance improvement and goals reaching (Linderbaum & Levy, 2010). This was also reported by King et al. (2009) who found that in an educational context the perceived utility regarding teacher feedback was based on the motivational factors of teacher feedback, its importance for improvement and students´ tendency to listen to and reflect on teacher feedback.

In this current study, a broader range of variables was identified regarding utility in a peer-feedback context where both the feedback orientation of the receiver as well as the provider were included (e.g. learning with feedback, creating meaning, feedback being tailor made, the moment of receiving and providing feedback, gaining new perspectives, learning from receiving as well as providing). These findings match those of Nicol et al. (2009) who found that students value receiving feedback because it showed them other perspectives and spots for improvement. Similar to King et al. (2009) possible concerns regarding the usefulness of receiving feedback were expressed.

Linderbaum and Levy (2010) defined accountability as "an individual's tendency to feel a sense of obligation to react to and follow up on feedback" (p. 1377). Although in line with this definition, the results of this study indicated that in peer-feedback, students not only feel responsible to act on the feedback they receive but also for the feedback they provide. Peer-feedback was described as a reciprocal and unselfish process in which students try to support their peers However, it was also stated that some students may have concerns regarding the fairness and seriousness of their peers during the peer-feedback process. Good students were attributed to being more serious than weaker students. In the IFOS by King et al. (2009) accountability is not a separate dimension. A possible explanation might be that teacher feedback is not seen as optional remark on student performance but seen as compulsory expert feedback.

Contrary to the results of Linderbaum and Levy (2010), social awareness was not solely defined by others' impressions about yourself and how you are perceived by others but rather by the social bond between students and the atmosphere in the group. In a peer-feedback context, social awareness was seen as a very relevant dimension, due to the co-dependency between students being both receiver as well as provider of feedback. Hierarchy between students resulting from differences in domain knowledge and social positioning in the group were stated as relevant factors for the social awareness dimension. In a face-to-face context, social awareness was reported as being higher as opposed to an online context due to the direct contact and relates with the accountability dimension. The IFOS does not contain a social awareness dimension, however their students' 'sensitivity' dimension includes elements that are similar to the findings of this study (i.e. feeling threatened, hurt and stressed by corrective feedback from the teacher) (King et al., 2009). Compared to teacher feedback, peer-feedback makes students

co-dependent of each other, which can influence their (social) behaviour and the manner in which they provide feedback.

In the work environment, self-efficacy was defined as "an individual's tendency to have confidence in dealing with feedback situations and feedback" (Linderbaum & Levy, 2010, pp. 1386). The underlying variables focus on the feedback receivers´ ability to handle, receive and respond to feedback. Again, compared to the FOS (Linderbaum & Levy, 2010), feedback orientation in a peer-feedback context focuses on both the provider as well as the receiver. This distinction is relevant since students' self-efficacy can vary across tasks (providing vs. receiving) and topics (being more/less knowledgeable in a certain topic). Elements such as fear for criticism, fear of being vulnerable and negative experiences with peer-feedback can negatively influence students' self-efficacy and thus their openness to receive. A student who is not able to receive feedback because of fear, will likely not see any value in it. Students fear of (corrective) feedback was also described by the feedback sensitivity dimension by King et al. (2009). Although self-efficacy is not a separate scale in the IFOS (King et al., 2009) elements were still included in the form of feedback retention (i.e. student ability to recall and remember teacher feedback).

The findings support the hypothesis that feedback orientation is a universal concept however its implementation is dependent on the context, the parties involved and the function of feedback. Therefore, further investigating the dimensions underlying students' feedback orientation towards peer-feedback seems relevant and promising. Comparing the findings of this study with related feedback orientation scales (King et al., 2009; Linderbaum & Levy, 2010) appeared complex, given the differences in context (educational vs. work environment), stakeholders (student–student vs. teacher-student vs. employer-employee) and feedback function (mandatory formative peer-feedback vs. corrective teacher feedback vs. developmental feedback). As discussed, the findings of this study are both consistent as well as contrasting compared to the 'Feedback Orientation Scale' and the 'Instructional Feedback Orientation Scale'.

12.7   Limitations of the Study and Recommendations for Future Research

The major limitation of the study was the small sample size of the participants involved in the research and the limitations to draw the sample only from a Dutch Higher Education context. This decision has been taken for practical reasons, but we might have identified some specific experiences or traits which are especially relevant in this context, but not in others. Future research will need to confirm the findings of this study and the follow-up study (Kasch et al., 2021a, 2021b) to be generalizable beyond the current context.

Additional research will be needed in terms of identifying meaningful differences in students with regard to peer-feedback orientation. While some individual

differences can be identified they do not need a differentiated approach for students. At the same time, specific dispositions may need actions which may help students to overcome for example a negative attitude or prior experience with peer-feedback.

12.8   Conclusions

This paper contributes to the theory development for peer feedback orientation and proposes a new conceptualisation of peer feedback orientation. Based on our findings, students' peer-feedback orientation relates to providing as well as receiving feedback, the relationship students have with each other and their skills. The findings have been used as a source for the development and testing of a preliminary 'Peer-Feedback Orientation Scale', useful for getting insight into students' dispositions or orientations/openness, towards receiving and providing peer-feedback (Kasch et al., 2021a, 2021b). Being aware and informed about students' peer-feedback orientation, especially at the beginning of a learning activity, course or even semester can provide teachers with the opportunity to address issues around student perspectives and experiences regarding the utility of providing and receiving peer-feedback, feelings of accountability, social awareness and self-efficacy.

This chapter has provided a documentation of the first step of a 2-step-study exploratory sequential mixed method design with the goal to develop a reliable and valid instrument to measure peer-feedback orientation of students in higher education. The second step of this research has been published already (Kasch et al., 2021a, 2021b). The final goal of the research is to offer options for practitioners to react to individual differences in students regarding their preparedness for peer-feedback activities and to avoid negative experiences with peer-feedback.

Appendix A

List of (personal) elements that influence students' openness to provide and receive peer-feedback provided by interviewees (N = 13) during think-aloud part of a semi-structured interview.

Interviewee	Elements influencing students' openness to provide and receive peer-feedback
Student 1	• Feedback previously received from the teacher • Amount of time invested in the task (on which you will receive feedback) • Self-confidence • Getting on well with the other students of the peer-feedback group • Providing positive feedback to receive positive feedback as well
Student 2	• Being afraid to hurt the other person • Feeling unsecure to provide feedback • Feeling unsecure to receive feedback • Providing feedback in an objective way

Interviewee	Elements influencing students' openness to provide and receive peer-feedback
Student 3	• Attitude of peer-feedback receiver • Group context/structure • Confidence • New feedback • The way you receive and provide feedback • Having specific moments were you provide and receive feedback • Mutual effort in providing feedback • Explaining feedback
Student 4	• Own knowledge • Justified feedback • Knowledge level of feedback provider • Boosting participation score • Relationship with the other person • Confidence level of feedback provider
Teacher 1	• Introvert • Experience with peer-feedback • Factual knowledge • Tactical game between students • Ratio of knowledge in the group • Atmosphere in the group
Teacher 2	• Uncertainty about the procedure • Fear of criticism • Uncertainty over content knowledge
Teacher 3	• Life experience/maturity • Self-confidence in providing and receiving peer-feedback • Previous experience with peer-feedback (in a formal and informal way) • Familiarity with the scientific process of peer-review • Social sensitivity (introvert/extrovert)
Teacher 4	• Self-image • Self-confidence • Alleged knowledge in the field in question • Number of siblings • Position in the group/class • Emotional age/matureness • Experienced consequences of the peer-feedback activity • Sex • Language skills • Previous experience with peer-feedback • Mood • Cultural background • Extrovert/introvert

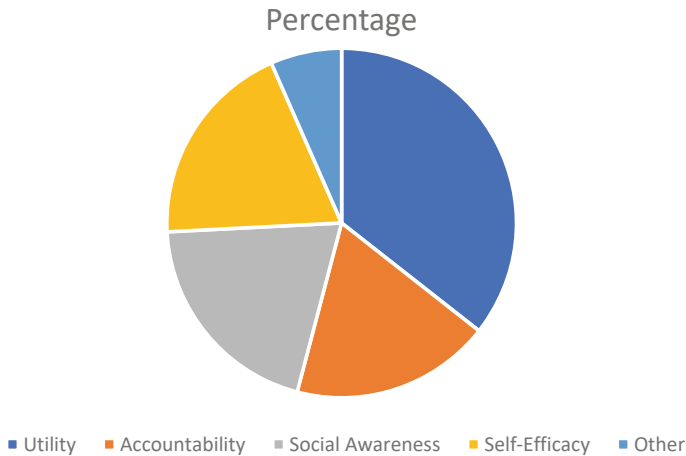| Interviewee	Elements influencing students' openness to provide and receive peer-feedback
Researcher 1	• Familiarity with the peer • Position in the group • Being open-minded and receptive • Working one-on-one or in a group • Character from the peer • Culture • Safe environment • Expertise of the peer • Hierarchy • Mental state • Moment of the learning process • Feedback on the task vs. feedback on the process • Summative vs. formative feedback • Reliability of the feedback • Quality of the work one has to review • Added value for own learning • Amount of feedback one is receiving
Researcher 2	• Self-awareness • Judgement of learning • Confidence • Positive attitude • Perseverance • Time to be spent on peer-feedback • Formative way • Personality (introvert/extrovert) • Curiosity/eagerness to learn • Trust in others • Knowledge level • Sex • Perfectionism
Researcher 3	• Need for structure, expectations and a clear goal • Self-confidence • Unfamiliarity with the group • Negative association with peer-feedback • Previous experiences with peer-feedback • Motivation • Interactivity (social skills) • Content knowledge
Researcher 4	• Previous experience, both positive and negative with peer-feedback • Whether the feedback meets your needs or not • Self-image about yourself as a human being • Self-image about your knowledge and skills • Your state of being/ current mood • Your strength • Your view of the other persons' knowledge and skills • Your opinion about the other person

Appendix B

Themes, subthemes and the main corresponding codes (originally in Dutch and translated for publication).

Theme: Utility	Theme: Accountability	Theme: Social Awareness	Theme: Self-efficacy
Subtheme: Teacher role	**Subtheme: Influence of the process on your accountability**	**Subtheme: On a group level**	**Subtheme: Your role as a giver**
Teacher control	Being approachable	Higher in face-to-face	Feeling competent to add something
Equivalency student and teacher feedback	Talking about feedback	Trust in the feedback provider	Time investment
Grading feedback	Uncertainty about the feedback process	Position in the group	Feedback on your feedback
Subtheme: Learning with feedback	Teacher making students feel accountable	**Subtheme: Behaviour that contributes positively to social awareness**	The way you receive feedback
Feedback on the process	**Subtheme: Things you hold the other accountable for**	Empathise	**Subtheme: Your role as a receiver**
Aimed at the receiver	Reciprocity	Balance between tips and tops	Fear for criticism
Getting another perspective	Benefitting from my feedback	Being open to different points of view	Wanting to receive feedback
Added value of feedback	Responsible for own learning process	Psychological safety	Previous experiences
Learning from each other	Familiarity of the peer	**Subtheme: Behaviour impairing social awareness**	Testing whether feedback is justified
Subtheme: Feedback is tailor-made	Taking peer-feedback seriously	If you can get along with the other person	There is no black and white answer
Insecurity about own knowledge	**Subtheme: Things you hold yourself accountable for**	Other person is benefitting from my work	Being vulnerable
Needs of receiver and provider	Trust in your own abilities	Tactical moves	Being able to process feedback
Being ready to receive	Unselfish	Anonymity	**Subtheme: Self-efficacy for giver and receiver**
Receiving too much	Responsible for own learning process		Having enough content knowledge
Content related and constructive feedback	Doing something with the feedback		Self-image
Providing value			Self-confidence
Being able to recognize value			**Subtheme: Context prerequisites for self-efficacy**
What happens next?			Transparency of the process
Subtheme: Creating meaning			Feedback as a skill
Same frame of reference			Training peer-feedback
Familiarity with the scientific process			
Explaining feedback			
Small groups			
Goal of feedback			
Subtheme: Feedback moment			
Feedback on completed work			
Feedback on draft version			
Time to use feedback			
Several feedback moments			

Appendix C

Percentage distribution of all peer-feedback orientation themes.

Appendix D

Frequencies and percentages of the peer-feedback orientation themes and corresponding subthemes.

Peer-Feedback Orientation (Sub-)Themes	Frequency	Within a subtheme		Across all subthemes	
		Relative Frequency	%	Relative Frequency	%
Utility					
Definition	16	0.080	8	0.028	3
Subtheme 1 'Teacher role'	43	0.215	22	0.077	8
Subtheme 2 'Learning with feedback'	47	0.235	24	0.084	8
Subtheme 3 'Feedback is tailormade'	39	0.195	20	0.069	7
Subtheme 4 'Creating Meaning'	36	0.180	18	0.064	6
Subtheme 5 'Feedback Moment'	19	0.095	10	0.034	3
Total	**200**	**1.000**	**100**	**0.356**	**36**
Accountability					
Definition	17	0.163	16	0.030	3
Subtheme 1 'Influence of the process on your accountability'	19	0.183	18	0.034	3

Peer-Feedback Orientation (Sub-)Themes	Frequency	Within a subtheme		Across all subthemes	
		Relative Frequency	%	Relative Frequency	%
Subtheme 2 'Things you hold the other accountable for'	49	0.471	47	0.087	9
Subtheme 3 'Things you hold yourself accountable for'	19	0.183	18	0.034	3
Total	**104**	**1.000**	**100**	**0.185**	**19**
Social Awareness					
Definition	13	0.115	12	0.023	2
Subtheme 1 'On a group level'	41	0.363	36	0.073	7
Subtheme 2 'Behaviour that contributes positively to social awareness'	36	0.319	32	0.064	6
Subtheme 3 'Behaviour imparing social awareness'	23	0.204	20	0.041	4
Total	**113**	**1.000**	**100**	**0.201**	**20**
Self-efficacy					
Definition	12	0.111	11	0.021	2
Subtheme 1 'Your role as a giver'	16	0.148	15	0.028	3
Subtheme 2 'Your role as a receiver'	30	0.278	28	0.053	5
Subtheme 3 'Self-efficacy for giver'	31	0.287	29	0.055	6
Subtheme 4 'Context/prerequisites for self-efficacy'	19	0.176	18	0.034	3
Total	**108**	**1.000**	**100**	**0.192**	**19**
Other	37	1.000	100	0.066	7
Total all	**562**				

References

Alqassab, M., Strijbos, J. W., & Ufer, S. (2019). Preservice mathematics teachers' beliefs about peer feedback, perceptions of their peer feedback message, and emotions as predictors of peer feedback accuracy and comprehension of the learning task. *Assessment & Evaluation in Higher Education, 44*(1), 139–154.

Berman, E. A. (2017). An exploratory sequential mixed methods approach to understanding researchers' data management practices at UVM: Integrated findings to develop research data

services. *Journal of eScience Librarianship, 6*(1), 1–24. https://doi.org/10.7191/jeslib.2017.1098

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education, 38*(6), 698–712. https://doi.org/10.1080/02602938.2012.691462

Carless, D. (2017). Scaling up assessment for learning: progress and prospects. In D. Carless, S. M. Bridges, C. K. Y. Chan & R. Glofechski (Eds.), *Scaling up assessment for learning in higher education* (pp. 3–17). Springer Nature.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Creswell, J. W., & Plano Clark, V. L. (2011). Choosing a mixed methods design. *Designing and Conducting Mixed Methods Research, 2*, 53–106.

Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D., & Molloy, E. (2019). What makes for effective feedback: Staff and student perspectives. *Assessment & Evaluation in Higher Education, 44*(1), 25–36. https://doi.org/10.1080/02602938.2018.1467877

Esterhazy, R., & Damşa, C. (2019). Unpacking the feedback process: An analysis of undergraduate students' interactional meaning-making of feedback comments. *Studies in Higher Education, 44*(2), 260–274. https://doi.org/10.1080/03075079.2017.1359249

Huisman, B., Saab, N., Van Driel, J., & Van Den Broek, P. (2019). A questionnaire to assess students' beliefs about peer-feedback. *Innovations in Education and Teaching International, 1–11*,. https://doi.org/10.1080/14703297.2019.1630294

Hulleman, C. S., Durik, A. M., Schweigert, S. A., & Harackiewicz, J. M. (2008). Task values, achievement goals, and interest: An integrative analysis. *Journal of Educational Psychology, 100*(2), 398.

Kasch, J., Van Rosmalen, P., & Kalz, M. (2017). A framework towards educational scalability of open online courses. *Journal of Universal Computer Science, 23*(9), 845–867.

Kasch, J., Van Rosmalen, P., Henderikx, M., & Kalz, M. (2021a). The factor structure of the peer-feedback orientation scale (PFOS): Toward a measure for assessing students' peer-feedback dispositions. *Assessment & Evaluation in Higher Education*, *47*(1), 15–28. https://doi.org/10.1080/02602938.2021.1893650.

Kasch, J., Van Rosmalen, P., & Kalz, M. (2021b). Educational scalability in MOOCs: Analysing instructional designs to find best practices. *Computers & Education, 161*, 104054. https://doi.org/10.1016/j.compedu.2020.104054

King, P. E., Schrodt, P., & Weisel, J. J. (2009). The instructional feedback orientation scale: Conceptualizing and validating a new measure for assessing perceptions of instructional feedback. *Communication Education, 58*(2), 235–261. https://doi.org/10.1080/03634520802515705

Latifi, S., Noroozi, O., Hatami, J., & Biemans, H. J. A. (2021). How does online peer feedback improve argumentative essay writing and learning? *Innovations in Education and Teaching International, 58*(2), 195–206. https://doi.org/10.1080/14703297.2019.1687005

Latifi, S., Noroozi, O., & Talaee, E. (2020). Worked example or scripting? Fostering students' online argumentative peer feedback, essay writing and learning. *Interactive Learning Environments, 1–15*. https://doi.org/10.1080/10494820.2020.1799032

Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education*, *11*(3), 279–290. http://hdl.handle.net/10722/54282.

Linderbaum, B. A., & Levy, P. E. (2010). The development and validation of the Feedback Orientation Scale (FOS). *Journal of Management, 36*(6), 1372–1405.

Li, J., & De Luca, R. (2014). Review of assessment feedback. *Studies in Higher Education, 39*(2), 378–393.

London, M., & Smither, J. W. (2002). Feedback orientation, feedback culture, and the longitudinal performance management process. *Human Resource Management Review, 12*(1), 81–100.

McConlogue, T. (2015). Making judgements: Investigating the process of composing and receiving peer feedback. *Studies in Higher Education, 40*(9), 1495–1506.

Moore, C., & Teather, S. (2013). Engaging students in peer review: Feedback as learning. *Issues in Educational Research, 23*(2), 196–211.

Mulder, R. A., Pearce, J. M., & Baik, C. (2014). Peer review in higher education: Student perceptions before and after participation. *Active Learning in Higher Education, 15*(2), 157–171.

Mulliner, E., & Tucker, M. (2017). Feedback on feedback practice: Perceptions of students and academics. *Assessment & Evaluation in Higher Education, 42*(2), 266–288. https://doi.org/10.1080/02602938.2015.1103365

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education, 39*(1), 102–122. https://doi.org/10.1080/02602938.2013.795518

Palmer, B., & Major, C. H. (2008). Using reciprocal peer review to help graduate students develop scholarly writing skills. *The Journal of Faculty Development, 22*(3), 163.

Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: How students respond to peers' texts of varying quality. *Instructional Science, 43*(5), 591–614.

Pearce, J., Mulder, R., & Baik, C. (2010). *Involving students in peer review: Case studies and practical strategies for university teaching.* http://www.cshe.unimelb.edu.au/resources_teach/teaching_in_practice/docs/Student_Peer_Review.pdf.

Price, M., Handley, K., Millar, J., & O'donovan, B. (2010). Feedback: All that effort, but what is the effect? *Assessment & Evaluation in Higher Education*, *35*(3), 277–289.

Price, M., Handley, K., & Millar, J. (2011). Feedback: Focusing attention on engagement. *Studies in Higher Education, 36*(8), 879–896.

Saito, H., & Fujita, T. (2004). Characteristics and user acceptance of peer rating in EFL writing classrooms. *Language Teaching Research, 8*(1), 31–54.

Schillings, M., Roebertsen, H., Savelberg, H., van Dijk, A., & Dolmans, D. (2021). Improving the understanding of written peer feedback through face-to-face peer dialogue: Students' perspective. *Higher Education Research & Development, 40*(5), 1100–1116. https://doi.org/10.1080/07294360.2020.1798889

SURF. (2020). *Stimuleringsregeling Open Online Onderwijs.* https://www.surf.nl/en/node/2511.

Srichanyachon, N. (2012). An investigation of university EFL students attitudes toward peer and teacher feedback. *Educational Research and Reviews, 7*(26), 558–562.

Steelman, L. A., Levy, P. E., & Snell, A. F. (2004). The feedback environment scale: Construct definition, measurement, and validation. *Educational and Psychological Measurement, 64*(1), 165–184.

Strijbos, J. W., Pat-El, R., & Narciss, S. (2021). Structural validity and invariance of the feedback perceptions questionnaire. *Studies in Educational Evaluation, 68*, 100980. https://doi.org/10.1016/j.stueduc.2021.100980

Taghizadeh Kerman, N., Noroozi, O., Banihashem, S. K., Karami, M. & Biemans, H. H. J. A. (2022). Online peer feedback patterns of success and failure in argumentative essay writing. *Interactive Learning Environments, 1–10.* https://doi.org/10.1080/10494820.2022.2093914.

van Gennip, N. A., Segers, M. S., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review, 4*(1), 41–54.

Wen, M. L., & Tsai, C. C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education, 51*(1), 27–44.

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist, 52*(1), 17–37. https://doi.org/10.1080/00461520.2016.1207538

Giving Feedback to Peers in an Online Inquiry-Learning Environment

13

Natasha Dmoshinskaia and Hannie Gijlers

13.1 Introduction

Traditionally, feedback from peers has been used when teachers cannot provide proper feedback themselves, usually because of large class sizes (e.g., Falchikov & Goldfinch, 2000). This way of using peer assessment has not been fully adopted by teachers, due to its low reliability and validity (e.g., Liu & Carless, 2006). However, there has been a shift in goals, from using peer feedback as an assessment tool replacing teacher feedback, to using it for the benefits of learning, as a learning tool (e.g., Adachi et al., 2018; van Popta et al., 2017). As peer assessment consists of two processes—giving feedback and receiving feedback—both can (and do) contribute to learning (e.g., Li et al., 2020). However, several studies have indicated that giving feedback can lead to comparable or even greater learning than receiving feedback (e.g., Ion et al., 2019; Li & Grion, 2019; Phillips, 2016). Therefore, when used as a learning tool, giving feedback to peers can be a learning experience for feedback providers (or reviewers). Even though the contribution that giving feedback makes to learning has been shown, that part of the peer assessment process has been less studied than the receiving-feedback part.

N. Dmoshinskaia (✉) · H. Gijlers
Department of Instructional Technology, University of Twente, Enschede, The Netherlands
e-mail: n.dmoshinskaia@utwente.nl

289

Separating these two parts and focusing only on giving feedback could lead to better understanding of the factors that might influence reviewers' learning.

Such learning can be attributed to several factors. One is that while giving feedback to peers, students need to be cognitively involved with the material and the task. They need to compare the product to be reviewed with their own understanding and/or self-created product; this comparison leads to deeper thinking and thereby to learning (Nicol & McCallum, 2022). Another factor that can lead to learning is the process of thinking of and formulating appropriate feedback, which can again stimulate deeper thing about the material (e.g., Lundstrom & Baker, 2009).

To maximize learning originating from giving feedback to peers, it is important first to study how the design of the feedback-giving procedure can influence its outcomes. And to do that, it is important to deconstruct this process to study each phase separately. We conceptualised the feedback-giving process using the model suggested by Sluijsmans (2002) that includes three steps: define assessment criteria, judge the performance of a peer, and provide feedback for future learning.

Usually a feedback-giving task is included in a course as separate activity that requires specially allocated time because it covers bigger scale products, such as essays, reports, or team projects. This also means that such an activity should be planned appropriately—there should be enough time given to it, so quite often it is set as homework or self-study. Teachers may be reluctant to include giving feedback to peers in their courses, as the feedback can be unreliable and too time-consuming for students (e.g., of Liu & Carless, 2006). However, if giving feedback can fit into a regular 50-min class and have a formative nature, this would give students an opportunity to interact with the material at a meta-level and learn from it, and still proceed with the usual classroom activities. To achieve this goal, this activity must be designed so that the feedback-giving moment is not too long. Therefore, the reviewed products should be relatively small, so that giving feedback on them does not occupy too much time.

Using smaller scale products and a shorter feedback-giving interaction could also influence the learning of feedback providers. Therefore, studying what learning can be triggered by reviewing such products is valuable for practice. Below, the results of a series of (quasi-)experimental studies investigating the process of giving feedback on smaller products are presented. Each study focused on one particular design feature related to one of the steps of the feedback-giving model by Sluijsmans (2002) mentioned above: defining assessment criteria, judging the performance of a peer, and providing feedback for future learning. The following features of the feedback-giving process were studied: being provided with assessment criteria (Step 1); the quality and the type of reviewed product (Step 2); and the form of providing feedback (Step 3). The rationale for each study is described below.

Step 1: When faced with the task of giving feedback, students can either be provided with assessment criteria or come up with their own. There is no clear opinion as to which of these is more beneficial for learning. Some studies have indicated that thinking of their own assessment criteria leads to greater ownership

of learning for students and results in more involvement and, thus, more learning (e.g., Canty et al., 2017; Tsivitanidou et al., 2011). However, other studies have suggested that assessment criteria can guide students in the process of giving feedback and provide them with required structure (e.g., Gan & Hattie, 2014; Panadero et al., 2013). Therefore, the question about the role that the source of assessment criteria plays in reviewers' learning is not clearly answered.

Step 2: The quality and type of a reviewed product can influence the quality and content of the feedback that reviewers provide and, as a result, the learning that arises from it (e.g., Patchan & Schunn, 2015). Some studies have shown that giving feedback on higher quality products can lead to more learning, as students see good examples and understand the material better (e.g., Alqassab et al., 2018a; Tsivitanidou et al., 2018). However, if the level of the reviewed product is too high, students may not be able to find mistakes and, thus, learn (e.g., Cho & Cho, 2011), which may mean that products of mediocre quality can stimulate learning more than those of high quality. Similarly, the type of product may affect learning. For example, students may find familiar and straightforward products, such as answers to open-ended questions, easy to review, as they understand the format and the expectations, and can find more mistakes. Some research has shown that identifying more mistakes in a reviewed product leads to more learning (e.g., Adams et al., 2019). However, reviewing a more challenging product such as a concept map may lead to more conceptual understanding and trigger more learning (e.g., Chen & Allen, 2017). This makes the effect of different levels and types of reviewed products on learning interesting to investigate.

Step 3: Giving feedback can be done in the form of comments or grades. Previous research has shown that providing cognitive feedback, that is, focusing on the task and not on the evaluation, and identifying mistakes in the reviewed products leads to learning for feedback providers (e.g., Lu & Law, 2012; Lu & Zhang, 2012; Wooley et al., 2008). However, it is not clear if the form of giving feedback will influence the learning when reviewing smaller scale products.

In all of the studies conducted, there was one factor that was considered thoroughly—students' prior knowledge. Previous research has shown that reviewers' prior knowledge can influence the way they interact with the material and the feedback they give (e.g., Alqassab, 2017; Patchan & Schunn, 2015). This is most obvious if we look at the quality of the reviewed product—the same product can too difficult and not understandable for lower prior-knowledge students, and stimulating and inspiring for higher prior-knowledge students. This means that the first case would lead to less cognitive involvement and, thus, less learning, while the second case could trigger more cognitive engagement and, thus, more learning. Similar influences can be seen for the other steps of the feedback-giving process. Therefore, the level of prior knowledge of feedback providers was taken into account in the analyses.

Nowadays, giving feedback to peers is often done with the help of technologies—online platforms, apps, or specially developed tools—a plug-in in Canvas (an LMS), Eduflow, or PeerGrade, to name just a few. One distinguishing feature of using such products is the possibility for the teacher to adjust and adapt the

process of giving feedback to their current goals by changing several parameters: anonymous or not, synchronous or not, using specific assessment criteria or not, reciprocal peer feedback or not—the list of adjustable parameters goes on. Moreover, such settings can be applied to all students or to specific groups of students. Therefore, knowing what settings lead to more learning for a specific group of students or in a specific context can have a clear translation into practice. This makes investigating the feedback-giving process conducted with a technology-based tool quite topical, as we use established methods to study the feedback-giving process in a new context to enrich both theoretical knowledge about it and practical implementation procedures.

In the sections below, we present our research, the goal of which was to investigate the learning of feedback providers in an online environment and how to increase such learning by designing the feedback-giving process in a particular way. First, we describe the studies conducted and the unique features they had. Second, we introduce the findings and their meaning for classroom practice. Finally, we draw conclusions and indicate the limitations of the studies, as well as directions for future research.

13.1.1 Design of the Studies Conducted

13.1.1.1 Common Features

The studies were conducted in an online inquiry-learning context, with each of the four studies focusing on one of the steps of the feedback-giving process:

- Comparing learning from giving feedback to peers while being provided with assessment criteria or not—Step 1;
- Investigating the effect of the level and type of reviewed products on reviewers' learning—Step 2;
- Comparing reviewers' learning when providing feedback in the form of comments and grades—Step 3.

In all studies, students gave feedback using an online tool. According to a meta-analysis by H. Li et al. (2020), computer-facilitated methods of giving feedback had positive effects on students' learning, in some cases even more than paper-based methods. In our contexts, the choice of an online tool was also supported by several considerations. First, with the help of this tool, students could give feedback anonymously. Previous research has shown that interpersonal relationships can influence the process of giving feedback, and anonymity helps to eliminate possible negative influence (e.g., Rotsaert et al., 2018). Second, students could give feedback at their own pace, which not only makes it convenient, but could also increase their ownership of their learning (e.g., Rosa et al., 2016). Giving students an opportunity to work at their own pace can be especially welcome during a standard lesson, as it is not always easy to differentiate students' work in this way.

Finally, the use of an online tool for giving feedback allowed smooth embedding in an inquiry-learning lesson.

Inquiry learning imitates the scientific research cycle and facilitates students' following of this cycle. Inquiry learning with appropriate guidance can be beneficial for students' cognitive development; for example, a meta-analysis by Furtak et al. (2012) reported an overall mean effect size of 0.5. Adding a feedback-giving activity in an inquiry-learning context makes the inquiry-learning cycle even closer to the real research cycle, as giving feedback on peers' products (such as articles, presentations, proposals, etc.) is a natural part of scientists' work. Critiquing peers' learning products and providing suggestions for their improvement allow students to develop conceptual understanding of a topic and scientific reasoning skills (e.g., Dunbar, 2000; Friesen & Scott, 2013). Moreover, giving feedback on peers' products provides students with another opportunity to reflect on and revise their own products, which may also stimulate learning. Therefore, studying the process and learning outcomes of giving feedback to peers in an inquiry-learning context might lead to better understanding of the different aspects involved in giving feedback than studying it in the context of traditional instruction.

Students gave feedback on concept maps in all four studies. This product was chosen for several reasons. First, as creating a concept map is a natural activity during the conceptualisation phase of an inquiry cycle, including this exercise did not break the flow of the lesson (e.g., Pedaste et al., 2015). Second, the product is quite compact, but at the same time requires understanding of the topic. Therefore, reviewing a concept map may be a relatively brief task, yet demonstrate a deeper level of understanding. Finally, research has shown that reviewing concept maps can add conceptual understanding compared to reviewing other products or just creating a concept map (e.g., Chen & Allen, 2017).

13.1.2 Participants

All studies were conducted with upper secondary-school students as participants, who are not the usual target group. Studies on peer feedback more often involve university students. There can be different reasons for that: researchers teaching at a university may have easier access to this audience, university students may seem to be more ready for feedback-giving activities, or university courses may seem more fit for such tasks than school lessons. The present series of studies allows for better understanding of the feedback-giving process in secondary school and the factors that influence the learning stimulated by it.

Participants were secondary school children (14–15 years old) from Dutch and Russian schools. They worked on a lesson on physics or chemistry from their curriculum in which a feedback-giving activity was included. For each study, students in each class were randomly assigned to the experimental conditions of that particular study. This was done to balance a possible difference between the classes.

13.1.3 Design and Procedure

The studies were experimental, using a pre-test post-test design. Participants worked individually in an online inquiry-learning environment that covered a topic from their physics or chemistry curriculum. The environment was built using the Go-Lab ecosystem (www.golabz.eu) and followed the stages of an inquiry cycle: orientation, conceptualisation, experimentation, conclusion and discussion (Pedaste et al., 2015). In each stage students were provided with some guidance for the inquiry process via specifically designed tools, but the learning process was still regulated by students themselves, as they could decide how to interact with the material and at what pace to move through it.

In the conceptualisation phase, students were asked to create a concept map with the key concepts of the topic they were studying. They made their concept maps using a special tool—Concept Mapper. The tool had some pre-defined concepts and link names, but also gave students an opportunity to add new concepts and link names. A view of the tool is given in Fig. 13.1.

In the investigation phase, students worked in an online lab checking the hypotheses they had created to answer the research question for the lesson. Figure 13.2 presents an example of an online lab.

In the discussion phase, students were asked to give feedback on two learning products (mainly concept maps; answers to open-ended questions were used in one

Fig. 13.1 View of the concept Mapper tool (covering the topic of Study 3)
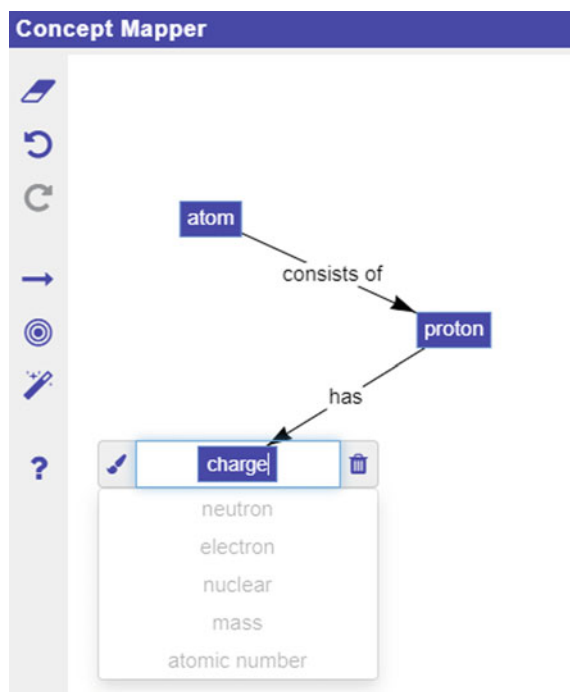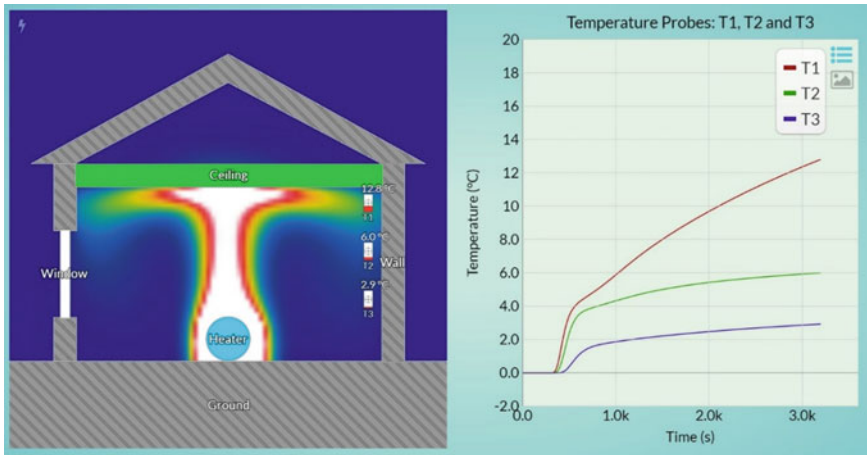
Fig. 13.2 View of the online lab "Vertical temperature gradients" (used in Study 4). Images by The Concord Consortium, licensed under CC-BY 4.0. https://concord.org/

condition in one study) by fictitious peers created by the researchers. To make the context more realistic, students were told that these concept maps came from students from a different class or a different school. Creating the concept maps was done in collaboration with the teachers of participating classes. One reason for that was to ensure that the products were similar to ones created by students and fit the learning material. The other reason was to make the products to be reviewed have a specific level of quality. In particular, all learning products (concept maps and answers to open-ended questions) included some misconceptions and had some room for improvement. Students were guided through the feedback-giving process by the assessment criteria (apart from one condition in one study) formulated in the form of questions and aimed at indicating the desired features of the product. Such prompts have been shown to be helpful for the feedback-giving process (e.g., Gan & Hattie, 2014). The whole process of giving feedback was done in a special peer-assessment tool. This tool allowed students to see the reviewed product and the assessment criteria, and to provide their comments about the product. An example of a fictitious-peer concept map (covering the topic of Study 3) with assessment criteria is given in Fig. 13.3.

After providing feedback on peers' products, students were encouraged but not obliged to revisit their own concept map and change it based on their newly acquired knowledge.

The design of the studies and their target group create a unique combination that allows us to see in what ways giving feedback to peers can be used in less-usual settings (such as an online inquiry environment), and what lessons can be learned for more general usage. This contributes to knowledge about and understanding of the feedback-giving process.
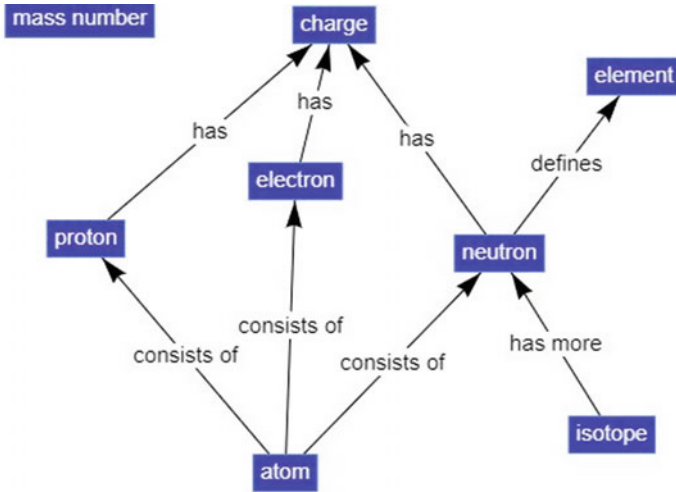
Fig. 13.3 View of the feedback tool

13.1.4 Results and Recommendations for Practice

For each study, this section presents a rationale based on the existing research, a brief description of the specific details distinguishing it from other studies, results obtained and our recommendations based on the results obtained.

13.1.5 The Role of Assessment Criteria

The first step of the feedback-giving model used in the studies is to define assessment criteria (Sluijsmans, 2002). The literature presents two opposite approaches.

According to several studies, assessment criteria can support and guide students in the evaluation process, as they indicate the desired characteristics of the reviewed products; students need such guidance, as providing meaningful feedback can be a challenging task (e.g., Gan & Hattie, 2014; Gielen & De Wever, 2015; Panadero et al., 2013). One approach, therefore, takes providing assessment criteria as necessary for better learning results. The other approach points out that using students' own criteria might be easier for them than understanding ones that are given, especially for complex subjects (e.g., Jones & Alcock, 2014; Orsmond et al., 2000). And if students cannot interpret criteria that are given because these criteria are too difficult or abstract for their level of knowledge, they cannot provide feedback and learn from that process.

Previous research has not led to a clear conclusion about the contribution that being provided with assessment criteria makes to reviewers' learning, and our study did not clarify the situation. In the study investigating that aspect, one group of students ($n = 49$) gave feedback on concept maps using provided assessment criteria. These assessment criteria were not topic-dependent, but focused on the important features of concept maps instead (see Fig. 13.3 for a view of assessment criteria). The other group of students ($n = 44$) had to come up with their own assessment criteria to review the same concept maps. We found no statistically significant difference in post-test scores (controlling for prior knowledge) between the participants who had been provided with assessment criteria and those who had not. However, the results indicated that students could still give meaningful and content-related feedback even if they were not supported by assessment criteria. These findings are in line with previous research suggesting that secondary school students do not necessarily have to be given assessment criteria to provide usable feedback to peers (e.g., Tsivitanidou et al., 2011).

These results are important for designing a feedback-giving activity in a real classroom. As there was no difference found in learning between two conditions, we can say that in our case, not providing students with assessment criteria did not lead to less learning. In other words, this may suggest that teachers can choose whether to give assessment criteria or not, depending on the situation. For small-scale products, not giving assessment criteria may even be more time-efficient, as teachers and students do not spend time on explaining and understanding the criteria. Teachers may instead focus their effort on explaining to students the benefits of giving feedback or discussing what helpful feedback can look like.

13.1.6 The Role of the Quality and Type of Reviewed Products

The second step of the feedback-giving model concerns judging a peer's performance or product. Two studies were conducted to investigate this step.

The first study zoomed in on reviewing products of different quality. According to Hattie and Timperley (2007), evaluating peers' products includes several cognitive activities, such as analysing the existing state of a product, comparing it against assessment criteria, and thinking of directions for improvement based on

identified problems or mistakes. These activities can definitely be influenced by the quality of the products under review. Low- and high-quality products not only have a different number of mistakes, but the mistakes (or areas for improvement) are different and may require different types of analysis and solutions. In other words, they may require different thinking processes from a reviewer and different content in the feedback provided. To fully interact with products of different levels, students should have enough knowledge and understanding to give meaningful feedback (e.g., Alqassab et al., 2018b), which may mean that reviewers' prior knowledge can play a role in the reviewing process and its outcomes. The same product may be challenging yet understandable for a student with higher prior knowledge, and beyond understanding for a student with lower prior knowledge. In such a case, the former student may learn a lot by analysing the product and thinking of possible improvements, while the latter may be overwhelmed and quit the process. However, finding mistakes and providing recommendations is not the only way of learning by reviewing. Students can learn when reviewing good examples, as they can see successful strategies for completing the task and may implement them later (Alqassab et al., 2018a; Tsivitanidou et al., 2018).

In our study about the level of reviewed products, students had to review one of three pairs of concept maps: two low-quality concept maps (29 students), two high-quality concept maps (25 students) or a mixed-quality set (23 students). The results showed that students reviewing a lower quality set had higher post-test scores (controlling for prior knowledge) than students reviewing a higher quality set [$p = 0.048$; $M_{LOW} = 6.39$, $SE = 0.50$, $M_{HIGH} = 5.01$, $SE = 0.47$]. In addition, the quality of the feedback provided by these students was also higher than in the other two conditions, with a statistically significant difference between groups reviewing low-quality and mixed-quality concept maps [$p = 0.033$; $M_{LOW} = 2.43$, $SD = 1.07$, $M_{MIXED} = 1.82$, $SD = 0.90$].

A similar rationale led to studying learning from reviewing different types of products—the contribution to the reviewer's learning could differ. In this study, one group of students was asked to give feedback on concept maps ($n = 66$), while the other group reviewed answers to open-ended questions ($n = 61$). On the one hand, concept maps can stimulate deeper thinking because of their nature. Giving feedback on a product that visualises connections between key concepts for the topic may lead to deeper understanding and, thus, to greater conceptual learning (e.g., Chen & Allen, 2017). On the other hand, identifying mistakes or misconceptions in such a complex product as a concept map can be more (or too) challenging than in a more straightforward and familiar product such as answers to opened-ended questions. As the ability to spot mistakes and provide suggestions is connected to learning (e.g., Adams et al., 2019), reviewing a more complex product (concept map) could lead to less learning than reviewing a less complex one (answers to test questions).

The study did not show a statistically significant difference in mean post-test scores (controlling for prior knowledge) between the conditions reviewing concept maps and answers to open-ended questions. However, it is noteworthy that the quality of the feedback provided was found to predict post-test scores for both

conditions [$F(2, 122) = 7.95$, $p < 0.01$, $R^2 = 0.12$], with a regression coefficient of 0.57. And this quality was higher in the condition reviewing answers to tests questions than in the condition reviewing concept maps [$t(123) = -2.37$, $p = 0.019$; $M_{TEST} = 3.18$, $SD = 1.90$, $M_{CONCEPT} = 2.53$, $SD = 1.14$].

These findings could suggest that students felt more comfortable with and, as a result, were better at giving feedback on lower quality and more familiar and straightforward products than on higher quality and more complex ones, as they could see more mistakes and make more suggestions. Being able to give better feedback led to better learning outcomes.

There are several implications for practice based on these results. First, as the quality of the feedback given predicted reviewers' learning, it is important to encourage students to give feedback thoughtfully. Second, the type of product to review does not seem to influence learning as long as students give high-quality feedback. There is no known universal way to increase the quality of feedback provided by students. Apart from explaining to students the benefits of giving feedback, teachers may introduce elements of evaluative judgement into a classroom routine as a way to practice this. In this way, students may develop their assessment skills without being specifically trained for peer assessment. Finally, to maximise reviewers' learning, they should be providing feedback on products of the same or lower level of quality than their own current level of performance. This means that if teachers use fictitious-peer work for reviewing, they need to find pieces at the average or below-average level. And if they implement a full peer-assessment process, their matching strategy should assign students of approximately the same level to give feedback to each other.

13.1.7 The Way of Giving Feedback

The third step of the model is to provide feedback for future learning. The form this feedback takes can influence the learning arising from it. In our study, one group of students gave feedback in the form of comments ($n = 46$), while the other group provided feedback with grades using smileys ($n = 47$). In both conditions, students were supported by assessment criteria, which were formulated as questions for the comment condition and as statements for the smiley condition.

Several studies have shown that commenting leads to more learning by reviewers than grading (e.g., Wooley et al., 2008; Xiao & Lucking, 2008). This body of research suggests that commenting triggers more learning, as students are more cognitively involved with the material for a longer time than while grading, as they not only had to evaluate their peer's work and identify areas of improvement, but also had to think of solutions. However, with smaller scale products, the difference in time (and probably effort) between reviewing by commenting and by grading might be not so obvious as with a larger scale product. Therefore, checking if these findings still stand for small-scale products can enrich our understanding of the feedback-giving process.

Our study confirmed the existing point of view—students in the commenting condition had higher post-test scores (controlling for prior knowledge) than students who graded peers' concept maps with smileys [$F(1, 87) = 5.84$, $p = 0.018$, $\eta_p^2 = 0.06$; $M_{COMMENT} = 5.23$, $SD = 0.33$; $M_{SMILEY} = 4.09$, $SD = 0.34$]. Moreover, a differential effect of commenting for different prior knowledge groups was found, with low-prior-knowledge students benefiting from commenting the most [$F(2, 87) = 4.19$, $p = 0.018$, $\eta_p^2 = 0.09$]. This backs up our idea that prior knowledge can be an influential factor in the learning of feedback providers. Obviously, students need to be knowledgeable enough to provide meaningful feedback (e.g., Alqassab et al., 2018b; van Zundert et al., 2012), but apparently commenting helped even low-prior-knowledge students to get cognitively involved with the concept maps. The fact that they could see some mistakes and comment on them was most likely enough to trigger their learning. These findings support our belief that students with any level of prior knowledge can benefit from giving feedback if this process is properly designed.

These results can be used as a basis for recommendations on incorporating the feedback-giving process into classroom practice. First, teachers should be aware of the fact that students may learn differently from giving feedback depending on their prior knowledge. When organising a feedback-giving activity in an online platform, this can be taken into account by using different settings for different groups of students. And second, as commenting was shown to contribute to reviewers' learning more than grading, students should be given an opportunity to write comments when asked to provide feedback. Reviewing small-scale products is a brief activity that can fit within the usual classroom routine, but still confer all of the benefits of reviewing for students' learning.

13.2 Conclusion

When properly organized, giving feedback to peers can be a learning experience for a feedback provider even when reviewing a small-scale product. This makes giving peer feedback more applicable in a real classroom situation, as teachers do not have to change a lot in the lesson to include a feedback-giving activity for a smaller product. This may allow students not only to be cognitively involved with the material, but also to be involved at a meta-level, as evaluating a peer's product with given or self-created assessment criteria and providing appropriate feedback may require higher order thinking than just completing a task. Peer feedback can also be a valuable addition to an inquiry-learning lesson, as it allows students to reflect on their exploration process and in that way to deepen it.

Using online platforms (such as www.golabz.eu) can make giving feedback more natural and easier than in traditional instruction, due to the ability to configure parameters of the feedback-giving process according to the learning goals. Although research on this topic is ongoing, peer assessment should be implemented in secondary schools more often, with a view to benefiting feedback providers.

There are several limitations or considerations regarding the studies conducted. First, the studies isolated the feedback-giving part of the peer-assessment process, while in a real-life situation students usually fulfil both roles: feedback provider and feedback recipient. In a real classroom, teachers have two choices: they can either follow the experimental settings and ask students to give feedback only (for example, on learning products from previous cohorts), or they can use a full peer-assessment process with the idea that at least the feedback-giving part could stimulate learning. Moreover, an interesting direction for further research in this area can be checking the findings of these studies in the situation of a reciprocal peer-assessment process.

Second, the experimental studies used fictitious products. The limitation associated with this is that even though the products to be reviewed were created in cooperation with teachers, they might still have differed from those created by students. Therefore, an interesting follow-up of this series of studies could be an experiment comparing students' feedback given on fictitious and real peers' products. This will help to explore if the students' responses differ and in what way. If teachers would like to use the results of the conducted studies and control the quality of reviewed products in a real classroom, they can do so by using pieces of work by students from previous cohorts, for example.

Third, the instruments used to measure students' learning were researcher-developed, and differed in different studies. As our intention was to study the process of giving feedback to peers in as natural an environment as possible, we always developed lessons based on the curriculum used by the participating classes. Using different STEM topics covered in secondary school supports the idea that our approach can be implemented for different domains. However, the drawback of this approach was that we could not use the same testing instruments and they had to be developed specifically to address the learning of the content in an isolated lesson or a series of lessons used for the studies. It could be interesting to validate these instruments by conducting a larger scale study; however, it could also be quite challenging in practice.

Finally, due to the scarce number of studies conducted with secondary schoolchildren as a target group, we sometimes used findings obtained for university students to set the expectations for our studies. The differences between these target groups may pose risks to the external validity of the studies conducted. This means that more experimental studies should be carried out in the field of peer assessment aiming at different target groups and domains to enrich our knowledge about this process.

At a more general level, further research on the feedback-giving process can take several directions. First, as higher quality feedback provided by students was associated with higher learning gains for them, it is important to investigate factors that lead to giving poor-quality feedback. Knowing this may help with developing ways to increase the quality of feedback given. Second, as the inquiry-learning context could have provided a unique and quite natural context for giving feedback, it could be interesting to check whether the results obtained in these studies hold for giving feedback on other products in an inquiry context. Finally, several studies

indicated a positive effect of training students in giving feedback, but these studies targeted a quite elaborate procedure of giving feedback on bigger scale products, such as an essay, a report, or even a thesis. With complex and elaborated products training seems like an important contributor to learning, but it is worth studying whether training is equally important when feedback is given on smaller scale products and what the desired format for such training could be.

References

Adachi, C., Tai, J.H.-M., & Dawson, P. (2018). Academics' perceptions of the benefits and challenges of self and peer assessment in higher education. *Assessment & Evaluation in Higher Education, 43*, 294–306. https://doi.org/10.1080/02602938.2017.1339775

Adams, R., Jackson, P., Lenton, K., Dugdale, M., Whittaker, C., Lasry, N., & Charles, E. S. (2019, May). *Error-detection tasks and peer feedback for engaging physics students.* Paper presented at the Fifteenth Conference on Education and Training in Optics and Photonics: ETOP 2019, Quebec City, Quebec.

Alqassab, M. (2017). *Peer feedback provision and mathematical proofs: Role of domain knowledge, beliefs, perceptions, epistemic emotions, and peer feedback content [Unpublished doctoral dissertation].* Ludwig-Maximilians University.

Alqassab, M., Strijbos, J. W., & Ufer, S. (2018a). The impact of peer solution quality on peer-feedback provision on geometry proofs: Evidence from eye-movement analysis. *Learning and Instruction, 58*, 182–192. https://doi.org/10.1016/j.learninstruc.2018.07.003

Alqassab, M., Strijbos, J. W., & Ufer, S. (2018b). Training peer-feedback skills on geometric construction tasks: Role of domain knowledge and peer-feedback levels. *European Journal of Psychology of Education, 33*, 11–30. https://doi.org/10.1007/s10212-017-0342-0

Canty, D., Seery, N., Hartell, E., & Doyle, A. (2017, July 10–14). *Integrating peer assessment in technology education through adaptive comparative judgment.* Paper presented at the PATT34, Millersville University, PA, USA.

Chen, W., & Allen, C. (2017). Concept mapping: Providing assessment of, for, and as learning. *Medical Science Educator, 27*, 149–153. https://doi.org/10.1007/s40670-016-0365-1

Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science, 39*, 629–643. https://doi.org/10.1007/s11251-010-9146-1

Dunbar, K. (2000). How scientists think in the real world: Implications for science education. *Journal of Applied Developmental Psychology, 21*, 49–58. https://doi.org/10.1016/S0193-3973(99)00050-7

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*, 287–322. https://doi.org/10.3102/00346543070003287

Friesen, S., & Scott, D. (2013). *Inquiry-based learning: A review of the research literature.* Paper prepared for the Alberta Ministry of Education.

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research, 82*(3), 300–329. https://doi.org/10.3102/0034654312457206

Gan, M. J. S., & Hattie, J. (2014). Prompting secondary students' use of criteria, feedback specificity and feedback levels during an investigative task. *Instructional Science, 42*(6), 861–878. https://doi.org/10.1007/s11251-014-9319-4

Gielen, M., & De Wever, B. (2015). Structuring the peer assessment process: A multilevel approach for the impact on product improvement and peer feedback quality. *Journal of Computer Assisted Learning, 31*, 435–449. https://doi.org/10.1111/jcal.12096

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112. https://doi.org/10.3102/003465430298487

Ion, G., Sánchez-Martí, A., & Agud-Morell, I. (2019). Giving or receiving feedback: Which is more beneficial to students' learning? *Assessment & Evaluation in Higher Education, 44*, 124–138. https://doi.org/10.1080/02602938.2018.1484881

Jones, I., & Alcock, L. (2014). Peer assessment without assessment criteria. *Studies in Higher Education, 39*, 1774–1787. https://doi.org/10.1080/03075079.2013.821974

Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education, 45*, 193–211. https://doi.org/10.1080/02602938.2019.1620679

Li, L., & Grion, V. (2019). The power of giving feedback and receiving feedback in peer assessment. *AISHE-J: The All Ireland Journal of Teaching and Learning in Higher Education, 11*(2).

Liu, N.-F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*, 279–290. https://doi.org/10.1080/13562510600680582

Lu, J., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science, 40*, 257–275. https://doi.org/10.1007/s11251-011-9177-2

Lu, J., & Zhang, Z. (2012). Understanding the effectiveness of online peer assessment: A path model. *Journal of Educational Computing Research, 46*, 313–333. https://doi.org/10.2190/EC.46.3.f

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing, 18*, 30–43. https://doi.org/10.1016/j.jslw.2008.06.002

Nicol, D., & McCallum, S. (2022). Making internal feedback explicit: Exploiting the multiple comparisons that occur during peer review. *Assessment & Evaluation in Higher Education, 47*(3), 424–443. https://doi.org/10.1080/02602938.2021.1924620

Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education, 25*, 23–38. https://doi.org/10.1080/02602930050025006

Panadero, E., Romero, M., & Strijbos, J.-W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation, 39*(4), 195–203. https://doi.org/10.1016/j.stueduc.2013.10.005

Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: How students respond to peers' texts of varying quality. *Instructional Science, 43*, 591–614. https://doi.org/10.1007/s11251-015-9353-x

Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review, 14*, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003

Phillips, F. (2016). The power of giving feedback: Outcomes from implementing an online peer assessment system. *Issues in Accounting Education, 31*, 1–15. https://doi.org/10.2308/iace-50754

Rosa, S. S., Coutinho, C. P., & Flores, M. A. (2016). Online peer assessment: Method and digital technologies. *Procedia-Social and Behavioral Sciences, 228*, 418–423. https://doi.org/10.1016/j.sbspro.2016.07.064

Rotsaert, T., Panadero, E., & Schellens, T. (2018). Anonymity as an instructional scaffold in peer assessment: Its effects on peer feedback quality and evolution in students' perceptions about peer assessment skills. *European Journal of Psychology of Education, 33*(1), 75–99. https://doi.org/10.1007/s10212-017-0339-8

Sluijsmans, D. M. A. (2002). *Student involvement in assessment. The training of peer assessment skills* [Unpublished doctoral dissertation]. Open University of the Netherlands, The Netherlands.

Tsivitanidou, O. E., Constantinou, C. P., Labudde, P., Rönnebeck, S., & Ropohl, M. (2018). Reciprocal peer assessment as a learning tool for secondary school students in modeling-based

learning. *European Journal of Psychology of Education, 33*, 51–73. https://doi.org/10.1007/s10212-017-0341-1

Tsivitanidou, O. E., Zacharia, Z. C., & Hovardas, T. (2011). Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction, 21*, 506–519. https://doi.org/10.1016/j.learninstruc.2010.08.002

van Popta, E., Kral, M., Camp, G., Martens, R. L., & Simons, P. R.-J. (2017). Exploring the value of peer feedback in online learning for the provider. *Educational Research Review, 20*, 24–34. https://doi.org/10.1016/j.edurev.2016.10.003

van Zundert, M. J., Könings, K. D., Sluijsmans, D. M. A., & van Merriënboer, J. J. G. (2012). Teaching domain-specific skills before peer assessment skills is superior to teaching them simultaneously. *Educational Studies, 38*, 541–557. https://doi.org/10.1080/03055698.2012.654920

Wooley, R., Was, C., Schunn, C. D., & Dalton, D. (2008, July). *The effects of feedback elaboration on the giver of feedback.* Paper presented at the 30th Annual Meeting of the Cognitive Science Society.

Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *The Internet and Higher Education, 11*, 186–193. https://doi.org/10.1016/j.iheduc.2008.06.005

14

Peer Interaction Types for Social and Academic Integration and Institutional Attachment in First Year Undergraduates

Emmeline Byl, Keith J. Topping, Katrien Struyven, and Nadine Engels

14.1 Introduction

The educational agenda is more than ever dominated by heightened demands on student attraction, retention and graduation (Vossensteyn et al., 2015). Newly entering students increasingly struggle with issues related to their transition and adjustment to university education (Berger et al., 2012; Hagedorn, 2006; Tinto, 2003, 2010). The lack of opportunities for safe and frequent social interaction hinders newcomers' socialisation and learning (Lowe & Cook, 2003; Wu, 2013). This is particularly true for those who commute to university, work and/or attend part-time (Gillies & Mifsud, 2016).

Higher education institutions internationally are increasingly implementing various peer tutoring and peer mentoring strategies to support newly enrolled students' transition into university, aiming to reduce drop-out and improve persistence. More

E. Byl (✉)
Imec.Istart, Imec, Ghent, Belgium
e-mail: Emmeline.Byl@imec.be

K. J. Topping
School of Education, University of Dundee, Dundee, Scotland

K. Struyven
School for Educational Studies, Hasselt University, Hasselt, Belgium

N. Engels
Teacher Education Department, Vrije Universiteit Brussel, Brussels, Belgium

305

and more studies report significant outcomes from such initiatives in student integration, commitment or persistence (e.g., Andrews & Clark, 2009; Pleschová & McAlpine, 2015). However, it has been relatively rare for different systems to be directly compared within the same institution.

This chapter investigates the impact of different strategies on opportunities for social development and academic learning in the first year of higher education in one university. Student-to-student interactions were developed face-to-face and technology was used to increase students' initial and continued participation. The aim was twofold: (1) to examine and compare the effects of peer mentoring and peer tutoring on students' perceptions of their social integration, academic integration and persistence; (2) to explore which aspects of each intervention students considered successful or otherwise, and explore suggestions for improving effectiveness.

14.2 Previous Research

Studies show that the more students are academically and socially involved, the more likely they are to persist and graduate (Tinto & Pusser, 2006). Fostering students' integration has become an important educational objective, certainly since adequate academic integration is often considered to deepen learning and is correlated with more active cognitive processing, better understanding and improved performance (Torenbeek, et al., 2010; Zepke, et al., 2006).

Also emphasized in research is the role of student support and needs-based aid for university students' social and academic integration, particularly in the first year of study (Astin, 1993; Carter, et al., 2013; Tinto, 2003) and achievement (Crosier, et al., 2007; Dukakis, et al., 2007). The literature contains many recommendations of strategies and initiatives that support students in becoming more active participants in all facets of university life (Tinto & Pusser, 2006). Researchers recently suggested that peer tutoring and mentoring are effective and relatively simple ways for students to become more active members of the university community (Rayle & Chung, 2008; Torenbeek, et al., 2010; Wilcox, et al., 2005).

As a working definition, we adopt that of Topping et al. (2017), i.e., peer tutoring is "people from similar social groupings who are not professional teachers helping each other to learn and learning themselves by teaching" (p. 10). It is characterized by specific role-taking and high focus on curriculum content. By contrast, Topping's definition of peer mentoring is "an encouraging and supportive one-to-one relationship with a more experienced worker (who is not a line manager) in a joint area of interest, characterized by positive role modelling, promoting raised aspirations, positive reinforcement, open-ended counselling and joint problem-solving" (Topping & Ehly, 1998, p. 9). It engages with broader issues than curriculum content. Both students benefit when they are able to help each other (Copeland, et al., 2002).

Previous research provides empirical support for the positive effects of peer tutoring in diverse instructional settings with outcomes ranging from cognitive and meta-cognitive gains to affective and social-motivational benefits for both peer tutors and tutees (Falchikov, 2001; Topping, 2005). Peer tutoring participants demonstrate better performance and higher academic achievement (Bronstein, 2008; Fayowski & MacMillan, 2008; Ning & Downing, 2010; Peterfreund et al., 2007). These often come from improved understanding of content (Dobbie & Joyce, 2008; Smith et al., 2007; van der Meer & Scott, 2009), critical thinking (Stigmar, 2016), transfer and autonomy of learning (Stigmar, 2016), profound knowledge-construction after applying deeper and strategic learning strategies (Dobbie & Joyce, 2008; Smith et al., 2007; van der Meer & Scott, 2009) and development of transferable academic skills (Court & Molesworth, 2008; Ning & Downing, 2010). Students perceive peer tutoring settings as safe learning environments that stimulate tutors' and tutees' self-confidence (Ford, et al., 2015), heighten their wellbeing (Bronstein, 2008), and lower any uncertainty (Court & Molesworth, 2008).

Additionally, peer tutoring appears to result in higher motivation (Stigmar, 2016) and connectedness (Dobbie & Joyce, 2008; Smith et al., 2007; van der Meer & Scott, 2009), as well as increased academic satisfaction (Robinson et al., 2005). Peer tutoring participants further report social benefits (Court & Molesworth, 2008) and improved communication behaviour (Ford et al., 2015). Although positive effects on students' retention are reported (Bronstein, 2008; Peterfreund et al., 2007), these cannot always be confirmed in other studies (e.g., Carr et al., 2016). By contrast, research clearly confirms students' appreciation of peer tutoring, both when providing and when receiving academic help (Ginsburg-Block, et al., 2006; Griffin & Griffin, 1998; Topping et al., 1997).

During the last decade, educational research has also provided empirical support for the positive effects of peer mentoring in higher education settings, including performance, intellectual and skills gains but also emotional benefits and other non-cognitive results for both peer mentors and mentees (Outhred & Chester, 2010). Peer mentoring participants demonstrate better performance (Amaral & Vala, 2009; Fox, et al., 2010; Goff, 2011; Smith, 2009) and higher academic knowledge (Bullen et al., 2010). By providing positive role models for the students (Lahman, 1999; Twomey, 1991), peer mentoring is often related to increased development of values and skills (Bullen et al., 2010; Hall & Jaugietis, 2010), as well as more profound listening skills (Lee et al., 2010). As a result of working and socializing with peers, students' participation in classes and extra-curricular activities is higher (Bittich & Rongen, 2007; Copeland, et al., 2002) and they are academically and socially more integrated (Elster, 2014; Pascarella & Terenzini, 2005).

What sort of tasks provided the context for peer interaction? In peer tutoring the tasks were determined by the tutee but drawn from the curriculum context set by the instructors, and in that sense were more aligned to the instructional design literature (Biggs, 1996). In peer mentoring the tasks were those the mentee chose to raise according to the extent to which they were concerning, and in this

sense related more to constructivist learning theory (Biggs, 1996), a family of theories all having in common the centrality of the learner's activities in creating meaning. Biggs's notion of "constructive alignment" sought to marry the two thrusts of instructional design and constructivism. Likewise, Wenger invented and then explored the concept of "community of practice" (Farnsworth et al., 2016), conceptualising identity and participation in order to develop a social theory of learning in which power and boundaries are inherent. This clearly relates to a context in which peers form communities in which interaction with each other regarding social and academic objectives and problems is regarded as natural.

As peer interaction strategies became more widespread and popular (Duron, et al., 2006), more diverse programmes and formats appeared, as well as new lines of research (Maheady & Gard, 2010; Maheady, et al., 2006; Roscoe & Chi, 2007). Peer interaction is becoming increasingly significant (Aljohani, 2016; James, et al., 2010; Muldoon & Wijeyewardene, 2012), and has become key for both the acquisition of innovation and creativity (Johansson, 2004) and interdisciplinary thinking (Johansson, 2004). Spontaneous forms of peer interaction might have potential that seem to be underestimated or underused. By taking a focus on peer interaction in informal learning environments and outside-class contexts, particularly during the first semester at university, in this study we contribute to current research.

In the literature, the concept of student integration is mostly discussed via environmental and (symbolic) interactional theories of social and academic integrative learning (Tinto, 1993). In contrast to the notions of cognitive development (e.g., Piaget, 1987), learning is conceived of as a collective process of maturation (Burgess, 2016). People are perceived as active agents and contributors to social life, with the ability to negotiate, share and create a distinct peer culture in collusion with other 'more experienced' others (Corsaro, 2005) to absorb the norms and values of the surrounding society (Burgess, 2016). "The more students are academically and socially engaged, the more likely they are to succeed. Such engagements lead not only to social affiliations and the social and emotional support they provide, but also to greater involvement in learning activities and the learning they produce. Both lead to success in the classroom" (Tinto, 2006).

Integrative learning, as recently formulated by Tinto (2012), focuses attention on the integration and translation of academic spheres and divergent domains of knowledge, culture, and social practice. According to Tinto (2015) "academic integration is the extent to which students adapt to the academic way-of-life." (Tinto, 1993). Academically well-integrated students have the *willingness* to belong to a group and the *ability* to belong to one (Severiens & Wolff, 2008). Social integration is the degree to which students adapt to and familiarize themselves with the social university environment (Rienties et al., 2012). Successful socially-integrated students have many friends at university, feel at home, take part in extra-curricular activities and feel connected to fellow students and teachers (Bittich & Rongen, 2007; Severiens & Wolff, 2008).

Consequently, while the present study will discuss both academic and social integration in university, the main focus will be on social integration and its crossover effects on academic integration.

14.3   Method

14.3.1 Design

The study used a sequential, mixed-methods design (Creswell & Clark, 2011), which was intended to maximise participation and triangulate findings. Quantitative and qualitative data were collected throughout the years with students registered for the first-time in the first year of a bachelor programme at the Faculty of Psychology and Educational Sciences, at a Dutch-speaking university in a large city in the north of Belgium, in three consecutive academic years. A clear control design could not be used due to ethical considerations. Therefore, a non-randomised control group of non-participants was formed for all comparisons. The design was thus quasi-experimental.

14.3.2 Sample

We invited all 842 students to volunteer to participate in the survey, and of these 731 (87%) eventually completed the survey. From these 731 students, students who had been studying for more than one year in the faculty (n = 285) were removed from the student population because they did not meet the inclusion criteria. In the end, a sample size of 446 (61%) unique students were included. The sample population of first-year students were 26% in the Department of Educational Sciences (N = 115) and 74% in the Department of Psychology (N = 360). The majority (65%) were registered for the first time in a bachelor degree programme in higher education (N = 291). There were four times as many female students (81%; N = 360) as male students (19%; N = 86).

Participants were recruited via a snowball technique which relied heavily on email and text messages. First students were invited face-to-face to complete the survey. Then we asked them by email and via the e-learning platform. After three weeks, the students received a reminder. After six weeks, the students who did not fill in the survey were personally reminded by email. After nine weeks, the students who still did not fill in the survey were personally reminded by mobile phone text message. This was followed-up with invitations to a Facebook group. In this way, students who would not have noticed traditional messaging participated, and the number of participants was close to the total numbers in these departments.

For the qualitative follow-up semi-structured interviews, participants were 39 self-selected volunteers stratified from each of the peer interaction initiatives. Of the 39 respondents, 23 (59%) were students who participated in peer mentoring, and 20 (51%) were students who participated in peer tutoring. Academic disciplines were almost equally divided between psychology (49%; N = 19) and educational science (51%; N = 20). There were three times as many female students (74%; N = 29) as male students (26%; N = 10).

14.3.3 Measures

In the quantitative part, data collection involved the development, delivery and collection of online questionnaires using the Qualtrics software. Surveys were administered at the start of the second year to investigate newcomers' perceptions after one year of experience of higher education. In the qualitative part, individual in-depth interviews with first-year students who participated in peer mentoring or peer tutoring or both were conducted in order to obtain a deeper understanding of their experiences.

The online questionnaire incorporated measures of social integration, academic integration, academic commitment, commitment attitude and institutional attachment. Three instruments of known reliability were administered: the Social Adjustment, Academic Adjustment and Institutional Attachment subscales of the Adaptation to College Questionnaire (Baker & Siryk, 1989); the Commitment subscale of the Revised Academic Hardiness Scale (Benishek, et al., 2005); and the Commitment Attitude Scale (Solinger, et al., 2015). A seven-point Likert scale, on a continuum ranging from 1 (does not apply to me at all) to 7 (applies to me very well) was used. The subsequent reliability of these measures (Cronbach's Alpha) was high for Social Integration and Social Adjustment (0.90), Social Engagement (0.83) and Institutional Attachment (0.82), but less so for Academic Integration and Motivation (0.74), Academic Application (0.75) and Academic Performance (0.80).

Interviews were based on the three stages of Appreciative Inquiry (AI): Discovery, Dream and Design (Barrett, 1995; Cooperrider, et al., 2003; Whitney & Trosten-Bloom, 2010). AI is an innovative participative research approach (Czarniawska-Joerges, 1996) that differs from other current research methodologies (Cooperrider & Srivastva, 1987) by its affordance of a positive, holistic and appreciative lens. It "involves a wondering that can touch the soul" (Kung, et al., 2014). Through its focus on successes and their potential influences in co-creating desired futures, it opens participants' experiences in a generative manner towards ongoing and deepening reflections and move deficit discourse towards deep engagement and contemplative insight within oneself and with others (Kung, et al., 2014). As a form of social constructivist evaluation, AI aims to enable those involved in evaluation to make sense of educational change through dialogue, reflection and interaction.

The interview instrument included only open-ended questions. The first questions to be posed (discovery) asked the participants to focus on their stories of best practice, positive moments, greatest learning and successful processes related to their experiences with one of the activities in which they participated. They were then asked to 'dream' about how those kinds of support systems could be even better (Watkins & Mohr, 2001). Particular attention was paid to asking reflective AI questions related to the question 'when': "the first four weeks", "after one month" and "in the last four weeks". The researcher in the first phase ('discovery') asked the participants to focus on particular experiences that they would describe as being positive and life-centric in nature and to share the essence of

their stories as a means of remembering specific practices, events, processes. In the second phase (dream), she asked them to 'dream' about how it could be even better (Watkins & Mohr, 2001) and imagine an ideal future. In order to determine the strategies to assist them in realising common needs, during the last phase ('design') the researcher encouraged the participants to think about their expectations related to actions and decision in order to make the vision become reality, in the form of an action plan for future practice.

14.3.4 Analysis

The questionnaire data were summarised in SPSS and subjected to analyses seeking to determine whether the responses were statistically different from a random distribution. All interviews were audio-recorded and transcribed by one researcher. Verbatim quotes of frequently occurring issues were documented with hand-written notes by the interviewer throughout the interview process. To help reduce socially desirable answers, each AI phase began with a one-minute independent writing activity in which individual responses related to the open-ended questions were documented with hand-written notes by each participant. The interviews lasted 20–30 min and each question lasted around 3 min. A combined inductive-deductive content-analysis technique was coupled with a thematic analysis technique. One coder reviewed all the interviews twice. MAXQDA 11 was used to analyse the data. In the results we indicate frequency of response themes, providing illuminative quotations.

Analysis of interviews was based on the transcripts, hand-written detailed reports of the researchers and the handwritten one-minute preparations of the students. The transcripts were all conjointly analysed by two researchers using thematic analysis technique, identifying "powerful" themes (van Manen, 1990) in relation to the participants' life-centric experiences using MAXQDATA. This programme had the advantage of making the process of axial coding easier by ordering, dividing and clustering codes into categories, and recognizing structures or patterns. The phases were analysed separately. Inter-rater reliability exceeded 90% (Miles et al., 2018), and informal discussions ensured consensus. We used Hycner's (1985, 1999) systematic procedures to identify essential features and relationships: repeatedly reading each interview, identifying statements of research phenomena, grouping units of meaning to identify significant topics or central themes, checking back with the data to ensure the content had been correctly captured, summarising the transcript of interview, and identifying "themes common to most or all of the interviews as well as the individual variations" (Hycner, 1999), and writing a composite summary.

14.3.5 Technology

Quite apart from the extensive use of technology in the sampling process, it tended to be less important at the beginning when students lacked self-esteem. However, as peer interaction activities developed, technology became more and more important. Later, social media (especially Facebook) played an important role. Indeed, social networking media became indispensable for some. Not only were they used to build friendships and maintain social relations, they were also used to process subject matter and exchange summaries. It was striking that the fear of feelings of loneliness and anxiety were very prominent among certain students.

Technology also had a role in academic integration. Learning did not exclusively take place during particular activities (e.g., revising for an exam) or in particular environments (e.g., the classroom), but was embodied in social environments and everyday life. Higher-year students could, for instance, help with administrative tasks, system navigation or educational knowledge later in the year (e.g., for examinations), by delivering relevant information in consolidated timely bursts via text messages, Facebook groups and emails. Students were involved in a range of community groups, physical places, virtual spaces and social networks in relation to their personal interests. These networks were individually selected by students, shaped and re-negotiated, and spread across physical spaces, friends and peer groups, as well as virtual spaces and online learning platforms.

Thus, there was no direction regarding which applications to use for maintaining contact, and indeed no way in which the institution could sensibly control this. Fashionable applications for exchanging messages change quite quickly, and the students needed to use those for which they were motivated and with which they were familiar. There was no way the institution could keep up with this, and issues of digital ethics could not be policed. Indeed, the fact that such applications clearly belonged to the students and were not part of the institution probably added to the sense of "community of practice". Of course, this may raise issues of privacy, equality and responsibility.

14.4   Results

14.4.1 Quantitative

Peer mentoring participants reported a significantly higher level of social adjustment and social engagement than nonparticipants (t = −2.480, df = 425, p < 0.05). Mentoring participation had a moderate effect on social adjustment (Effect Size d = 0.370) and had a small effect on the difference in social engagement (Effect Size d = 0.315). There were also clear differences between participant and non-participant group scores on peer tutoring, with participants achieving a higher average level of social adjustment and social engagement than non-participants, but this difference did not reach statistical significance. For academic integration, there was a slight difference in the level of academic motivation,

academic application and academic performance between participants and non-participants of peer mentoring, but none of these was significant. An independent samples t-test showed that for students who participated in peer tutoring, academic application scores appeared to indicate a marginal significance for the difference between the participation groups ($t = -1.715$, df = 429, $p < 0.1$). Participation in peer tutoring thus had a small effect on academic application (Effect Size d = 0.306). The differences in average institutional attachment scores between participants and non-participants in peer mentoring were not significant. Although peer tutoring students showed a higher level of institutional attachment compared to non-participants, the differences were not significant.

Thus, as far as the survey data could tell us, peer mentoring appeared better for social integration, peer tutoring appeared better for academic integration, and neither appeared to affect institutional attachment.

14.4.2 Qualitative

Taking peer mentoring first, almost two-thirds of the respondents who participated in peer mentoring claimed that they had a connection with a higher-year student, and more than one-quarter indicated that they had built up friendships. Over half of respondents saw speed dating as very valuable. They claimed that such activities were fundamental contact-making mechanisms between new students which could become sustainable. Over half of respondents claimed that due to peer mentoring they had a connection with the student community and believed that a certain level of similar interests in psychology and human development, together with taking the same courses and/or study/life path, was what bound students together. This is clear in this quote:

> I'm always coming back to the same conclusion: to get connected with the right people. Those students who have the same effort and energy or willingness as you have. (Student 164: woman, first-time, regular student, large group learning context - LGLC).

The speed-dating activities were reported as an effective strategy for networking and searching for a mentor. Less than half of the respondents commented that this was due to the possibility of meeting different students at different times and in a structured manner. One respondent clearly describes this:

> It was good to keep some activities simply cosy; so that people can have simply a talk with each other, enjoying their time, and having fun is paramount. This breaking-the ice is almost everything. So that is also very important. (Student 185: woman, first-time, regular student, small group learning context - SGLC).

Almost two-thirds of the respondents were satisfied with peer mentoring because they were given the opportunity to get in touch with higher year students as well as students of the same year and the same study programme. They

saw this as a unique opportunity to build trust and more personal relationships with those who were willing to share expertise and experiences with each other in common:

> The peer mentoring walks to the centre of Brussels are very captivating because then you get to know each other better, build up a connection. The development of this bonding is important to gain trust in each other, and to enable you to share your problems or difficulties more easily and immediately, or to ask for help if you do not understand course contents. (Student 154: woman, first-time, regular student, SGLC).

For almost two-thirds of the respondents, connecting with higher-year students in their own department was the reason for valuing their experience with peer mentoring. This was also evident in this quotation:

> If you are fretting for weeks, then you have someone you trust and you can call on, and you don't have to think, for example, 'who should I now badger again with my troubles? They shouldn't have time for me,' therefore. (Student 143: woman, first-time, regular student, LGLC).

Because students experienced pleasure and value with their mentors during peer mentoring, they often also spent their free time together. Almost two-thirds of the respondents claimed that such activities were necessary especially for those who were interested in getting involved in social life on campus and wanting to get in touch with peers with whom they could discover life on campus. Almost all mentioned that peer mentoring enabled them to experience the power of social interactions with experienced peers who had recently taken the same path:

> This is imperative for me. Because if you are befriended with higher-year students, you are closer to them and you get more help. It makes you feel better about yourself, and feel safer and more relaxed if you have someone around. And in turn, you will also provide more help to others, which again increases your wellbeing. So yes, these relationships determine if I'm satisfied and experience a certain level of happiness or not, and this consequently predicts the extent to which I will be happy and satisfied with my study and study situation. (Student 185: woman, first-time, regular student, LGLC).

Turning to peer tutoring, nearly all the respondents asserted they connected with other classmates with whom they had the same social learning experience. Almost half of the respondents saw peer tutoring's focus on courses such as 'Logic' and 'Statistics' as very valuable. They claimed that such focus on difficult courses was a fundamental binding mechanism between new students:

> At the beginning of the year, you cannot really imagine how you will pass this course. Further experiences with fast-paced teaching professors and difficult, challenging courses just strengthened this first impression. In such a context, when the understanding of the content and the intent to persist fully becomes the responsibility of the student, peer tutoring makes a big difference when shared learning experiences were provided. (Student 127: man, first-time, regular student, SGLC).

A certain degree of willingness to be involved and persist in learning is what attracted students to peer tutoring and what bound them together, creating the sense of safety needed to start a conversation, to help each other, and to work together. Experiencing the effectiveness of social learning and collaborative colleague support was invariably appreciated by most respondents. The satisfaction that arose from getting the opportunity from the start of their study careers to make contact with their classmates who were open to share experiences was common: peer tutoring was of particular importance for first-year students to increase their social engagement in the faculty.

> Experiences like those in peer tutoring enhance the willingness to interact and to share and to help other members of the faculty. As a student in the third year of the academic bachelor (programme), you can also experience difficulties. Peer tutoring is then also highly relevant. (Student 193: woman, first-time, regular student, LGLC).

Approximately one third of the respondents said that the way they experienced enjoyment with their classmates during the peer tutoring sessions led them to spend more time together to learn. Over two-thirds of the respondents reported that peer tutoring was particularly needed for those interested in the academic challenge of studying. The majority of the respondents indicated that typical, whole-class tutoring at many educational institutions could not match peer tutoring because the latter empowered learning and social integration:

> When you enter university, you hardly know anybody. So, peer tutoring for Logic was great; I had just arrived and in no time at all everybody was helping each other. There were higher-year students who were engaged in facilitating the sessions and helping us with difficulties. We were searching for the correct answers together. This provided us with an opportunity to experience positive interactions with classmates and to build up relationships with more sustainable potential friendships. (Student 169: woman, first-time, regular student, LGLC).

That peer mentoring also had an impact on students' institutional attachment was evident from the interviews. Almost invariably, peer mentoring was emphasized as crucial for the initial decision of students to study at university. The attachment to university that arose was common among many respondents:

> It makes the university unique in this way. And, also, the competitive position with respect to other universities. They do not offer a support network of senior students of the faculty or where you can get a mentor scheme. Then you know in particular that you have a safety net. This was and is very important for me. And this is also the reason why students choose this university. (Student 154: woman, first-time, regular student, SGLC).

Respondents who participated in peer mentoring indicated that they now felt more attached to the university and were more motivated to remain enrolled. The role of peer mentoring in promoting counselling and finding informal expertise became clear:

The antisocial atmosphere and my roommates, they were the reason why I felt depressed. This period was really hard for me. Imagine, you come home from the lessons, but they do not interest you anymore. You cannot motivate yourself to study, and you are all alone, miserable. If my mentor was not with me, I think that would be the reason why I no longer lived in the dorms and why I left university. (Student 177: woman, first-time, regular student, LGLC, SA).

14.5 Discussion

14.5.1 Summary

Results indicated that peer mentoring (as compared to peer tutoring) was the most effective and efficient means to enhance social integration. Participants particularly mentioned that activities such as speed dating and mentoring days were important, since they brought them into contact with classmates and senior students and provided more opportunities for further social interaction. Participants mentioned the potential of such methods, since through this they could build up self-esteem, which stimulated students to ask questions of higher-year students and participate in other cross-age peer mentoring programmes. Although peer tutoring was not as effective in social integration, it was significantly important in relation to academic integration. However, participants emphasised the importance of class-based peer mentoring for social and academic integration, because classmates would experience the same study trajectory for the following three years. The availability of support over the longer term was seen as important. Using out-of-campus locations, appreciation-based narratives, and regular class-based social events were identified as examples of best practice.

14.5.2 Limitations

The use of a single cohort of psychology and educational students from one university inevitably raises limits on the transferability of the findings to other institutions and student groups. Nevertheless, triangulating data through questionnaires and interviews has provided rich descriptions and will raise the validity and credibility of the findings (Cresswell & Miller, 2000). Another important limitation is the absence of randomly selected controls for dealing with variables. The fact that we did not check for the multilevel effects of students being nested within the class groups is another limitation. A further limitation is that we did not enter covariates such as gender, age, socio-economic status or migration background to assess effects of variables other than contextual ones. Nor did we check the effects of implementation fidelity of the intervention as this was not the aim of the study. The dependent variables were only capturing self-reported data, sometimes recalled from previous integration experiences. Not all students can remember

events accurately or completely. Finally, the degree of differences between participants and non-participants could have originated from inter-individual differences at start. These differences might be the result of selection bias.

14.6   Relationship to Previous Research

Firstly, concerning peer mentoring, it was further evident from students' comments that it was not primarily higher-year students who were responsible for creating the benefits for social integration, but it was particularly activities such as speed dating and mentoring days that helped bring them into contact with classmates and eased them into social integration. This is partly in line with Daloz and Holt's (1988) suggestion that peer mentoring organisations need to set up social events for those participating in the programme, as these events provide opportunities for increased social interactions between mentors and mentees. The findings of this study showed that that such events provide more opportunities for social interactions between mentors and mentees. Secondly, it was further evident that students who reported the most beneficial experiences with peer mentoring were mainly those who belonged to student organisations or lived nearby or close to their mentor (either on campus or at home). Since the development of social relationships is correlated with regular and frequent meetings between mentor and mentee, this finding is not surprising (Colvin, 2007; Cornelius, et al., 2016). Indeed, it reveals some of the factors related to the problems inherent in building Wenger's "communities of practice" (Farnsworth, et al., 2016).

Secondly, results indicated that students' social integration between those who participated in peer tutoring and those who did not were not significantly different. Some respondents stated it was relatively hard for first-year students to make connections and work collaboratively together when social connections were not promoted in initial phases. Students needed to make connections assertively and to try to find someone at a similar stage of progress and achievement level in order to get the most help out of these contacts and to experience collaborative learning positively. In this respect, firstly, it is suggested that it should help for facilitators to encourage students to spend a few moments socialising with each other before each session begins. Future research needs to clarify whether or not this makes any difference to social integration outcomes.

Thirdly, it is argued that the development of social relations can be fostered by making connections and making students' needs or abilities apparent to peers: these needs and abilities being topics with which students need help, or for subjects where students want to provide help, for example. This finding fits in with recent research that peer tutoring activities must incorporate some means of ensuring that tutees and tutors are well matched (Evans & Cosnefroy, 2013). This closely relates to what Ito et al. (2013) recently described as connected learning, which aims to support interest-driven activities, whereby learning is driven through social interactions with other like-minded people. As such, peer tutoring is based on connected learning principles, because students can exchange experiences and make

friends—a promising approach to promoting both social and academic integration and learning in the first year of university (Rayle & Chung, 2008).

Our study confirms previous findings from Sosik and Godshalk (2000), which also suggested that age, gender, ethnicity, language preferences and education need to be taken into consideration. It also confirms findings from Bozeman and Feeney (2007), further suggesting that having similar backgrounds, interest and life experiences should be taken into consideration when pairing mentors and mentees.

14.7 Conclusion

This study extends prior research by exploring the potential influence of peer mentoring and peer tutoring on social integration, academic integration and institutional attachment with first year students. Using a mixed methods approach involving both quantitative and qualitative methods, the study compared the impact of both peer tutoring and peer mentoring approaches. Results indicated that friendship resulting from the accelerating integration was created in both groups of peer mentoring and peer tutoring participants. Both experienced informal learning in contrast to other non-participating students who did not create such friendships. However, peer mentoring seemed more powerful in terms of effects on social integration and peer tutoring was more powerful regarding academic integration. Another important conclusion of our study is that as spontaneously indicated by the students, both peer mentoring and peer tutoring increase self-esteem. There are thus evidence-based action implications for educational practice, policy-making and future researchers. It will be important in planning future strategies to enhance social and academic integration and institutional attachment that student opinions are firmly taken into account.

References

Aljohani, O. (2016). A comprehensive review of the major studies and theoretical models of student retention in higher education. *Higher Education Studies, 6*(2), 1–18.

Amaral, K. E., & Vala, M. (2009). What teaching teaches: Mentoring and the performance gains of mentors. *Journal of Chemical Education, 86*(5), 630–633.

Andrews, J., & Clark, R. (2009). *Peer mentoring in higher education: A literature review.* (CLIPP working paper series; Vol. 109). Aston University.

Astin, A. W. (1993). *What matters in college? Four critical years revisited*. Jossey-Bass.

Baker, R. W., & Siryk, B. (1989). *Student adaptation to college questionnaire (SACQ).* Western Psychological Services.

Barrett, F. J. (1995). Creating appreciative learning cultures. *Organizational Dynamics, 24*(2), 36–49.

Benishek, L. A., Feldman, J. M., Shipon, R. W., Mecham, S. D., & Lopez, F. G. (2005). Development and evaluation of the revised academic hardiness scale. *Journal of Career Assessment, 13*(1), 59–76.

Berger, J. B., Ramírez, G., & Lyon, S. (2012). Past to present: A historical look at retention. In A. Seidman (Ed.), *College student retention: Formula for student success* (pp. 7–34). Rowman & Littlefield.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*, 347–364. https://doi.org/10.1007/BF00138871

Bittich S., & Rongen L. (2007). Factoren voor studiesucces, rapport deskresearch en QuickScan. In *Model-beleidsplan ter bevordering van sociale integratie in hogeschoolopleidingen*. Hogeschool Zeeland.

Bozeman, B., & Feeney, M. K. (2007). Toward a useful theory of mentoring: A conceptual analysis and critique. *Administration & Society, 39*(6), 719–739.

Bronstein, S. B. (2008). Supplemental instruction: Supporting persistence in barrier courses. *Learning Assistance Review, 13*(1), 31–45.

Bullen, P., Farruggia, S. P., Gómez, C. R., Hebaishi, G. H. K., & Mahmood, M. (2010). Meeting the graduating teacher standards: The added benefits for undergraduate university students who mentor youth. *Educational Horizons, 89*(1), 47–61.

Burgess, M. (2016). *The 'whole of the wall': A micro-analytic study of informal, computer-mediated interaction between children from a marginalised community*. Newcastle University.

Carr, S. E., Brand, G., Wei, L., Wright, H., Nicol, P., Metcalfe, H., & Foley, L. (2016). Helping someone with a skill sharpens it in your own mind: A mixed method study exploring health professions students' experiences of Peer Assisted Learning (PAL). *BMC Medical Education, 16*(1), 1–48.

Carter, D. F., Locks, A. M., & Winkle-Wagner, R. (2013). From when and where I enter: Theoretical and empirical considerations of minority students' transition to college. In M. B. Paulsen (Ed.), *Higher education: Handbook of theory and research* (Vol. 28). Springer.

Colvin, J. W. (2007). Peer tutoring and social dynamics in higher education. *Mentoring & Tutoring, 15*(2), 165–181.

Cooperrider, D. L., & Srivastva, S. (1987). Appreciative inquiry in organizational life. *Research in Organizational Change and Development, 1*(1), 129–169.

Cooperrider, D. L., Whitney, D. K., & Stavros, J. M. (2003). *Appreciative Inquiry handbook*. Berrett-Koehler Publishers.

Copeland, S. R., McCall, J., Williams, C. R., Guth, C., et al. (2002). High school peer buddies. *Teaching Exceptional Children, 35*(1), 16.

Cornelius, V., Wood, L., & Lai, J. (2016). Implementation and evaluation of a formal academic-peer-mentoring programme in higher education. *Active Learning in Higher Education, 10*(1), 7–25.

Corsaro, W. A. (2005). *The sociology of childhood*. Pine Forge Press.

Court, S., & Molesworth, M. (2008). Course-specific learning in peer assisted learning schemes: A case study of creative media production courses. *Research in Post-Compulsory Education, 13*(1), 123–134.

Creswell, J. W., & Clark, V. L. P. (2011). *Designing and conducting mixed methods research*. SAGE publications.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice, 39*(3), 1–130.

Crosier, D., Purser, L., & Smidt, H. (2007). *Trends V: Universities shaping the European Higher Education Area*. European University Association.

Czarniawska-Joerges, B. (1996). Book review: Realities and relationships. Soundings in social construction by K. J. Gergen. *Scandinavian Journal of Management*, (12)4, 468–470.

Daloz, L. A., & Holt, M. E. (1988). Effective teaching and mentoring. *The Journal of Continuing Higher Education, 36*(1), 28–29.

Dobbie, M., & Joyce, S. (2008). Peer-assisted learning in accounting: A qualitative assessment. *Asian Social Science, 4*(3), 18–25.

Dukakis, K., Bellm, D., Seer, N., & Lee, Y. (2007). *Chutes or ladders. Creating support services to help early childhood students succeed in Higher Education*. Center for the Study of Child Care Employment, University of California at Berkeley.

Duron, R., Limbach, B., & Waugh, W. (2006). Critical thinking framework for any discipline. *International Journal of Teaching and Learning in Higher Education, 17*(2), 160–166.

Elster, D. (2014). First-year students' priorities and choices in STEM studies—IRIS findings from Germany and Austria. *Science Education International, 25*(1), 52–59.

Evans, L., & Cosnefroy, L. (2013). The dawn of a new professionalism in the French academy? Academics facing the challenges of change. *Studies in Higher Education, 38*(8), 1201–1221.

Falchikov, N. (2001). *Learning together: Peer tutoring in higher education.* Routledge Kegan Paul.

Farnsworth, V., Kleanthous, I., & Wenger-Trayner, E. (2016). Communities of practice as a social theory of learning: A conversation with Etienne Wenger. *British Journal of Educational Studies, 64*(2), 139–160. https://doi.org/10.1080/00071005.2015.1133799

Fayowski, V., & MacMillan, P. D. (2008). An evaluation of the supplemental instruction programme in a first-year calculus course. *International Journal of Mathematical Education in Science and Technology, 39*(7), 843–855.

Ford, N., Thackeray, C., Barnes, P., & Hendrickx, K. (2015). Peer learning leaders: Developing employability through facilitating the learning of other students. *Journal of Learning Development in Higher Education, 8*, 1–23.

Fox, A., Stevenson, L., Connelly, P., Duff, A., & Dunlop, A. (2010). Peer-mentoring undergraduate accounting students: The influence on approaches to learning and academic performance. *Active Learning in Higher Education, 11*(2), 145–156.

Gillies, D., & Mifsud, D. (2016). Policy in transition: The emergence of tackling early school leaving (ESL) as EU policy priority. *Journal of Education Policy, 31*(6), 819–832.

Ginsburg-Block, M. D., Rohrbeck, C. A., & Fantuzzo, J. W. (2006). A meta-analytic review of social, self-concept, and behavioral outcomes of peer-assisted learning. *Journal of Educational Psychology, 98*(4), 732–749.

Goff, L. (2011). Evaluating the outcomes of a peer-mentoring program for students transitioning to postsecondary education. *The Canadian Journal for the Scholarship of Teaching and Learning, 2*(2), 1–13.

Griffin, M. M., & Griffin, B. W. (1998). An investigation of the effects of reciprocal peer tutoring on achievement, self-efficacy, and test anxiety. *Contemporary Educational Psychology, 23*(3), 298–311.

Hagedorn, L. S. (2006). *How to define retention: New look at an old problem. Transfer and retention of urban community college students.* TRUCCS Research Center.

Hall, R., & Jaugietis, Z. (2010). Developing peer mentoring through evaluation. *Innovative Higher Education, 36*(1), 41–52.

Hycner, R. H. (1985). Some guidelines for the phenomenological analysis of interview data. *Human Studies, 8*(3), 279–303.

Hycner, R. H. (1999). Some guidelines for the phenomenological analysis of interview data. In A. Bryman & R. G. Burgess (Eds.), *Qualitative research* (Vol. 3, pp. 143–164). Sage.

Ito, M., Gutiérrez, K., Livingstone, S., Penuel, B., Rhodes, J., Salen, K., Schor, J., Sefton-Green, J., & Watkins, S. C. (2013). *Connected learning: An agenda for research and design.* Digital Media and Learning Research Hub.

James, R., Krause, & Jennings. (2010). *The first year experience in Australian universities: Findings from a decade of national studies.* Centre for the Study of Higher Education.

Johansson, F. (2004). *The Medici effect: Breakthrough insights at the intersection of ideas, concepts, and cultures.* Harvard Business School Press.

Kung, S., Giles, D., & Hagan, B. (2014). Applying an appreciative Inquiry process to a course evaluation in higher education. *International Journal of Teaching and Learning in Higher Education, 25*(1), 29–37.

Lahman, M. P. (1999). *To what extent does a peer mentoring program aid in student retention?* Paper presented at the annual meeting of the National Communication Association Conference (85th, Chicago, Illinois, November 4–7).

Lee, J. M., Germain, L. J., Lawrence, E. C., & Marshall, J. H. (2010). It opened my mind, my eyes. It was good. Supporting college students' navigation of difference in a youth mentoring program. *Educational Horizons, 89*(1), 33–46.

Lowe, H., & Cook, A. (2003). Mind the gap: Are students prepared for higher education? *Journal of Further and Higher Education, 27*(1), 53–76.

Maheady, L., & Gard, J. (2010). Classwide peer tutoring: Practice, theory, research, and personal narrative. *Intervention in School and Clinic, 46*(2), 71–78.

Maheady, L., Mallette, B., & Harper, G. F. (2006). Four classwide peer tutoring models: Similarities, differences, and implications for research and practice. *Reading & Writing Quarterly, 22*(1), 65–89.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2018). *Qualitative data analysis: A methods sourcebook*. Sage.

Muldoon, R., & Wijeyewardene, I. (2012). Two approaches to mentoring students into academic practice at university. *Journal of the Australian and New Zealand Student Services Association, 39*, 21–31.

Ning, H. K., & Downing, K. (2010). The impact of supplemental instruction on learning competence and academic performance. *Studies in Higher Education, 35*(8), 921–939.

Outhred, T., & Chester, A. (2010). The experience of class tutors in a peer tutoring programme: A novel theoretical framework. *Journal of Peer Learning, 3*(1), 12–23.

Pascarella, E. T., & Terenzini, P. T. (2005). How college affects students (Vol. 2): A third decade of research (review). *Journal of College Student Development, 47*(5), 589–592. https://doi.org/10.1353/csd.2006.0055.

Peterfreund, A. R., Rath, K. A., Xenos, S. P., & Bayliss, F. (2007). The impact of supplemental instruction on students in STEM courses: Results from San Francisco State University. *Journal of College Student Retention: Research, Theory and Practice, 9*(4), 487–503.

Piaget., J. (1987). *Possibility and necessity*. University of Minnesota Press.

Pleschová, G., & McAlpine, L. (2015). Enhancing university teaching and learning through mentoring: A systematic review of the literature. *International Journal of Mentoring and Coaching in Education, 4*(2), 107–125.

Rayle, A. D., & Chung, K.-Y. (2008). Revisiting first-year college students' mattering: Social support, academic stress, and the mattering experience. *Journal of College Student Retention: Research, Theory & Practice, 9*(1), 21–37.

Rienties, B., Beausaert, S., Grohnert, T., Niemantsverdriet, S., & Kommers, P. (2012). Understanding academic performance of international students: The role of ethnicity, academic and social integration. *Higher Education, 63*(6), 685–700.

Robinson, D. R., Schofield, J. W., & Steers-Wentzell, K. L. (2005). Peer and cross-age tutoring in math: Outcomes and their design implications. *Educational Psychology Review, 17*(4), 327–362.

Roscoe, R. D., & Chi, M. T. H. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research, 77*(4), 534–574.

Severiens, S., & Wolff, R. (2008). A comparison of ethnic minority and majority students: Social and academic integration, and quality of learning. *Studies in Higher Education, 33*(3), 253–266.

Smith, D. I. (2009). Changes in transitions: The role of mobility, class and gender. *Journal of Education and Work, 22*(5), 369–390.

Smith, J., May, S., & Burke, L. (2007). Peer assisted learning: A case study into the value to student mentors and mentees. *Practice and Evidence of the Scholarship of Teaching and Learning in Higher Education, 2*(2), 80–109.

Solinger, O. N., Hofmans, J., Bal, P. M., & Jansen, P. G. W. (2015). Bouncing back from psychological contract breach: How commitment recovers over time. *Journal of Organizational Behaviour, 37*(4), 494–514.

Sosik, J. J., & Godshalk, V. M. (2000). Leadership styles, mentoring functions received, and job-related stress: A conceptual model and preliminary study. *Journal of Organizational Behavior, 21*(4), 365–390.

Stigmar, M. (2016). Peer-to-peer teaching in higher education: A critical literature review. *Mentoring & Tutoring: Partnership in Learning, 24*(2), 124–136.

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). University of Chicago Press.

Tinto, V. (2003). Promoting student retention through classroom practice. In *Enhancing student retention: Using international policy and practice.* International Conference Sponsored by the European Access Network and the Institute for Access Studies. Staffordshire University.

Tinto, V. (2006). Research and practice of student retention: What next? *Journal of College Student Retention: Research, Theory & Practice, 8*(1), 1–19.

Tinto, V. (2010). From theory to action: Exploring the institutional conditions for student retention. In J. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 25). Springer.

Tinto, V. (2012). Enhancing student success: Taking the classroom success seriously. *The International Journal of the First Year in Higher Education, 3*(1), 1–8.

Tinto, V. (2015). Through the eyes of students. *Journal of College Student Retention: Research, Theory & Practice, 19*(3), 254–269.

Tinto, V., & Pusser, B. (2006). *Moving from theory to action: Building a model of institutional action for student success.* National Postsecondary Education Cooperative. Retrieved from: https://nces.ed.gov/npec/pdf/tinto_pusser_report.

Topping, K. (2005). Trends in peer learning. *Educational Psychology, 25*(6), 631–645.

Topping, K. (2015). Peer tutoring: Old method, new developments/Tutoría entre iguales: Método antiguo, nuevos avances. *Infancia y Aprendizaje, 38*(1), 1–29.

Topping, K., & Ehly, S. (1998). *Peer-assisted learning.* Routledge/Taylor & Francis.

Topping, K. J., Duran, D., & Van Keer, H. (2016). *Using peer tutoring to improve reading skills.* Routledge.

Topping, K., Hill, S., McKaig, A., Rogers, C., Rushi, N., & Young, D. (1997). Paired reciprocal peer tutoring in undergraduate economics. *Innovations in Education and Training International, 34*(2), 96–113.

Torenbeek, M., Jansen, E., & Hofman, A. (2010). The effect of the fit between secondary and university education on first-year student achievement. *Studies in Higher Education, 35*(6), 659–675.

Twomey, J. L. (1991). *Academic performance and retention in a peer mentor program at a two-year campus of a four-year institution.* New Mexico State University.

Van der Meer, J., & Scott, C. (2009). Students' experiences and perceptions of peer assisted study sessions: Towards ongoing improvement. *Australasian Journal of Peer Learning, 2*(1), 3–22.

Van Manen, M. (1990). *Researching lived experience: Human science for an action sensitive pedagogy.* State University of New York Press.

Vossensteyn, J. L., Kottmann, A., & Jongbloed, B. W. A. (2015). *Dropout and completion in higher education in Europe.* Publications Office of the European Union.

Watkins, J. M., & Mohr, B. J. (2001). *Appreciative inquiry: Change at the speed of imagination.* Jossey-Bass/Pfeiffer.

Whitney, D. D., & Trosten-Bloom, A. (2010). *The power of appreciative inquiry: A practical guide to positive change.* Berrett-Koehler Publishers.

Wilcox, P., Winn, S., & Fyvie-Gauld, M. (2005). It was nothing to do with the university, it was just the people: The role of social support in the first-year experience of higher education. *Studies in Higher Education, 30*(6), 707–722.

Wu, P. (2013). *Bridging the transition process for first-year students in distance construction programs: A case study in Australia.* FYHE Conference Committee.

Zepke, N., Leach, L., & Prebble, T. (2006). Being learner centred: One way to improve student retention? *Studies in Higher Education, 31*(5), 587–600.

How to Make Students Feel Safe and Confident? Designing an Online Training Targeting the Social Nature of Peer Feedback

Morgane Senden, Dominique De Jaeger, Tijs Rotsaert, Fréderic Leroy, and Liesje Coertjens

15.1 Peer Feedback and Its Social Nature

15.1.1 The Benefits of Peer Feedback

How to foster student learning is a crucial question in educational research. Feedback—defined as "a process through which learners make sense of information from various sources and use it to enhance their work or learning strategies" (Carless & Boud, 2018, p. 1)—has been proposed as an important tool for student learning and numerous studies have confirmed this (e.g. Black & Wiliam, 1998; Hattie & Timperley, 2007). Indeed, in a recent meta-analysis of 435 studies, Wisniewski and his colleagues (2020) found that feedback has a moderate size effect (d = 0.48) on students' learning.

A particularly effective kind of feedback is feedback from peers. Indeed, empirical evidence supports the value of peer feedback and suggests it can even be a more useful tool for learning than teacher feedback. Wisniewski et al. (2020) found

M. Senden (✉) · L. Coertjens
Psychological Sciences Research Institute (IPSY), Université Catholique de Louvain, Louvain-La-Neuve, Belgium
e-mail: morgane.senden@uclouvain.be

D. De Jaeger
Institute of Neuroscience (IoNS), Faculty of Motor Sciences, Université Catholique de Louvain, Louvain-La-Neuve, Belgium

T. Rotsaert
Department of Educational Studies, Ghent University, Ghent, Belgium

F. Leroy
Faculty of Motor Sciences, Université Catholique de Louvain, Louvain-La-Neuve, Belgium

student-to-student feedback to be more efficient than teacher-to-student feedback. In another recent meta-analysis, Double and his colleagues (2020) addressed the effect of peer assessment (which sometimes, but not always include peer feedback). Based on 54 (quasi-) experimental studies, this meta-analysis concludes that peer assessment impacts student learning more positively than teacher assessment.

The value of peer feedback does not only lies in the fact it helps students to developed specific subject-related learning, it also pushes them to develop more general feedback skills (Carless et al., 2011). By giving opportunities for students to practice making judgements, peer feedback contributes, for example, to the development of evaluative judgment, which is defined as "the capacity to make decisions about the quality of work of self and others (Tai et al., 2018, p. 471)". Evaluative judgement is a necessary skill for students to become independent lifelong learners, which should be a goal of higher education (Tai et al., 2018).

15.1.2 Students' Concerns and How to Take Them into Account

Given that peer feedback can be a very beneficial activity for students' learning, we could expect students to be eager to participate in peer assessment activities. However, this is only partly the case. Indeed, the majority of students report to like peer assessment and to find it useful (Hanrahan & Isaacs, 2001; Mulder et al., 2014). However, they also express a series of concerns. These concerns are various but have a common element: they emerge from the fact that peer assessment is a social experience (e.g. Hanrahan & Isaacs, 2001; Mulder et al., 2014; Wilson et al., 2015). Some students fear, for example, that their peers will be biased or will not put enough effort into their assessment, feel they lack the skills to evaluate their peers, find it difficult to be objective and feel uncomfortable evaluating their peers and being evaluated by them (Hanrahan & Isaacs, 2001; Mulder et al., 2014; Wilson et al., 2015). These concerns are not anecdotic: in Stanier (1997)'s study, 40% of the students found peer assessment to be an uncomfortable experience and in Mostert and Snowball (2013)'s study, 29% felt that their peers did not engage enough in the activity and 19% did not trust their peers as assessors.

The main suggestion in literature to overcome students' concerns linked to the social nature of peer feedback is anonymity. In Yu and Liu (2009)'s study, for example, students preferred using a surname rather than their real name in a peer assessment activity and, in Vanderhoven et al. (2015)'s study, students experienced less peer pressure and fear of disapproval in an anonymous peer assessment activity compared to a non-anonymous one. Rotsaert et al. (2018) have found that, when peer feedback is used multiple times in a course, fading anonymity can be used as an instructional scaffold. When students first had the opportunity to experience peer feedback anonymously, they continue to provide feedback of the same quality and to feel safe when anonymity is removed and the importance they place on anonymity decreases.

However, expecting anonymity to relieve every tension created by peer assessment is unrealistic (Panadero & Alqassab, 2019). Panadero and Alqassab (2019)'s

literature review on the effects of anonymity in peer feedback shows mixed results, with only a slight positive tendency towards anonymity. Their main conclusions are that more research on the effects of anonymity is needed and that the instructional context and goals need to be considered (Panadero & Alqassab, 2019).

Moreover, anonymity is not always possible in peer feedback (e.g. feedback on an oral presentation), and, even when possible, it is not always desirable. Indeed, anonymity necessarily means that students can not interact with one another and discuss the feedback, which removed the richest part of feedback if we see it as a dialogical process (Ajjawi & Boud, 2017). Additionally, not all potential undesirable effects of the social nature of peer feedback can be cancelled out by the use of anonymity (e.g. the fear of not being able to provide valuable feedback). Therefore, it is necessary to find other ways to create an environment in which students feel safe and comfortable to participate in peer feedback activities.

To ease tensions linked to the social aspects of peer feedback, students could be trained on these aspects. In Li (2017)'s study, for example, students were assigned to three groups: identity group (the identity of assessors and assesses was revealed to each other), anonymity group (both assessors and assesses were anonymous), and the training group (the identities were known, but students followed a training, aimed at controlling the possible negative impact of having their identities revealed). Results indicated that both the training and the anonymity groups showed a larger improvement in their performance than the identity group. Moreover, regarding their perception, students in the training group valued peer assessment activities more and experienced less pressure from peers than the students in the two other groups. Thus, it seems that when anonymity is not feasible, training could counteract the negative impact of having students' identities revealed.

15.1.3 Trust and Psychological Safety

Li (2017)'s study suggests that the provision of peer feedback training could be useful, but in her study, she focused on peer pressure, which is not the only important variable linked to the social nature of peer feedback. Based on the literature on collaborative learning and group work, van Gennip and her colleagues (2009; 2010) identified four variables that could be of importance in a peer assessment activity, namely:

1. Psychological safety: the belief shared by members of a group that they can take interpersonal risks in this group
2. Value congruency: the similarities of team members' opinions about what the tasks, missions and goals of their team should be
3. Interdependence: the fact that everyone needs to participate actively in the assessment task because, if some students do not provide feedback it will have an impact on the students whose work they assessed

4. Trust: (1) the confidence that a student has in his/her own ability to assess their peers' work—i.e., trust in oneself—and (2) the trust in their peers' capacity to assess his/her work—i.e., trust in peer

Of these four variables, trust and psychological safety appear to be key to consider when implementing a peer feedback activity. Indeed, van Gennip and colleagues (2012) found that, in the context of secondary-vocational education, a high level of psychological safety and trust (in the self and the peer) had a positive impact on perceived learning. For value congruency and interdependence, the relationship with perceived learning was less clear. Panadero (2016) confirms this: the relevance of trust and psychological safety is more evident than that of value congruency and interdependence, as these two latter variables are relevant in contexts where students have shared goals, which is not necessarily the case in the context of peer feedback (e.g. with online anonymous peer feedback). Consequently, for the training, we focused upon trust and psychological safety.

Regarding the term 'trust", it is important to highlight its two facets: trust in oneself and trust in peers (van Gennip et al., 2010). A student can trust his or her ability but not the ability of his or her peers, or vice versa. In a study by Cheng and Tsai (2012), for example, the majority of students (74%) trusted their abilities to assess their peers. A smaller percentage of students (57%) also trusted their peers' ability.

The notion of psychological safety originally came from organizational psychology where it can be defined as the "perceptions of the consequences of taking interpersonal risks in a particular context" (Edmondson & Lei, 2014, p. 24). Linked to its origin, the majority of research on psychological safety has been conducted in the working environment (e.g. Edmondson et al., 2007); however, even in the working environment, an emphasis was placed on learning behavior. When people feel psychologically safe, they are less afraid to take interpersonal risks, which means that they are more willing to express themselves without worrying about possible negative reactions from other members of their team. Therefore, in a psychologically safe environment, team members are more willing to carry out learning behavior like seeking feedback, asking for help or talking about errors (Edmondson, 1999). De Stobbeleir and colleagues (2019) confirmed this: when employees perceived their environment as psychologically safe, they seek more feedback from their peers.

In an educational context, Soares and Lopes (2020) have shown that psychological safety has a positive influence on academic performance. Psychological safety creates an environment where students feel comfortable discussing their performance and errors and asking for feedback, which has a positive impact on their learning. Hence, psychological safety is a requirement for peer feedback. In the context of peer assessment, psychological safety is defined by Panadero (2016, p. 251) as "the extent to which students feel safe to give sincere feedback as an assessor and do not fear inappropriate negative feedback as an assessee".

15.1.4 Online Training

Online courses, such as MOOCs, have gained popularity over the last decade (Shah, 2020). For such MOOCs, it is challenging to exceed knowledge transfer and provide a real educational experience for students (Suen, 2014). To achieve this and keep the workload for teaching staff feasible, peer feedback is frequently used (Suen, 2014): it allows students to obtain feedback to enhance their learning.

Yet, students' concerns linked to the social nature of peer feedback may be amplified in MOOC settings. As Suen (2014) explains, peer feedback in MOOCs takes place in a context where there is, at best, few instructor mediation, supervision or guidance and where students have little incentive to take peer feedback activities seriously. In this context, students are often dissatisfied with the use of peer feedback and complain that their peers give them superficial or inconsistent feedback (Hew, 2018). Therefore, taking students' concerns into account and training them before a peer feedback activity could be even more important in MOOCs than in traditional on-campus courses.

15.1.5 The Present Study

Although some leads exist on how to take into account the interpersonal context when designing peer feedback activities, only a few studies have been conducted (e.g. Rotsaert et al., 2018; Vanderhoven et al., 2015). It remains veiled how trust and psychological safety can be stimulated in an online setting. To fill this gap, we set out to design an online training targeting these two aspects, to optimize students' learning from peer feedback.

The purpose of this chapter is to present the training and the rationale behind its different components. In the section "Training design", we will explain how the literature was explored to find effective tools for training. Subsequently, in the section "Training procedure", a detailed outline of the different stages of the training will be presented. Moreover, even though an evaluation of the training is beyond the scope of this article, we will give some elements on how the training was received by students in the section "Students' perceptions of the training session". Finally, we will discuss some limitations and perspectives.

It is important to specify that the training is conceived for peer feedback activities where the feedback are performance feedback, i.e. feedback on students' performance, and not process feedback, i.e. feedback on the way students performed a task (Gabelica et al., 2012). Indeed the training is based on research done on performance feedback and its specific challenges, whose results cannot necessarily be generalized to process feedback (Gabelica et al., 2012).

15.2   Training Design

15.2.1 Training Objectives

Our purpose is to design a training to tackle students' feelings of distrust and psychological unsafety before participating in a peer feedback activity. Given that it is recommended to train students on how to provide peer feedback (van Zundert et al., 2010), we also included a more general part to our training to help students provide effective feedback. Therefore, the training has six objectives:

1. to clarify for the students what the objectives and advantages of peer feedback are
2. to increase students' skills for providing effective feedback
3. to address students' concerns about peer feedback
4. to increase students' feeling of psychological safety
5. to increase students' trust in their ability to assess others
6. to increase students' trust in their peers' ability to assess them

For our two first objectives (objectives 1 and 2) the learning outcomes are knowledge and skills. The learning outcomes of the last four objectives (objectives 3–6) can be considered as attitudes, which are defined as "beliefs and opinions that support or inhibit behavior" (Oskamp, cited by Blanchard & Thacker, 2013, p. 37). Because the training mostly targets students' attitudes, it must allow the active participation of students. Therefore, the training consisted mainly of role-plays and discussions, two effective methods to transform attitudes (Blanchard & Thacker, 2013).

15.2.2 Inspirations from Existing Training

As it is common to train students before a peer feedback activity and some training in other contexts than peer feedback may have targeted trust and psychological safety, we additionally explored this literature.

It has been shown that training students before a peer assessment activity improved the reliability of peer assessment, peer assessment skills and students' attitudes towards peer assessment (van Zundert et al., 2010). Training can focus on various aspects, like how to decide what is important to assess, how to judge a performance or how to provide feedback for future learning (Sluijsmans et al., 2004). Their length may also vary, some of them being very comprehensive (e.g. Sluijsmans et al., 2002), while others are shorter but can still be effective, as shown by Alqassab and colleagues (2018). It is from the latter that we drew to design our training, which we want to keep short enough to fit it into already busy schedules: it seems unrealistic that more than one session will be dedicated to peer feedback training in a course where peer feedback is only a way to help students learn and not an objective in itself.

The training created by Alqassab et al. (2018) consisted of a general discussion on peer assessment, a lecture on Hattie and Timperley (2007)'s framework, group and individual exercises to integrate the theory and practice sessions. Alqassab and colleagues (2018) found that, for medium or high-achieving students, the training increased the proportion of self-regulation feedback, which are considered higher-level feedback and are more effective (Hattie & Timperley, 2007). The training did not affect low-achieving students.

In addition, Dusenberry and Robinson (2020) created a training session to increase students' feeling of psychological safety before working in small groups. Their training lasted 50 min and was composed of a video lecture, a short discussion and a hands-on exercise. Contrary to their hypothesis, the level of psychological safety was not higher for the students who followed the training than for the students in the control group. They identified several limits in their training, which could explain this absence of a significant effect. A first limitation is that their training was not context-specific. The same training was given to a variety of students, working on various projects, and there was no direct link between the training and the teams' projects. A second limitation is that the video lecture took up about half of the training time, which did not leave much time for more active learning methods. Consequently, if we want to make sure our training is effective, it seems important to avoid these pitfalls by making our training context-specific and by favouring active learning methods, as also recommended by Blanchard and Thacker (2013).

15.3 Training Procedure

We provided the peer feedback training to third-year university students in physical education following a seminar on acrobatic sports didactics. Forty-one students were enrolled in this mandatory seminar, but six students did not participate in either the training or the peer feedback activity (even though both were mandatory). During this seminar, students had to create an instruction sheet illustrating a gymnastic exercise, assess the instruction sheet of seven of their peers and improve their work based on the feedback they received. The training took place during the second seminar session. It was provided by a researcher (first author) in collaboration with the course's professor (fourth author). The online training was composed of five stages (see Table 15.1): discovery of student's representation, lecture on how to provide effective feedback, peer feedback practice, role-play and discussion in small groups, and summary of key learning points.

We conducted the training online, through the Microsoft-Teams platform. Like other videoconferencing applications (e.g. Zoom), Teams allows us to divide participants into break-out rooms, which was essential as most of the training time was spent in subgroups.

Doing the training online incited us to carefully plan it. As Bolinger and Stanton (2020) explained, it is possible to run synchronous online role-plays but the logistics are more difficult to manage. It is not possible, for example, to pass sheets

Table 15.1 Overview of the training session

Stage	Students' activities	Objectives	Timing
Discovery of students' representation	In the large group, students answer to a Wooclap (active)	To adapt the rest of the training To address students' concerns about peer feedback (objective 3)	10 min
Lecture on how to provide effective feedback	In the large group, students listen to the lecture (passive)	To increase students' skills for provideng effective feedback (objective 2)	10 min
Peer feedback practice	Individually and then in pairs, students practice giving feedback and discuss it (active)	To increase students' skills for provideng effective feedback (objective 2)	30 min
Role-play and discussion in small groups	In small groups, students perform role-play and discuss in small-groups (active)	To increase students' feeling of psychological safety (objective 4) To increase students' trust in their ability to assess others (objective 5) To increase students' trust in their peers' ability to assess them (objective 6)	50 min
Summary of key learning points	In the large group, students listen to and contribute to the synthesis (passive)	To clarify the objectives and advantages of peer feedback (objective 1) To address students' concerns about peer feedback (objective 3)	20 min

Note The first part (discovery of students' representations) took place a week before the training session

around or to easily identify which students have questions once they are in the break-out rooms. To address these difficulties, we tried to make the instructions as explicit as possible (e.g. what students were expected to do, how much time they had…) and we gave them a very detailed roadmap for the role-plays. We also made sure that there were two instructors present online, which made it possible to visit all sub-groups to answer questions.

As mentioned above, it was challenging to find a balance between the importance of having comprehensive training and the necessity of keeping it time-efficient, so it can be relatively easily inserted into courses. Table 15.1 presents an overview of the training, with the timing devoted to each activity of this two-hour session. As recommended by Blanchard and Thacker (2013), we started by targeting the knowledge and skills before moving on to attitudes.

Fig. 15.1 Word cloud generated by students' answers regarding the disadvantages of peer feedback

15.3.1 Stage 1. Discovery of Students' Representations

A week before the training session, we asked students to answer some questions through an interactive platform (Wooclap). We asked them to write down in one word what peer feedback means to them, to write down the advantages and disadvantages of peer feedback and to tell us if they had any concerns or fear linked to the use of peer feedback. For the three first questions, students saw the responses of other students appear live and could *like* them. As an example, Fig. 15.1 is the word cloud generated by students' answers regarding the disadvantages of peer feedback.

Discovering students' representations allowed us to tailor the training to this specific group of students, which could enhance training efficiency (Dusenberry & Robinson, 2020). This group of students were concerned that their peers are not qualified enough to assess them, that their peers are not objective enough to assess them (more precisely they fear that they will be "too nice") and that they themselves are not qualified enough to assess their peers. These results confirmed the relevance of providing a training targeting the notion of trust. Some aspects linked to psychological safety also emerged, although to a lesser extent (e.g. the fear of being judged as stupid).

15.3.2 Stage 2. Lecture on How to Provide Effective Feedback

At the start of the two-hour training, we explained the six objectives (see Table 15.1) and linked them to students' concerns based on the Wooclap responses. Then we gave a short lecture on how to provide effective feedback based on the framework of Hattie and Timperley (2007). We explained to students that the purpose of feedback is to reduce the gap between actual and desired performance and it

should therefore contain an answer to the three following questions: Where am I going? How am I going? And Where to next? (Hattie & Timperley, 2007). We also described the four feedback levels (self, task, process and auto-regulation), explained why the former two were less effective and gave examples of feedback at each level. The lecture format allowed us to convey essential knowledge to students, but we kept it under 20 min as to not lose students' attention (Blanchard & Thacker, 2013).

15.3.3 Stage 3. Peer Feedback Practice

For stage 3, students had the opportunity to practice giving feedback and to familiarize themselves with the rubric they will use afterwards for the real peer feedback activity. The day before the training session, students had to hand in an assignment (similar but not identical to the main assignment). During the training, we paired them in breakout rooms and randomly assigned them two assignments. Students had to individually assess the assignments with a rubric and then, in pairs, compare their assessments and discuss possible disagreements. After returning to the large group, time was set aside for them to ask questions. We also asked them to give examples of feedback they would provide and think together about how to make them as effective as possible.

15.3.4 Stage 4. Role-Play and Discussion in Small Groups

As explained by De Ketele et al. (2007), role-play is a training method in which participants interpret the role of different characters in a specific situation, to allow an analysis of the representations, feelings and attitudes related to this situation. What distinguishes role-play from other simulations is its emphasis on interpersonal interactions (Bolinger & Stanton, 2020), which makes it particularly relevant for training on trust (in others) and psychological safety.

To the best of our knowledge, there are no existing role-plays on trust and psychological safety in peer feedback described in the literature. Consequently, we designed them ourselves, following general guidelines provided by Bolinger and Stanton (2020). The role-plays' scenarios were conceived to bring students to project themselves into peer feedback situations and to identify problems that may emerge in these situations. Based on the Wooclap responses (see stage 1), we selected the two most appropriate role-plays among several that we had created. The first role-play consisted of three friends who participated in a peer feedback activity but were all dissatisfied with the received written comments for different reasons (e.g. the feedback were only positive, without any suggestion for improvement). In the second role-play, students had to put themselves in the shoes of three students who had to decide what grade and feedback to give to peers who did a poor oral presentation.

Students were randomly divided into small groups, with each group performing the same role-play simultaneously (i.e. multiple role-plays, Blanchard & Thacker, 2013). This format allowed us to involve all students and to let various elements emerge for the discussion afterwards (given that the scenario will be played slightly different in every group) while taking much less time than if each group had played one after the other (Blanchard & Thacker, 2013).

Once split into groups of six, students received a detailed roadmap with two role-play scenarios and instructions on how to play and discuss them. For each role-play three students acted out the roles while the three others observed the role-play and took notes to inform the following discussion. Having two role-plays allowed each student to play one role, either in the first or second role-play.

After performing and discussing the two role-plays, they stayed in sub-groups to synthesize their discussions. More precisely, we asked them to identify the benefits and interpersonal risks of peer feedback, and what the professor and themselves as students can do to ensure that a peer feedback activity works well.

Participating in role-play simulations can bring discomfort to some students, especially if they are not used to role-playing in class (Bolinger & Stanton, 2020). Therefore, we tried to make the situation as comfortable as possible. Playing in small groups instead of in front of everyone should help students feel at ease. Moreover, by having two role-plays, students more reluctant to participate could observe the first one before actively participating in the second one. And finally, students could choose which role they wanted to play (some roles being more demanding than others).

15.3.5 Stage 5. Summary of Key Learning Points

The last training stage was an open discussion to synthesize all the sub-groups' ideas in the large group. This method is used to generate participation, find out what participants think or have learned and stimulate recall of relevant knowledge (Blanchard & Thacker, 2013). We asked a student from each group to report the key points of their discussions and took live notes on our slideshow so students could see how the discussion progressed. We used this moment to explain and justify the choices made by the professor for the organization of the peer feedback activity and to link these choices to the elements discussed by students and with the literature on peer feedback. For example, when a student said that they would feel more confident if they were assessed by more than one peer, we explained that this feeling was coherent with the literature (e.g. Sung et al., 2010) according to which when the number of assessors is large enough, peer feedback is as reliable as teacher feedback and that is why they had to provide feedback to seven of their peers for this course (unlike what they did during the training). We also used this moment to address any remaining concerns.

Based on the discussion, we made a mind map (see Fig. 15.2) that we sent to students a few days after the training session. This mind map allows students to keep a record of the key ideas identified together during the training under a visual
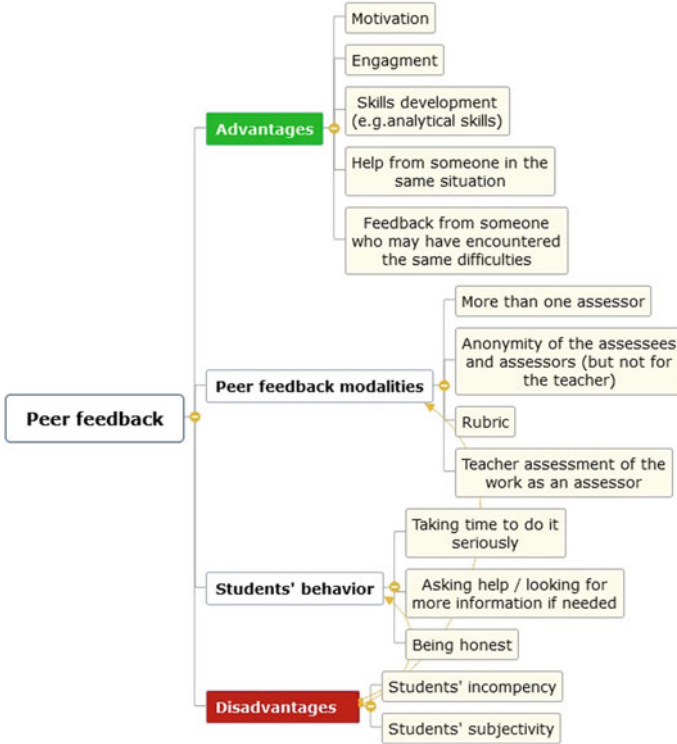
Fig. 15.2 Mind map summarizing the discussions

and accessible format. We also provided them with the slideshow used during the training, for further detail.

15.4 Students' Perceptions of the Training

A month after the training session, students answered a short questionnaire to assess it. At this point, students had had the opportunity to transfer knowledge into practice because they had already done the peer feedback activity. Of the 35 students who participated in the training and the peer feedback activity, 27 answered the questionnaire (response rate: 77%). At the end of the questionnaire, students could leave their contact information if they agreed to participate in an interview. We conducted semi-directed interviews with the five students who agreed to participate.

15.4.1 Questionnaire Conception and Interview Process

We constructed our questionnaire based on Grohmann and Kauffeld (2013)'s Questionnaire for professional training evaluation. This questionnaire is based on Kirkpatrick's framework (Kirkpatrick & Kirkpatrick, cited by Grohmann & Kauffeld, 2013) which distinguishes four levels: reaction, learning, behavior and organizational impact. Grohmann and Kauffeld (2013) have divided the reaction and organizational impact levels into two sub-levels which gives them six subscales (each one composed of two items): satisfaction (reaction level), utility (reaction level), knowledge (learning level), application to practice (behavior level), individual organizational results (organizational level) and global organizational results (organizational level). As the two last subscales are not relevant to our context, we limited our questionnaire to the first four subscales. The eight items of these subscales were subsequently adapted to the higher education context (e.g. the item "In my everyday work, I often use the knowledge I gained in the training" was replaced by "In the peer feedback activity, I used the knowledge I gained in the training") and translated to French.

In addition, we included four items from Holgado-Tello et al. (2006)'s training satisfaction rating scale, which measures participants' general impression of the training.

Our questionnaire is therefore composed of 12 items divided into five subscales (see Table 15.2 for details). In line with Grohmann and Kauffeld (2013), we used an 11-points response scale. The responses range from 0 per cent to 100 per cent, with steps of 10 per cent. The general impression scale is reliable with a Cronbach's alpha of 0.854. For the five other subscales, we calculate Spearman-Brown Coefficient, as it is recommended for two-item scales (Eisinga et al., 2013). All scales are reliable, with Spearman-Brown coefficients ranging between 0.752 and 0.929. At the end of the questionnaire, we allowed students to add a written comment.

The interviews were held using Teams and lasted approximately 30 min. We transcribed them and used N-Vivo (version 20.5) to code the data.

15.4.2 Insights from Questionnaire and Interview Data

As you can see in Fig. 15.3, students' general impression is generally positive (M = 66.5, SD = 16.4), with the majority of students rating the training around 70%. The satisfaction is a bit lower (see Fig. 15.4), with very high variability (M = 59.3, SD = 20). The same pattern is present for perceived utility (MD = 49.6, SD = 22.5), perceived learning (MD = 53.7, SD = 18.3) and behavioral changes (MD = 55.4, SD = 24) as shown in Figs. 15.5, 15.6 and 15.7.

The most striking result is the high variability of students' perceptions. For each scale (although to a lesser extent for the general impression), the standard deviation and range are wide, with some students who saw little value in the training (with

Table 15.2 Item examples, number of items and cronbachs' α of each subscales

Subscale	Adapted from	Example item	Number of items	Cronbach's α	Spearman-Brown coefficient
General impression	Holgado-Tello et al. (2006)	The training merits a good overall rating	4	0.854	/
Satisfaction	Grohmann and Kauffeld (2013)	I enjoyed the training session very much	2	0.874	0.884
Perceived utility		Participation in this kind of training is very useful for me	2	0.850	0.861
Perceived learning		I learned a lot of new things in the training	2	0.747	0.752
Behavioral changes		I successfully manage to apply the training contents in the peer feedback activity	2	0.929	0.929

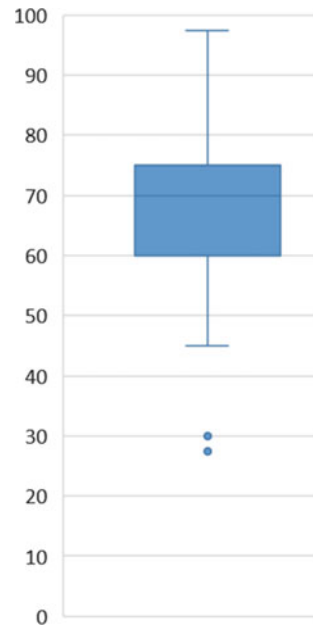**Fig. 15.3** Boxplot for the subscale general impression

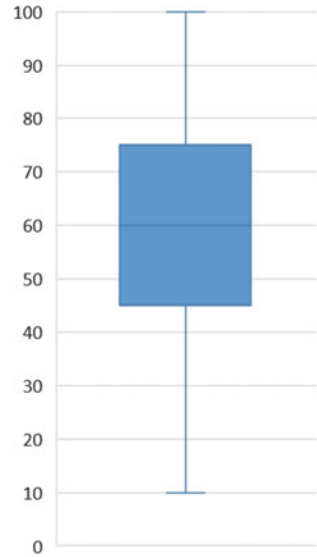**Fig. 15.4** Boxplot for the
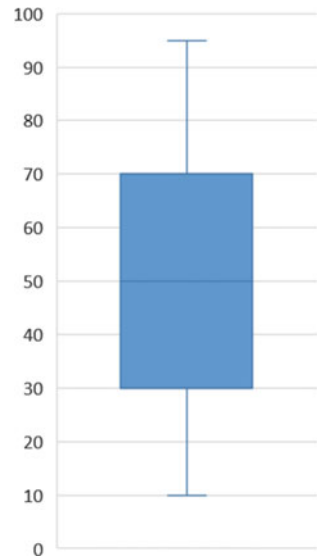subscale satisfaction

Fig. 15.5 Boxplot for the
subscale perceived utility

some aspects evaluated at only 10%) and others who seem to have very positive
perceptions of the training (with a score of 90 or 100%).

Fig. 15.6 Boxplot for the
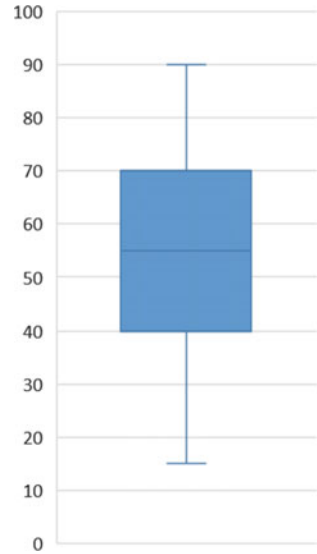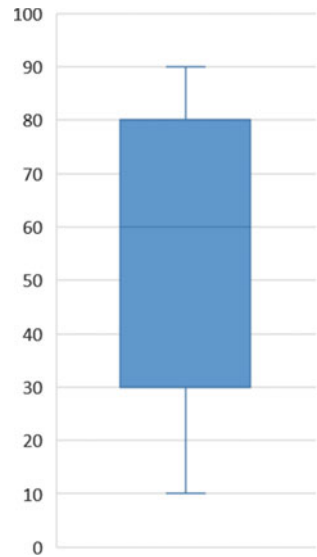subscale perceived learning

Fig. 15.7 Boxplot for the
subscale behavioral changes

We see two main reasons for the diversity of students' perceptions linked to the
training. The first concerns variability in students' needs as illustrated by quotes
from *David* and *Robin*.[1]

---

[1] Names have been changed to protect the participants' confidentiality.

«I want to point out that I was already familiar with the principle of peer feedback because I saw it in Research methodology in Master1. That is why I haven't learned as much as others, I think.» *David* (referring to a course in which the peer review process in research is explained; perceived learning: 55, perceived utility: 70).

«Before we had the course I clearly thought that I wasn't going to be able to… to provide feedback […] and then, with your course, I put this more into perspective, a lot more and I'll say that I was… I wanted to try and assess my peers, and I could see where I had to go.» *Robin* (perceived learning: 90, perceived utility: 95).

The variability in students' needs seems to stem from their previous experiences. While most students following this course were in their third year and only followed bachelor courses, some students, like *David*, were also taking some master courses (based on the number of ECTS they have acquired). Moreover, students also had different extracurricular experiences. Several physical education students had student jobs as sports instructors or coaches, for example, which enable them to develop assessment and feedback skills. Additionally, given that students practice various sports outside their courses, some students have a much higher level than others in acrobatic sports. This high level of expertise could lead them to overestimate their ability to easily provide feedback in this specific discipline. These factors could explain why some students expressed a strong need for guidance, while others felt they already had the necessary knowledge and skills before the training.

Another possible reason concerns the variability in students' implications in the training session and the seminar more generally. While some students were genuinely interested in the seminar contents, others only followed it because it was mandatory. Given that it was an online session and that students had their cameras off (as not to saturate their wifi), it was more difficult to discern if they were truly paying attention, or if they were even there. In an interview, for example, a student explained that for another session of the course he let his computer with Teams turned on to appear present and went for a run. We tried to make the session as interactive as possible to avoid this, but it is still possible that we lost some students at times.

Moreover, an important part of the training took place in small groups and we observed that some groups worked better than others did in this online context. Indeed, it is well-known that the physical presence of an educator is important for student engagement (e.g. Hunter, cited by Bolinger & Stanton, 2020). The set-up made it difficult for us to know whether the students were taking the role-play seriously and to quickly identify which groups needed help. Although there were two instructors to visit each break-out room, students were just among themselves most of the time and, while most groups seemed to work efficiently, we felt others needed a little push to keep working seriously. This feeling was confirmed by some of the comments in the interviews.

«Well, in the group I was in […] There was some misunderstanding in the group, and we botched the part where we were supposed to take the role.» *Raphaël*,

who liked the training (satisfaction score of 75), but did not feel that it was useful (perceived utility score of 30) or that he learned from it (perceived learning score of 20).

This variation in students' engagement could explain why some students felt they learned less from the training or said they did not really apply the training content while doing the peer feedback activity.

15.5   Conclusion and Implications

Our goal was to create an online training to tackle students' feelings of distrust and psychological unsafety before participating in a peer feedback activity. To this aim, we explored the literature to find effective learning methods. The training that we created was composed of five stages (discovery of students' representations, theory, practice, role-plays and summary), which allowed students' active participation. This training was implemented in the context of a physical education university seminar and we collected data on how students perceived it. Based on this, we can draw some conclusions, and stemming from them, implications for practice and perspectives for research.

When given the opportunity students express concerns linked to the social nature of peer feedback. Indeed, the gathered responses showed that, even though students saw various advantages to peer feedback, they also raised a series of concerns, like the fear of not being qualified enough or the fear that their peers will be "too nice" when assessing them. This confirms previous findings (e.g. Mostert & Snowball, 2013; Mulder et al., 2014; Wilson et al., 2015) and suggests that, when planning a peer feedback activity, it is essential to take time to let students express these concerns and to address them.

The literature on training methods (e.g. Blanchard & Thacker, 2013) and the studies we drew upon to create the training (e.g. Dusenberry & Robinson, 2020) converged on the idea that (inter)active learning methods are necessary. This training with active learning methods was delivered online, thus making it a potentially useful part of MOOCs (Suen, 2014). For other courses, while normally on campus, online alternatives had to be sought during the COVID-19 outbreak. Indeed, according to UNESCO (2021), more than 220 million tertiary students worldwide have been confronted with university closures and online courses. It is therefore important to conceive learning activities such as role-play that can take place online. In the present case, the group was small enough so we could quickly visit each online break-out room during the sub-groups activities, however, students are generally far more numerous in MOOCs. Future studies should investigate if the training is feasible with larger groups of students.

We obtained encouraging insights when asking students' opinions about the training. About half the students were positive: they said that they learned from it, found it useful and that it influenced their behavior during the peer feedback activity. Other students valued the training less. Differences in students' perceptions may be explained by factors like their prior knowledge or experience with

peer feedback or by their varying engagement with the training (due to the online context). These factors could be investigated in future studies. We expect students' perceptions to be less variable if the training is delivered on campus, as physical presence incites engagement (Hunter, 2004, as cited in Bolinger & Stanton, 2020).

The overall positive impact of the training will have to be confirmed in future studies. Indeed, an evaluation study was beyond the scope of this chapter, in which we focused on students' perceptions of the training. A quasi-experimental study, with a large sample and pre- and post-test should be conducted to verify whether the training has a positive impact on students' perceived level of trust and psychological safety. In addition, given that the goal of peer feedback is to develop student learning and to incite them to be proactive recipients of feedback, such a study could investigate the impact of the training on students' learning due to the peer feedback activity, as well as regarding their feedback literacy skills (Boud et al., 2022).

Now the training focuses predominantly on peer feedback provision; its objectives are to teach students how to provide effective feedback and to ensure they feel safe and confident while doing so. A perspective could be to redesign it so it also encompasses peer feedback processing. Indeed, no matter how good the received feedback is, students still need the support of an adequate learning context to efficiently use it to revise their work (Panadero & Lipnevich, 2022; Wichmann, 2018). Given that interpersonal factors—such as trust and psychological safety—play a role in feedback provision, but also in peer feedback processing (Aben et al., 2019), it would be interesting to create an intervention that explicitly considers the social aspects that play a role in peer feedback processing.

All in all, it seems that an online training with (inter)active methods such as role-plays is a promising way to address students' concerns raised by the social nature of peer feedback.

References

Aben, J. E. J., Dingyloudi, F., Timmermans, A. C., & Strijbos, J. (2019). Embracing errors for learning: Intrapersonal and interpersonal factors in feedback provision and processing in dyadic interactions. In *The impact of feedback in higher education: Improving assessment outcomes for learners* (pp. 107–125). https://doi.org/10.1007/978-3-030-25112-3_7.

Ajjawi, R., & Boud, D. (2017). Researching feedback dialogue: An interactional analysis approach. *Assessment and Evaluation in Higher Education, 42*(2), 252–265. https://doi.org/10.1080/02602938.2015.1102863

Alqassab, M., Strijbos, J., & Ufer, S. (2018). Training peer-feedback skills on geometric construction tasks: Role of domain knowledge and peer-feedback levels. *European Journal of Psychology of Education, 33*(1), 11–30. https://doi.org/10.1007/s10212-017-0342-0

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Blanchard, P. N. & Thacker, J. W. (2013). *Effective training. systems, strategies, and practices.* Pearson Education Limited.

Bolinger, A. R., & Stanton, J. V. (2020). *Role-play simulations.* Elgar.

Boud, D., Dawson, P., Yan, Z., Lipnevich, A., Tai, J., & Mahoney, P. (2022). *Measuring learners' capabilities to engage in feedback: The Feedback Literacy Scale* [Paper presentation]. EARLI SIG 1 & 4 joint conference 2022, Cadiz, Spain.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment and Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education, 36*(4), 395–407. https://doi.org/10.1080/03075071003642449

Cheng, K., & Tsai, C. (2012). Students' interpersonal perspectives on, conceptions of and approaches to learning in online peer assessment. *Australasian Journal of Educational Technology, 28*(4), 599–618. https://doi.org/10.14742/ajet.830.

De Ketele, J., Chastrette, M., & Cros, D. (2007). *Guide du formateur* (3e éd. ed.). De Boeck.

De Stobbeleir, K., Ashford, S., & Zhang, C. (2020). Shifting focus: Antecedents and outcomes of proactive feedback seeking from peers. *Human Relations, 73*(3), 303–325. https://doi.org/10.1177/0018726719828448

Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review, 32*(2), 481–509. https://doi.org/10.1007/s10648-019-09510-3

Dusenberry, L., & Robinson, J. (2020). Building psychological safety through training interventions: Manage the team, not just the project. *IEEE Transactions of Professional Communication, 63*(3). https://doi.org/10.1109/TPC.2020.3014483.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly, 44*(2), 350–383. https://doi.org/10.2307/2666999

Edmondson, A., Dillon, J., & Roloff, K. (2007). Three perspectives on team learning: outcome improvement, task mastery, and group process. In A. Brief, & J. Walsh, J. (Eds.), *The academy of management annals, Volume 1. Acad Manag Ann* (Vol. 1, pp. 269–314). https://doi.org/10.1080/078559811.

Edmondson, A., & Lei, Z. (2014). Psychological safety: The history, renaissance, and future of an interpersonal construct. In *Annual review of organizational psychology and organizational behavior* (Vol. 1, pp. 23–43). Annual Reviews Inc. https://doi.org/10.1146/annurev-orgpsych-031413-091305.

Eisinga, R. N., Grotenhuis, H. F. T., & Pelzer, B. J. (2013). The reliability of a two-item scale: Pearson, cronbach or spearman-brown? *International Journal of Public Health, 58*(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3.

Gabelica, C., Bossche, P. V. D., Segers, M., & Gijselaers, W. (2012). Feedback, a powerful lever in teams: A review. *Educational Research Review, 7*(2), 123–144. https://doi.org/10.1016/j.edurev.2011.11.003.

Grohmann, A., & Kauffeld, S. (2013). Evaluating training programs: Development and correlates of the questionnaire for professional training evaluation. *International Journal of Training and Development, 17*(2), 135–155. https://doi.org/10.1111/ijtd.12005

Hanrahan, S., & Isaacs, G. (2001). Assessing self- and peer-assessment: The students' views. *Higher Education Research & Development, 20*(1), 53–68. https://doi.org/10.1080/07924360120043658

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/00346543029848

Hew, K. F. (2018). Unpacking the strategies of ten highly rated MOOCs: Implications for engaging students in large online courses. *Teachers College Record, 120*(1), 1–40. https://doi.org/10.1177/016146811812000107

Holgado Tello, F. P., Chacón Moscoso, S., Barbero García, I., & Sanduvete Chaves, S. (2006). Training satisfaction rating scale: Development of a measurement model using polychoric correlations. *European Journal of Psychological Assessment: Official Organ of the European Association of Psychological Assessment, 22*(4), 268–279. https://doi.org/10.1027/1015-5759.22.4.268

Li, L. (2017). The role of anonymity in peer assessment. *Assessment and Evaluation in Higher Education, 42*(4), 645–656. https://doi.org/10.1080/02602938.2016.1174766

Mostert, M., & Snowball, J. D. (2013). Where angels fear to tread: Online peer-assessment in a large first-year class. *Assessment and Evaluation in Higher Education.* https://doi.org/10.1080/02602938.2012.683770

Mulder, R. A., Pearce, J. M., & Baik, C. (2014). Peer review in higher education: Student perceptions before and after participation. *Active Learning in Higher Education.* https://doi.org/10.1177/1469787414527391

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 247–265). Routledge/Taylor & Francis.

Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment and Evaluation in Higher Education, 44*(8), 1253–1278. https://doi.org/10.1080/02602938.2019.1600186

Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review, 35*, 100416. https://doi.org/10.1016/j.edurev.2021.100416

Rotsaert, T., Panadero, E., & Schellens, T. (2018). Anonymity as an instructional scaffold in peer assessment: Its effects on peer feedback quality and evolution in students' perceptions about peer assessment skills. *European Journal of Psychology of Education*, *33*(1), 75–99. https://doi.org/10.1007/s10212-017-0339-8.

Shah, D. (2020). *By the numbers: MOOCs during the pandemic.* Retrieved from https://www.classcentral.com/report/mooc-stats-pandemic.

Sluijsmans, D. M. A., Brand-Gruwel, S., van Merriënboer, J. J. G., & Bastiaens, T. J. (2002). The training of peer assessment skills to promote the development of reflection skills in teacher education. *Studies in Educational Evaluation, 29*(1), 23–42. https://doi.org/10.1016/S0191-491X(03)90003-4

Sluijsmans, D. M. A., Brand-Gruwel, S., van Merriënboer, J. J. G., & Martens, R. L. (2004). Training teachers in peer-assessment skills: Effects on performance and perceptions. *Innovations in Education and Teaching International, 41*(1), 59–78. https://doi.org/10.1080/1470329032000172720

Soares, A. E., & Lopes, M. P. (2020). Are your students safe to learn? The role of lecturer's authentic leadership in the creation of psychologically safe environments and their impact on academic performance. *Active Learning in Higher Education, 21*(1), 65–78. https://doi.org/10.1177/1469787417742023

Suen, H. K. (2014). Peer assessment for massive open online courses (MOOCs). *International Review of Research in Open and Distance Learning, 15*(3), 312–327. https://doi.org/10.19173/irrodl.v15i3.1680.

Sung, Y., Chang, K., Chang, T., & Yu, W. (2009, 2010). How many heads are better than one? The reliability and validity of teenagers' self- and peer assessments. *Journal of Adolescence, 33*(1), 135–145. https://doi.org/10.1016/j.adolescence.2009.04.004.

Stanier, L. (1997). Peer assessment and group work as vehicles for student empowerment: A module evaluation. *Journal of Geography in Higher Education.* https://doi.org/10.1080/03098269708725413

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3.

UNESCO. (2021). *COVID-19: reopening and reimagining universities, survey on higher education through the UNESCO National Comissions.* Retrieved from: https://unesdoc.unesco.org/ark:/48223/pf0000378174.

van Gennip, N., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review.* https://doi.org/10.1016/j.edurev.2008.11.002

van Gennip, N., Segers, M. S. R. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction*, *20*(4), 280–290. https://doi.org/10.1016/j.learninstruc.2009.08.010.

van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20*(4), 270–279. https://doi.org/10.1016/j.learninstruc.2009.08.004

Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers and Education, 81*, 123–132. https://doi.org/10.1016/j.compedu.2014.10.001

Wichmann, A., Funk, A., & Rummel, N. (2018). Leveraging the potential of peer feedback in an academic writing activity through sense-making support. *European Journal of Psychology of Education, 33*(1), 165. https://doi.org/10.1007/s10212-017-0348-7

Wilson, M. J., Diao, M. M., & Huang, L. (2015). 'I'm not here to learn how to mark someone else's stuff': An investigation of an online peer-to-peer review workshop tool. *Assessment and Evaluation in Higher Education, 40*(1), 15–32. https://doi.org/10.1080/02602938.2014.881980

Wisniewski, B., Zierer, K., & Hattie, J. (2020, 2019). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology, 10*, 3087–3087. https://doi.org/10.3389/fpsyg.2019.03087.

Yu, F.-Y. Y., & Liu, Y.-H. H. (2009). Creating a psychologically safe online space for a student-generated questions learning activity via different identity revelation modes. *British Journal of Educational Technology*, *40*(6), 1109–1123. https://doi.org/10.1111/j.1467-8535.2008.00905.x.

The Relationship Among Students' Attitude Towards Peer Feedback, Peer Feedback Performance, and Uptake

16

Nafiseh Taghizadeh Kerman, Seyyed Kazem Banihashem, and Omid Noroozi

16.1 Introduction

The use of peer feedback in higher education, particularly in online classes with large size of students has been considerably growing (Latifi et al., 2021; Yang, 2016), especially in writing classes (e.g., Noroozi & Hatami, 2019; Shang, 2019). For example, in the context of argumentative essay writing, peer feedback is acknowledged as an active and effective learning activity since it involves students in a learning process where they deal with critical reading, critical reflection, and creating constructive knowledge that leads to enhancing peers' argumentative essay writing competence (Noroozi, 2018, 2022; Noroozi & Hatami, 2019; Tian & Zhou, 2020).

N. T. Kerman (✉)
Ferdowsi University, Mashhad, Iran
e-mail: na_ta249@mail.um.ac.ir

S. K. Banihashem
Wageningen University and Research, Wageningen, The Netherlands

Open Universiteit, Heerlen, The Netherlands

O. Noroozi
Education and Learning Sciences, Wageningen University and Research, Wageningen, The Netherlands

According to previous studies, using peer feedback in higher education can improve students' evaluation and judgment skills (Liu & Carless, 2006), self-regulation skills (Lin, 2018a, 2018b), communication, collaboration, and negotiation skills (e.g., Altınay, 2016; Bayat et al., 2022; Lai, 2016; Lai et al., 2020), critical thinking skills (e.g., Ekahitanond, 2013; Novakovich, 2016), engagement (e.g., Devon et al., 2015; Fan & Xu, 2020), motivation (e.g., Hsia et al., 2016; Zhang et al., 2014), and learning satisfaction (e.g., Donia et al., 2022; Zhang et al., 2014).

The success of peer feedback mainly depends on its quality (Carless et al. 2011; Er et al., 2021; Hattie & Timperley, 2007; Latifi et al., 2020; Taghizadeh et al., 2022; Shute, 2008). If students find the received feedback of high quality, they are more likely to uptake and implement it in their essays (Wu & Schunn, 2020). For the feedback to be effective, it should contain features such as affective statements (e.g., praise or compliment), a summary explanation of the work, identifications, and localization of the problem, and solutions and action plans to the identified problems and further improvements (Banihashem et al., 2022; Noroozi et al., 2012; Patchan et al., 2016; Wu & Schunn, 2021).

Empirical research has revealed a number of issues related to peer feedback (Latifi & Noroozi, 2021; Latifi et al., 2021; Noroozi et al., 2012, 2018; Panadero, 2016; Zhao, 2018; Zhu & Carless, 2018). One of the challenges is the perception of distrust in peers' competence to provide high-quality feedback (Kaufman & Schunn, 2011; Liu & Carless, 2006; Zhu & Carless, 2018). Students are skeptical in terms of receiving high-quality feedback from peers as they perceive peers' knowledge may not good enough to identify the problem or may not even their peers take it seriously to carefully read and provide constructive feedback (Hu, 2005; Panadero & Alonso-Tapia, 2013; Tsui & Ng, 2000; Vu & Dall'Alba, 2007). One reason is that students may have a different perceived level of domain knowledge and feedback proficiency that can cause a different impact on levels of contribution and motivation of students (Allen & Mills, 2016; Wu, 2019). For example, students with high feedback proficiency are demotivated because they have little faith in and perception of the quality of the feedback received from peers with low feedback proficiency (Jiang & Yu, 2014). Therefore, students' performance and uptake of peer feedback can be influenced by their attitude towards peer feedback.

Attitude is defined as the psychological evaluations a person makes of people, objects, or events (Gagne et al., 2005). Attitude towards peer feedback means how students perceive peer feedback and what they feel about providing or receiving peer feedback. Attitude towards peer feedback includes multiple components. For example, perceived fairness (Lin, 2018a, 2018b), perceived usefulness (Kuo, 2017), perceived learning outcomes (Chan & Lin, 2019; Lin et al., 2016, 2018; Noroozi & Mulder, 2017), and perceived ease to use (Kuo, 2017; Ge, 2019). Although attitudes are largely internal and particular to each person, they are socially impacted and changed by how other people behave (Bordens and Horowitz, 2008). Many factors change attitudes, especially attitudes toward peer feedback. For example, defining peer feedback goals (Topping, 2017), training and

the required instruction and direction (Falchikov, 2005; Morra and Romano, 2008, 2009), providing argumentative peer feedback (Noroozi & Hatami, 2019), using the mobile peer feedback strategy (Kuo, 2017), online peer feedback with TQM (Lin, 2016), anonymous condition (Lin, 2018), guided peer feedback (Noroozi & Mulder, 2017), using the blogging (Rahmany et al., 2013), accurate and specific feedback (Wang et al., 2019) caused attitudinal change towards online peer feedback and learning.

Prior studies also have shown that students' perceptions of peer feedback plays an influential role in their peer feedback performance and uptake (Chou, 2014; Collimore et al., 2014; Paré & Joordens, 2008; Prins et al., 2010; Wen & Tsai, 2006; Zou et al., 2017). If students have a positive attitude towards peer feedback, they are more likely to provide feedback and to take the received feedback more seriously into account, while a negative attitude towards peer feedback may not motivate them enough to actively participate in the peer feedback process (Azarnoosh, 2013; Lin et al., 2001). For example, Mishra et al. (2020) and Mulder et al. (2014) reported that students' attitude towards peers' competence in providing good feedback or even in a larger scope students' perceptions about the usefulness of the peer feedback is one of the key factors that can influence students' peer feedback performance and uptake. Because students who perceived peer feedback useful were more likely to accept it by acknowledging their mistakes, indicating that they want to change their material, and/or appreciating the effectiveness of the peer feedback (Misiejuk et al., 2021; Noroozi et al., 2016). Studies have shown that if students do not perceive peer feedback as a useful activity and if they do not perceive their peers as knowledgeable and reliable feedback providers, they are less likely to uptake feedback and implement it in their work (Harks et al., 2014; Noroozi & Mulder, 2017).

Although the evidence showed that students' attitude towards peer feedback and peer feedback performance and uptake can influence each other (e.g., Alhomaidan, 2016; Kuyyogsuy, 2019; Noroozi et al., 2022), this has not been largely investigated in online learning environments in the context of argumentative essay writing. Little is known how students' attitude towards peer feedback relates to students' peer feedback performance and uptake, in the context of argumentative essay writing in an online mode of education (Alhomaidan, 2016; Kuyyogsuy, 2019). There is also little known about how the quality of the received peer feedback can influence students' attitude towards peer feedback. For example, if students receive high-quality feedback from their peers can it improve students' attitude towards peer feedback in the context of argumentative essay writing.

16.2   Purpose of the Present Study

Therefore, this study was conducted to further explore this by answering the following research questions.

1. To what extent does students' attitude towards peer feedback predict peer feedback performance in the context of argumentative essay writing in online education?
2. To what extent does students' attitude towards peer feedback predict the uptake of peer feedback in the context of argumentative essay writing in online education?
3. To what extent does the quality of the received peer feedback predict students' attitude towards peer feedback in the context of argumentative essay writing in online education?

16.3   Method

16.3.1 Sample

In this study, 135 undergraduate students participated, however, only 101 students have completed the module. About 69% of participants were female (N = 70) and 31% of participants were male (N = 31). Out of 101 participants, 79 students completed the attitude towards peer feedback questionnaire. As a results, the sample size of 79 was analysis. To comply with ethical considerations, participants were informed about the research setup of the module. They were assured that no data can be linked to any individual participant. Furthermore, ethical approval from the Social Sciences Ethics Committee at Wageningen University and Research was obtained for this study.

16.4   Instrument

16.4.1 Students' Argumentative Essay Performance

To measure the quality of students' argumentative essay performance, a coding scheme adjusted based on Noroozi et al. (2016) instrument was used. This coding scheme was developed based on a high-quality argumentative essay structure which comprised of eight elements including (1) introduction on the topic, (2) taking a position on the topic, (3) arguments for the position, (4) justifications for arguments for the position, (5) arguments against the position, (6) justifications for arguments against the position, (7) response to counter-arguments, and (8) conclusion and implications. Each element is scored from 0 points (not mentioned at all) to 3 points (mentioned with the highest quality) (Table 16.1). All given points for these elements are summed up together and indicate the student's total score

for the quality of the written argumentative essay. This coding scheme was used in two phases. In the first phase, it was used to assess students' first draft of the essay and in the second phase, it was used to assess students' revised version of the essay. The quality of students' argumentative essays was assessed based on the differences in their performances in the first draft and revised draft of the essay. Two coders with expertise in education contributed to the coding of the quality of written argumentative essays. Cohen's kappa coefficient analysis was used to measure the inter-rater reliability between the coders and the results showed that there is a reliable agreement between the coders (*Kappa = 0.70, p < 0.001*). According to Landis and Koch (1977) and McHugh (2012) classification for Cohen's Kappa coefficients, 0.70 is substantial.

16.4.2 Students' Online Peer Feedback Performance

To measure the quality of students' online peer feedback, a coding scheme was designed by the authors based on the review of related previous studies mainly (e.g., Nelson & Schunn, 2009; Patchan et al., 2016; Wu & Schunn, 2020). This coding scheme entails four main categories including affective, cognitive (description, identification, and justification), and constructive features feedback. The coding scheme was scored from 0 points (poor) to 2 points (good) for all the categories. All points were summed up and determined the quality of online peer feedback performance (Table 16.2). Since each student provided and received two sets of feedback, the mean score of both feedback was identified as the quality of online peer feedback for each student. Similar to the argumentative essay analysis, the same two coders participated in the coding process for peer feedback analysis, and Cohen's kappa coefficient results for inter-rater reliability among coders were found to be significant (*Kappa = 0.60, p < 0.001*). According to Landis and Koch (1977) and McHugh (2012) classification for Cohen's Kappa coefficients, 0.60 is moderate and acceptable.

16.4.3 Students' Attitude Towards Peer Feedback

The authors developed a questionnaire with a 19-item to measure students' attitude towards peer feedback. All items of this questionnaire were designed on a five-point Likert scale ranging "*strongly disagree = 1,*" "*disagree = 2,*" "*neutral = 3,*" "*agree = 4*", and "*strongly agree = 5.*" This questionnaire entails four main sections including perceived usefulness of peer feedback, perceived motivation of peer feedback, perceived trustworthiness of peer feedback, and perceived fairness of peer feedback. The reliability coefficient was high for all four scales of this instrument (*Cronbach α = 0.82, 0.80, 0.76*, and *0.84*). Also, we did factor analysis with Lisrel software 8.80 for the students' attitude towards peer feedback questionnaire. If the vast majority of the indexes indicate a good fit, then there is

Table 16.1 Coding scheme to analyze the quality of students' argumentative essay writing

Variables	Points	Labels	Descriptions
Introduction on the topic	0	Not mentioned at all	Introduction on the topic is not presented at all
	1	Just mentioned	Introduction on the topic is just presented, but not elaborated and justified
	2	Mentioned and elaborated	Introduction on the topic is presented and elaborated, but not justified
	3	Mentioned, elaborated, and justified	Introduction on the topic is presented, elaborated, and justified
Taking a position on the topic	0	Not mentioned at all	Position on the topic is not presented at all
	1	Just mentioned	Position on the topic is just presented, but not elaborated and justified
	2	Mentioned and elaborated	Position on the topic is presented and elaborated, but not justified
	3	Mentioned, elaborated, and justified	Position on the topic is presented, elaborated, and justified
Arguments for the position	0	Not mentioned at all	No argument in favour of the position is presented
	1	Mentioned to a small extent	Only one argument in favour of the position is presented
	2	Mentioned to a moderate extent	Only two arguments in favour of the position are presented
	3	Mentioned to a great extent	More than two arguments in favour of the position are presented
Justifications for arguments for the position	0	Not justified at all	Justification for arguments for the position is not presented at all
	1	Justified to a small extent	Only one argument for the position is justified
	2	Justified to a moderate extent	Some but not all arguments for the position are justified

(continued)

Table 16.1   (continued)

Variables	Points	Labels	Descriptions
	3	Justified to a great extent	All arguments for the position are justified
Arguments against the position (counter-arguments)	0	Not mentioned at all	No argument against the position is presented
	1	Mentioned to a small extent	Only one argument against the position is presented
	2	Mentioned to a moderate extent	Only two arguments against the position are presented
	3	Mentioned to a great extent	More than two arguments against the position are presented
Justifications for arguments against the position	0	Not justified at all	Justification for arguments against the position is not presented at all
	1	Justified to a small extent	Only one argument against the position is justified
	2	Justified to a moderate extent	Some but not all arguments against the position are justified
	3	Justified to a great extent	All arguments against the position are justified
Response to counter-arguments	0	Not mentioned at all	Response to counter-arguments is not presented at all
	1	Just mentioned	Response to counter-arguments is just presented, but not elaborated and justified
	2	Mentioned and elaborated	Response to counter-arguments is presented and elaborated, but not justified
	3	Mentioned, elaborated, and justified	Response to counter-arguments is presented, elaborated, and justified
Conclusion and implications	0	Not mentioned at all	Conclusion and/or implications are not presented at all

Table 16.1 (continued)

Variables	Points	Labels	Descriptions
	1	Just mentioned	Conclusion and/or implications are just presented, but not elaborated and justified
	2	Mentioned and elaborated	Conclusion and/or implications are presented and elaborated, but not justified
	3	Mentioned, elaborated, and justified	Conclusion and/or implications are presented, elaborated, and justified

probably a good fit. Schreiber et al. (2006) suggested that for continuous data— $\chi2/df \leq 2$ or 3, CFI > 0.95, IFI > 0.95, GFI > 0.95, AGFI > 0.95, and RMSEA < 0.06 or 0.08. Our results revealed that standardized loading estimates of each element were greater than 0.70. Also, the result of Confirmatory Factor Analysis (CFA) for students' attitude towards peer feedback questionnaire showed that the single-factor model provides good fit indices [$\chi2$ (2) = 5.43, p > 0.05, $\chi2/df$ = 2.71, Comparative Fit Index (CFI) = 0.99, Incremental Fit Index (IFI) = 0.99, Goodness of Fit Index (GFI) = 0.99, Adjusted Goodness of Fit Index (AGFI) = 0.94, Root Mean Square Error of Approximation (RMSEA) = 0.08.

16.4.4 Design

This study is a part of a bigger project that took place at Wageningen University and Research in the 2020–2021 academic year. As a part of a bigger project, one course from Environmental Science was selected for this study, and the module called the "*Argumentative Essay Writing*" was designed and embedded in the course at the Brightspace platform. The module was followed by the students in three consecutive weeks and for each week they were requested to complete a specific task. In the first week, students were asked to write an argumentative essay on one of the three provided controversial topics including (a) the long-term impacts of Covid-19 on the environment, (b) the role of private actors in funding local and global biodiversity, and (c) bans on the use of single-use plastics. The word limit for this argumentative essay is 600 to 800 words (excluding references). All students were requested to write their essays within the determined work limit. Since all students were the same, therefore, all students performed their essays in the same condition, the effects of word count is controlled. In the second week, students were invited to provide feedback on the argumentative essays of two peers based on specific given criteria. Each student provided and received two sets of feedback (30 to 50 words for each element) on peers' essay performance based on

Table 16.2 Coding scheme to analyze the quality of students' online peer feedback performance

Nature of feedback	Feedback features	Points	Label	Description
Affective		0	Poor—discouraging	The comment included discouraging and negative emotions such as anger or disappointment
		1	Average—neutral/not mentioned	The comment did not include either negative or positive emotions
		2	Good—encouraging	The comment included encouraging and positive emotions such as praise or compliments
Cognitive	Description	0	Poor—not mentioned	The comment did not include a summary statement such as the description of the content or the taken action
		1	Average—mentioned to a small extent	The comment included a summary statement such as the description of the content or the taken action but to a small extent
		2	Good—mentioned to a large extent	The comment included a summary statement such as the description of the content or taken action to a large extent
	Identification	0	Poor—not mentioned	The comment did not include explicit identification of the problem
		1	Average—mentioned but not localized	The comment included identification of problem without localization of identified problem
		2	Good—mentioned and localized	The comment included explicit and localized identification of the problem

Table 16.2 (continued)

Nature of feedback	Feedback features	Points	Label	Description
	Justification	0	Poor—not mentioned	The comment did not include elaborations[a] and justifications[b] of the identified problem
		1	Average—mentioned, elaborated, but not justified	The comment included elaborations but not justifications of the identified problem
		2	Good—mentioned, elaborated, and justified	The comment included elaborations and justifications of the identified problem
Constructive		0	Poor—not mentioned	The comment did not include any recommendations or action plans for further improvements
		1	Average—only recommendation is mentioned	The comment included recommendations but not action plans for further improvements
		2	Good—both recommendation and action plan are mentioned	The comment included recommendations and action plans for further improvements

[a]Elaborations: refers to students' explanations, reasons to support "why the identified problem" should be taken into account by the feedback receiver
[b]Justifications: refers to the scientific facts, references, and reliable and valid examples to support elaborations

the criteria embedded in the FeedbackFruits app within the Brightspace platform. It should be noted that students did not receive more than two sets of feedback from their peers on their essays. In the third week, students were asked to revise their original argumentative essay based on the two received feedback sets provided by their peers. Students were informed that this module is a part of their course and it is necessary for them to complete all tasks offered within the proposed time and deadline. Students received an extra bonus for completing this module.

16.4.5 Analysis

In this study, descriptive analysis was used to show an overview of students' attitude towards peer feedback in the context of argumentative essay writing in an

online learning environment. The Kolmogorov–Smirnov test was used to determine whether the distribution of the data was normal or not and it was found that data were normally distributed ($p > 0.05$). Also, collinearity effects were checked in regression models. If Variance Inflation Factor (VIF) value was lower than the cut-off score 10 and Tolerance value was lower than the cut-off score 1, an indication that is no multicollinearity problem (Miles, 2014). Tests to see if the data met the assumption of collinearity in this study indicated that multicollinearity was not a concern (perceived usefulness of peer feedback Tolerance = 0.37, VIF = 2.64; perceived motivation/enjoyment of peer feedback Tolerance = 0.70, VIF = 1.41; perceived trustworthiness of peer feedback Tolerance = 0.33, VIF = 2.97; perceived fairness of peer feedback Tolerance = 0.56, VIF = 1.76). Then, a multiple linear regression test was used to answer the research questions.

16.5 Results

An overview of students' attitude towards peer feedback in the context of argumentative essay writing in an online learning environment is presented in Table 16.3. The percentages provided for each of the attitude components include perceived usefulness of peer feedback, perceived motivation/enjoyment of peer feedback, perceived trustworthiness of peer feedback, and perceived fairness of peer feedback. Almost 66% of students stated that they perceived feedback from peers as a useful learning activity. Almost 55% of students stated that peer feedback is motivational for them. About 60% of students stated that they trust feedback from peers. About 69% of students perceived peer feedback as fair as teacher feedback.

RQ1: To what extent does students' attitude towards peer feedback predict peer feedback performance in the context of argumentative essay writing in online education?

The results showed that students' attitude did not predict peer feedback performance ($F(4, 73) = 1.21$, $p = 0.31$) (Table 16.4). Students who had a better perception of peer feedback did not perform better in providing feedback to their peers.

RQ2: To what extent does students' attitude towards peer feedback predict the uptake of peer feedback in the context of argumentative essay writing in online education?

The results showed that students' attitude did not predict uptake of peer feedback ($F(4, 74) = 1.54$, $p = 0.19$). However, the perceived usefulness of peer feedback was a significant predictor for uptaking of peer feedback (Table 16.5). Students who perceived useful feedback from their peers significantly were more progress from pre-test to post-test in argumentative essay writing improvement.

RQ3: To what extent does the quality of the received peer feedback predict students' attitude towards peer feedback in the context of argumentative essay writing in online education?

The results showed that the quality of the received peer feedback including justification and constructive features of feedback can predict students' attitude

Table 16.3 Descriptive results for students' attitude towards peer feedback in the context of argumentative essay writing in online education (n = 79)[a]

Attitude towards peer feedback	Item	Mean	SD	Agreement N. (%)[b]	Disagreement N. (%)[c]	Neutral N. (%)
Perceived usefulness of peer feedback	Peer feedback was helpful for argumentative essay writing	3.96	0.85	62 (78.48)	5 (6.32)	12 (15.18)
	Peer feedback was as valuable as teacher's feedback	3.12	0.92	32 (40.50)	22 (27.84)	26 (32.91)
	Peer feedback helped me to better structure my argumentative essay	3.59	1.03	51 (64.55)	12 (15.18)	16 (20.25)
	I learned when I provided feedback to my peers' argumentative essays	3.83	0.74	60 (75.94)	5 (6.32)	14 (17.72)
	I learned when I received feedback from my peers on my argumentative essay	3.72	0.86	56 (70.88)	7 (8.86)	16 (20.25)
Perceived motivation of peer feedback	I enjoyed giving feedback to my peers' works	3.24	1.00	30 (37.97)	17 (21.51)	32 (40.50)
	I enjoyed receiving feedback from my peers on my works	3.60	0.88	47 (59.49)	7 (8.86)	25 (31.64)
	Peer feedback activities motivated me to engage in learning assignments	3.37	0.95	36 (45.56)	13 (16.45)	30 (37.97)

Table 16.3 (continued)

Attitude towards peer feedback	Item	Mean	SD	Agreement N. (%)[b]	Disagreement N. (%)[c]	Neutral N. (%)
	I felt proud when I receive positive peer feedback on my works	3.88	0.84	56 (70.88)	5 (6.32)	18 (22.78)
	I felt comfortable giving critical feedback to my peers' works	3.62	1.01	49 (62.02)	14 (17.72)	16 (20.25)
Perceived trustworthiness of peer feedback	I think my peers had enough knowledge to provide reliable feedback on my argumentative essay	3.50	0.88	41 (51.89)	8 (10.12)	30 (37.97)
	My peers evaluated my argumentative essay appropriately	3.75	0.78	57 (72.15)	7 (8.86)	15 (18.98)
	I was willing to have my argumentative essay reviewed by learning peers	4.10	0.77	68 (86.07)	3 (3.79)	8 (10.12)
	My learning peers were able to identify the mistakes and errors in my argumentative essay	3.65	0.86	52 (65.82)	7 (8.86)	20 (19.80)
	I trusted my learning peers as much as teachers when it comes to feedback on my argumentative essay	3.80	0.97	20 (25.31)	31 (39.24)	28 (35.44)

Table 16.3 (continued)

Attitude towards peer feedback	Item	Mean	SD	Agreement N. (%)[b]	Disagreement N. (%)[c]	Neutral N. (%)
Perceived fairness of peer feedback	The feedback I received from my peers on my argumentative essay was fair	4.05	0.74	63 (79.74)	2 (2.53)	14 (17.72)
	I deserved the feedback I received from my peers on my argumentative essay	3.94	0.65	64 (81.01)	2 (2.53)	13 (16.45)
	The feedback I received from my peers was as fair as the teacher's feedback	3.37	0.86	35 (44.30)	11 (13.92)	33 (41.77)
	I am satisfied with the level of fairness of feedback I received from my peers	3.81	0.75	56 (70.88)	4 (5.06)	19 (24.05)

Note [a] Based on a 5-point Likert scale (Strongly disagree, disagree, neutral, agree, and strongly agree)
[b] Agreement = Agree, and strongly agree
[c] Disagreement = Strongly disagree, disagree

Table 16.4 Students' attitude towards peer feedback and peer feedback performance in the context of argumentative essay writing in online education

Attitude towards peer feedback	Mean	SD	Results
Perceived usefulness of peer feedback	3.63	0.67	$t = -0.08$, $p = 0.92$
Perceived motivation of peer feedback	3.55	0.69	$t = 1.42$, $p = 0.15$
Perceived trustworthiness of peer feedback	3.57	0.62	$t = -1.16$, $p = 0.24$
Perceived fairness of peer feedback	3.80	0.63	$t = 1.44$, $p = 0.15$

Table 16.5 Students' attitude towards peer feedback and peer feedback uptake in the argumentative essay writing in the context of argumentative essay writing in online education

Attitude towards peer feedback	Mean	SD	Results (* = Sig)
Perceived usefulness of peer feedback	3.63	0.67	$t = 2.01$, $p < 0.05$[*]
Perceived motivation of peer feedback	3.55	0.69	$t = -1.57$, $p = 0.11$
Perceived trustworthiness of peer feedback	3.57	0.62	$t = -0.79$, $p = 0.43$
Perceived fairness of peer feedback	3.80	0.63	$t = -0.76$, $p = 0.44$

(F(5, 73) = 3.31, p < 0.01, R^2 = 0.18). The adjusted R square value indicated that 18% of the attitude difference could be explained by these factors, but only two predictors (i.e. justification and constructive features) were significant.

The quality of the received peer feedback including constructive feature of feedback can predict students' perceived usefulness of peer feedback (F(5, 73) = 4.80, p < 0.01, R^2 = 0.25). The adjusted R square value indicated that 25% of the students' perceived usefulness difference could be explained by these factors, but only one predictor (i.e. constructive features) was significant.

The results also showed that the quality of the received peer feedback cannot predict students' perceived motivation of peer feedback (F(5, 73) = 1.29, p = 0.27).

However, it was found that the quality of the received peer feedback including justification and constructive features of feedback can predict students' perceived trustworthiness of peer feedback (F(5, 73) = 2.35, p < 0.05, R^2 = 0.14). The adjusted R square value indicated that 14% of the students' perceived trustworthiness difference could be explained by these factors, but only two predictors (i.e. justification and constructive features) were significant.

The results also showed that the quality of the received peer feedback including justification and constructive features of feedback can predict students' perceived fairness of peer feedback (F(5, 73) = 3.00, p < 0.05, R^2 = 0.17). The adjusted R square value indicated that 17% of the students' perceived fairness difference could be explained by these factors, but only two predictors (i.e. justification and constructive features) were significant (Table 16.6).

16.6 Discussion

16.6.1 Discussions for Findings of the RQ1

The findings revealed that students' attitude towards peer feedback had no predictive impacts on peer feedback performance. This means that the quality of the feedback that students provided was not influenced by their attitude towards peer feedback. Even though students showed a positive attitude towards peer feedback (Table 16.3), this finding showed that this attitude did not significantly affect students' peer feedback performance. To explain this finding, it can be argued that providing feedback is more a behavioral act and it is considered a skill that students should acquire through practice. Previous research has shown that practice is crucial for the development of peer feedback skills (Sluijsmans et al., 2002). Students who have more practice with peer feedback, the more likely are to develop expertise in making a critical evaluation of peers' essays to provide constructive points for improvements (Panadero, 2016). Researchers indicated that when students have more opportunities to practice peer feedback during essay writing in classes, they improve their ability how to give and make use of feedback (Chang et al., 2015; Liang & Tsai, 2010; Tsai et al., 2002; Wen & Tsai, 2006). In other words, the more training and preparation students had, the better they appeared to participate

Table 16.6 The effects of quality of the received peer feedback on students' attitude towards peer feedback in the argumentative essay writing

Attitude towards peer feedback	Peer feedback feature		Mean	SD	Results (* = Sig)
Students' attitude towards peer feedback	Affective		1.64	0.16	t = 0.08, p = 0.92
	Cognitive	Description	1.35	0.33	t = 0.31, p = 0.75
		Identification	0.65	0.31	t = −1.49, p = 0.14
		Justification	0.04	0.06	t = 2.01, p < 0.05*
	Constructive		0.77	0.38	t = 3.31, p < 0.01**
Students' perceived usefulness of peer feedback	Affective		1.64	0.16	t = −1.44, p = 0.15
	Cognitive	Description	1.35	0.33	t = 0.76, p = 0.44
		Identification	0.65	0.31	t = −0.72, p = 0.47
		Justification	0.04	0.06	t = 1.26, p = 0.21
	Constructive		0.77	0.38	t = 3.94, p < 0.01**
Students' perceived motivation/ enjoyment of peer feedback	Affective		1.64	0.16	t = 0.72, p = 0.47
	Cognitive	Description	1.35	0.33	t = −0.01, p = 0.99
		Identification	0.65	0.31	t = −1.36, p = 0.17
		Justification	0.04	0.6	t = 1.10, p = 0.27
	Constructive		0.77	0.38	t = 1.83, p = 0.07
Students' perceived trustworthiness of peer feedback	Affective		1.64	0.16	t = −0.30, p = 0.76
	Cognitive	Description	1.35	0.33	t = 0.95, p = 0.34
		Identification	0.65	0.31	t = −0.93, p = 0.35
		Justification	0.04	0.6	t = 1.91, p < 0.05*

(continued)

Table 16.6 (continued)

Attitude towards peer feedback	Peer feedback feature		Mean	SD	Results (* = Sig)
	Constructive		0.77	0.38	t = 2.28, p < 0.01*
Students' perceived fairness of peer feedback	Affective		1.64	0.16	t = 1.53, p = 0.12
	Cognitive	Description	1.35	0.33	t = −0.90, p = 0.37
		Identification	0.65	0.31	t = −1.85, p = 0.06
		Justification	0.04	0.6	t = 2.37, p < 0.05*
	Constructive		0.77	0.38	t = 2.58, p < 0.05*

in the peer assessment activity. This suggests that students' opinions toward their practice are influenced by this preparation (Hansen & Liu, 2005). Also, Liu and Lee (2013) showed that the students made valuable modifications to their work with the help of feedback from others, and most of the students had a positive impression of peer feedback after participating in multiple rounds of online peer assessment activities. Therefore, what can be said here is that the quality of provided feedback by peers depends more on their practices and experiences with peer feedback than their attitude towards peer feedback. Also, review publications showed that a number of the round of peer feedback (Chen et al., 2020; Liu & Lee, 2013), scripting (Noroozi et al., 2016), worked example and scripting (Latifi et al., 2020), collaborative team of reviewers (Mandala et al. 2018), structured peer feedback (Wang & Wu, 2008), anonymous (Basheti et al., 2010; Lane et al., 2018), synchronous discussion (Zheng et al., 2017), video annotation peer feedback (Lai, 2016), type of provided feedback (Noroozi et al., 2016), and peer feedback mode (peer ratings plus peer comments) (Chen et al., 2020; Hsia et al., 2016) affect on peer feedback performance. For example, Hsia et al., (2016) showed that the integration of both peer rating and peer comments is an effective approach that can meet the students' expectations and help them improve peer-feedback quality, and peer-scoring correctness as well as their willingness to participate in online learning activities. And, Mandala et al. (2018) showed that a collaborative team of reviewers produced higher quality feedback than did individual reviewers. Collaboration improved student engagement in the process. Zheng et al., (2017) showed that synchronous discussion can significantly improve the quality of affective and metacognitive peer feedback messages. Also, Lin (2018a, 2018b) showed that students in the anonymous group provided significantly more cognitive feedback (i.e., vague suggestions, extension). As a result, based on previous research, it can

be said that improving peer feedback performance is more influenced by different educational mechanisms and approaches than students' attitudes toward peer feedback.

16.6.2 Discussions for Findings of the RQ2

The findings revealed that in general students' attitude towards peer feedback did not predict their feedback uptake in the context of argumentative essay writing in online education. However, the perceived usefulness of peer feedback was a significant predictor for uptaking of peer feedback in argumentative essay writing. This means that if students feel that the received peer feedback is useful to improve their argumentative essay writing, they are willing to implement the received feedback in their essays. This finding, in general, is consistent with the findings of Huisman et al. (2018), Kaufman and Schunn (2011), and Strijbos et al. (2010). In particular, this finding is consistent with the findings of Misiejuk et al. (2020) and Mulder et al. (2014) where a relationship was found between the perceived usefulness of peer feedback and uptake of peer feedback. One reason to explain why the perceived usefulness of peer feedback can predict uptake of peer feedback could be related to the fact that when students feel that the received peer feedback can truly improve the quality of their work, then they will be in favor of taking those feedback comments seriously (Harks et al., 2014). This is supported by Misiejuk et al. (2020) study where they reported that students who found the feedback useful tended to be more accepting by acknowledging their errors, intending to revise their text, and praising its usefulness, while students who found the feedback less useful tended to be more defensive by expressing that they were confused about its meaning, critical towards its form and focus, and in disagreement with the claims. In other words, Students who perceived peer feedback useful were more likely to accept it by acknowledging their mistakes, indicating that they want to change their material, and/or appreciating the effectiveness of the peer feedback (Misiejuk et al., 2021; Noroozi et al., 2016). Therefore, teachers need to use strategies and mechanisms in the classroom to help students provide useful feedback. Learner attributes such as knowledge of the activity's goals, capacity to apply feedback criteria, and evaluation of the strengths and shortcomings of feedback (Sluijsmans et al., 2002) are all critical drivers of a peer feedback activity's success or failure. Future research could explore the impact of peer feedback activities on the skills and characteristics of students.

16.6.3 Discussions for Findings of the RQ3

The findings revealed that the quality of the received peer feedback can influence students' attitude towards peer feedback. This finding is consistent with the findings of Noroozi and Mulder (2017) and Wang et al. (2019). The findings showed that feedback that is justified by facts, example, various pieces of evidence as well

as suggestions for improvement, makes students more likely to trust that feedback and understand it more fairly. Students also find feedback that contains suggestions for improving work more useful. These findings are supported by Chen et al. (2009) and Lin (2018a, 2018b). One reason for such findings can be related to the fact that when students find the received feedback of high quality, they are more likely to uptake and use the received feedback in their essays (Noroozi et al., 2023; Wu & Schunn, 2020). Especially if the feedback is constructive and has suggestions for performance improvement (Valero-Haro et al., 2019a, b, 2022). If the received peer feedback is not constructive, and if peer feedback lacks quality features such as justification of problems in the essay and suggestions for improvement, students are more likely to ignore rather than accept and implement the feedback (Dominguez et al., 2012; Patchan et al., 2016). Because students did not perceive such feedback as useful. Geilen et al. (2010) found that students that have received justified recommendations outperformed in their revised work which is an indication for uptaking of received peer feedback. This depicts that if students explain and support their comments and feedback, their peers can better understand feedback and the issues raised in the feedback. This is in line with the prior studies that highlight the importance of high-quality features of feedback in the uptake of feedback (Winstone et al., 2016; Yuan & Kim, 2015).

16.7 Conclusion, Limitations, and Future Research

This study contributes to extending our knowledge on students' attitude towards peer feedback, peer feedback performance, and uptake. This study provides insights into how students with different attitudes perform and uptake peer feedback and how students with different qualities of received feedback perceived peer feedback in the context of argumentative essay writing in online education. This study revealed that the nature and quality of the received feedback plays a critical role in students' attitude towards peer feedback. This study suggests that for improving students' attitude towards peer feedback, students should be encouraged to provide high-quality feedback including features such as cognitive and constructive feedback with justified elaborations.

Although in this study we explored what features of the received feedback can predict students' attitude towards peer feedback in essay writing, we did not explore the role of provided feedback features in students' argumentative essay writing. It would be interesting to explore this in future studies and compare the effectiveness of the received and provided feedback features on students' attitude towards peer feedback. This can provide insights into the role of the assessor and assessee in the feedback process and its impacts on students' attitude towards peer feedback in the context of essay writing in higher education.

Since peer feedback also contains an internal process where students reflect on their own mind by critically reading and reflecting on peers' argumentative essay writing (Huisman et al., 2018), it is suggested that future research examine individual factors such as gender, culture, previous experiences and knowledge in relation

to students' attitudes towards peer feedback. Also, more research on peer feedback perceptions and responses to various aspects of peer feedback implementation is required.

In this study, students' prior knowledge and experiences regarding peer feedback and argumentative essay writing have not been investigated. The results of this study might have been influenced by this factor. Due to this reason, we should cautiously interpret the results of this study. For future studies, we suggest exploring the relationship between students' peer feedback performance on argumentative essay writing, their background knowledge and experiences with peer feedback, and their attitudes toward peer feedback. Another of the limitations of this study is the workload needed to provide and utilize peer feedback, so student attitudes may also depend upon the "fatigue" which can be experienced by students in peer assessment arrangements and their perception of trade-offs between benefits envisaged or gained and costs.

References

Alhomaidan, A. M. A. (2016). ESL writing students attitudes towards peer feedback activities. *International Journal of Research and Review, 3*(3), 74–88.

Allen, D., & Mills, A. (2016). The impact of second language proficiency in dyadic peer feedback. *Language Teaching Research, 20*(4), 498–513. https://doi.org/10.1177/1362168814561902

Altınay, Z. (2016). Evaluating peer learning and assessment in online collaborative learning environments. *Behaviour & Information Technology, 36*(3), 312–320. https://doi.org/10.1080/0144929X.2016.1232752

Azarnoosh, M. (2013). Peer assessment in an EFL context: Attitudes and friendship bias. *Language Testing in Asia, 3*(1), 1–10. https://doi.org/10.1186/2229-0443-3-11/TABLES/5

Basheti, I. A., Ryan, G., Woulfe, J., & Bartimote-Aufflick, K. (2010). Anonymous Peer Assessment of Medication Management Reviews. *American Journal of Pharmaceutical Education, 74*(5), 1–8. https://doi.org/10.5688/AJ740577

Bayat, M., Banihashem, S. K., & Noroozi, O. (2022). The effects of collaborative reasoning strategies on improving primary school students' argumentative decision-making skills. *The Journal of Educational Research, 115*(6), 349–358. https://doi.org/10.1080/00220671.2022.2155602

Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review*, 100489. https://doi.org/10.1016/j.edurev.2022.100489

Bordens, K, S. & Horowitz, I, A. (2008). *Social psychology* (3rd edn). Freeload Press.

Chang, C., & Lin, H.-C.K. (2019). Effects of a mobile-based peer-assessment approach on enhancing language-learners' oral proficiency. *Innovations in Education and Teaching International, 57*(6), 668–679. https://doi.org/10.1080/14703297.2019.1612264

Chen, I. C., Hwang, G. J., Lai, C. L., & Wang, W. C. (2020). From design to reflection: Effects of peer-scoring and comments on students' behavioral patterns and learning outcomes in musical theater performance. *Computers & Education, 150*, 103856. https://doi.org/10.1016/J.COMPEDU.2020.103856

Chen, N. S., Wei, C. W., Wu, K. T., & Uden, L. (2009). Effects of high level prompts and peer assessment on online learners' reflection levels. *Computers & Education, 52*(2), 283–291. https://doi.org/10.1016/J.COMPEDU.2008.08.007

Chou, T.-C.R. (2014). A scale of University students' attitudes toward e-learning on the moodle system. *International Journal of Online Pedagogy and Course Design, 4*(3), 49–65. https://doi.org/10.4018/IJOPCD.2014070104

Collimore, L. M., Paré, D. E., & Joordens, S. (2014). SWDYT: So what do you think? Canadian students' attitudes about peerScholar, an online peer-assessment tool. *Learning Environments Research 2014 18:1*, *18*(1), 33–45. https://doi.org/10.1007/S10984-014-9170-1.

Devon, J., Paterson, J. H., Moffat, D. C., & McCrae, J. (2015). Evaluation of student engagement with peer feedback based on student-generated MCQs. *ITALICS Innovations in Teaching and Learning in Information and Computer Sciences, 11*(1), 27–37. https://doi.org/10.11120/ITAL.2012.11010027.

Dominguez, C., Cruz, G., Maia, A., Pedrosa, D., & Grams, G. (2012). Online peer assessment: An exploratory case study in a higher education civil engineering course. In *2012 15th International Conference on Interactive Collaborative Learning, ICL 2012.* https://doi.org/10.1109/ICL.2012.6402220.

Donia, M. B. L., Mach, M., O'Neill, T. A., & Brutus, S. (2022). Student satisfaction with use of an online peer feedback system. *Assessment and Evaluation in Higher Education, 47*(2), 269–283. https://doi.org/10.1080/02602938.2021.1912286

Ekahitanond, V. (2013). Promoting university students' critical thinking skills through peer feedback activity in an online discussion forum. *Alberta Journal of Educational Research*, *59*(2), 247–265. https://doi.org/10.11575/AJER.V59I2.55617.

Fan, Y., & Xu, J. (2020). Exploring student engagement with peer feedback on L2 writing. *Journal of Second Language Writing, 50*, 100775. https://doi.org/10.1016/J.JSLW.2020.100775

Falchikov, N. (2005). *Improving through student involvement*. Routledge-Falmer.

Gagne, R. M., Wager, W. W., Golas, K. C., Keller, J. M., & Russell, J. D. (2005). Principles of instructional design, 5th edition. *Performance Improvement*, *44*(2), 44–46. https://doi.org/10.1002/PFI.4140440211.

Ge, Z. G. (2019). Exploring the effect of video feedback from unknown peers on e-learners' English-Chinese translation performance. *Computer Assisted Language Learning, 35*(1–2), 169–189. https://doi.org/10.1080/09588221.2019.1677721

Hansen, J. G., & Liu, J. (2005). Guiding principles for effective peer response. *ELT Journal, 59*(1), 31–38. https://doi.org/10.1093/elt/cci004

Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2014). The effects of feedback on achievement, interest and self-evaluation: The role of feedback's perceived usefulness. *Educational Psychology, 34*(3), 269–290. https://doi.org/10.1080/01443410.2013.785384

Hsia, L. H., Huang, I., & Hwang, G. J. (2016). Effects of different online peer-feedback approaches on students' performance skills, motivation and self-efficacy in a dance course. *Computers & Education, 96*, 55–71. https://doi.org/10.1016/J.COMPEDU.2016.02.004

Hu, G. (2005). Using peer review with Chinese ESL student writers. *Language Teaching Research, 9*(3), 321–342. https://doi.org/10.1191/1362168805LR169OA

Huisman, B., Saab, N., Van Driel, J., & Van Den Broek, P. (2018). Peer feedback on academic writing: Undergraduate students' peer feedback role, peer feedback perceptions and essay performance. *Assessment & Evaluation in Higher Education, 43*(6), 955–968. https://doi.org/10.1080/02602938.2018.1424318

Jiang, J., & Yu, Y. (2014). The effectiveness of internet-based peer feedback training on Chinese EFL college students' writing proficiency. *International Journal of Information and Communication Technology Education, 10*(3), 34–46. https://doi.org/10.4018/IJICTE.2014070103

Kaufman, J. H., & Schunn, C. D. (2011). Students' perceptions about peer assessment for writing: Their origin and impact on revision work. In *Instructional science* (Vol. 39, Issue 3, pp. 387–406). Springer. https://doi.org/10.1007/s11251-010-9133-6.

Kuyyogsuy, S. (2019). Students' attitudes toward peer feedback: Paving a way for students' English writing improvement. *English Language Teaching, 12*(7), 107. https://doi.org/10.5539/elt.v12n7p107

Kuo, F. C., Chen, J. M., Chu, H. C., Yang, K. H., & Chen, Y. H. (2017). A peer-assessment mobile Kung Fu education approach to improving students' affective performances. *International Journal of Distance Education Technologies, 15*(1), 1–14. https://doi.org/10.4018/IJDET.2017010101

Lai, C. Y. (2016). Training nursing students' communication skills with online video peer assessment. *Computers and Education, 97*, 21–30. https://doi.org/10.1016/j.compedu.2016.02.017

Lai, C. Y., Chen, L. J., Yen, Y. C., & Lin, K. Y. (2020). Impact of video annotation on undergraduate nursing students' communication performance and commenting behaviour during an online peer-assessment activity. *Australasian Journal of Educational Technology*, *36*(2), 71–88. https://doi.org/10.14742/AJET.4341.

Lane, J. N., Ankenman, B., & Iravani, S. (2018). Insight into gender differences in higher education: Evidence from peer reviews in an introductory STEM course. *Industrial Engineering and Management Sciences, 10*(4), 442–456. https://doi.org/10.1287/SERV.2018.0224

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174. https://doi.org/10.2307/2529310

Latifi, S., Noroozi, O., & Talaee, E. (2020). Worked example or scripting? Fostering students' online argumentative peer feedback, essay writing and learning. *Interactive Learning Environments, 1–15*. https://doi.org/10.1080/10494820.2020.1799032

Latifi, S., Noroozi, O., & Talaee, E. (2021). Peer feedback or peer feedforward? Enhancing students' argumentative peer learning processes and outcomes. *British Journal of Educational Technology.* https://doi.org/10.1111/bjet.13054

Latifi, S., & Noroozi, O. (2021). Supporting argumentative essay writing through an online supported peer-review script. *Innovations in Education and Teaching International, 58*(5), 501–511. https://doi.org/10.1080/14703297.2021.1961097

Latifi, S., Noroozi, O., Hatami, J., & Biemans, H. J. A. (2021). How does online peer feedback improve argumentative essay writing and learning? *Innovations in Education and Teaching International, 58*(2), 195–206. https://doi.org/10.1080/14703297.2019.1687005.

Lin, G. Y. (2018a). Anonymous versus identified peer assessment via a Facebook-based learning application: Effects on quality of peer feedback, perceived learning, perceived fairness, and attitude toward the system. *Computers & Education, 116*, 81–92. https://doi.org/10.1016/J.COMPEDU.2017.08.010

Lin, J.-W. (2018b). Effects of an online team project-based learning environment with group awareness and peer evaluation on socially shared regulation of learning and self-regulated learning, *37*(5), 445–461. https://doi.org/10.1080/0144929X.2018.1451558.

Lin, G.-Y. (2016). Effects that Facebook-based online peer assessment with micro-teaching videos can have on attitudes toward peer assessment and perceived learning from peer assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, *12*(9), 2295–2307. https://doi.org/10.12973/EURASIA.2016.1280A.

Lin, S. S. J., Liu, E. Z. F., & Yuan, S. M. (2001). Web-based peer assessment: Feedback for students with various thinking-styles. *Journal of Computer Assisted Learning, 17*(4), 420–432. https://doi.org/10.1046/J.0266-4909.2001.00198.X

Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279–290. https://doi.org/10.1080/13562510600680582

Liu, E. Z.-F., & Lee, C.-Y. (2013). Using peer feedback to improve learning via online peer assessment. *Turkish Online Journal of Educational Technology, 12*(1), 187–199.

Mandala, M. (2018). Impact of collaborative team peer review on the quality of feedback in engineering design projects. *International Journal of Engineering Education, 34*(4), 1299–1313.

McHugh, M. (2012). Interrater reliability: the Kappa statistic, *Biochem Med*, 22(3), 276–282. https://doi.org/10.11613/BM.2012.031.

Misiejuk, K., Wasson, B., & Egelandsdal, K. (2020). Using learning analytics to understand student perceptions of peer feedback. *Computers in Human Behavior, 117*,. https://doi.org/10.1016/j.chb.2020.106658

Miles, J. (2014). Tolerance and variance inflation factor. *Wiley Statsref: Statistics Reference Online.* https://doi.org/10.1002/9781118445112.stat06593

Mulder, R. A., Pearce, J. M., & Baik, C. (2014). Peer review in higher education: Student perceptions before and after participation. *Active Learning in Higher Education, 15*(2), 157–171. https://doi.org/10.1177/1469787414527391

Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science, 37*(4), 375–401. https://doi.org/10.1007/s11251-008-9053-x

Noroozi, O. (2018). Considering students' epistemic beliefs to facilitate their argumentative discourse and attitudinal change with a digital dialogue game. *Innovations in Education and Teaching International, 55*(3), 357–365. https://doi.org/10.1080/14703297.2016.1208112.

Noroozi, O. (2022). The role of students' epistemic beliefs for their argumentation performance in higher education. *Innovations in Education and Teaching International*. 1–12. https://doi.org/10.1080/14703297.2022.2092188.

Noroozi, O., Banihashem, S. K., Biemans, H. J. A., Smits, M., Vervoort, M. T. W., & Verbaan, C. (2023). Design, implementation, and evaluation of an online supported peer feedback module to enhance students' argumentative essay quality. *Education and Information Technologies*, 1–28. https://doi.org/10.1007/s10639-023-11683-y

Noroozi, O., Banihashem, S. K., Taghizadeh Kerman, N., Parvaneh Akhteh Khaneh, M., Babayi, M., Ashrafi, H., & Biemans, H. J. A. (2022). Gender differences in students' argumentative essay writing, peer review performance and uptake in online learning environments. *Interactive Learning Environments*, 1–15. https://doi.org/10.1080/10494820.2022.2034887.

Noroozi, O., Biemans, H., & Mulder, M. (2016). Relations between scripted online peer feedback processes and quality of written argumentative essay. *Internet and Higher Education, 31*, 20–31. https://doi.org/10.1016/j.iheduc.2016.05.002

Noroozi, O., & Hatami, J. (2019). The effects of online peer feedback and epistemic beliefs on students' argumentation-based learning. *Innovations in Education and Teaching International, 56*(5), 548–557. https://doi.org/10.1080/14703297.2018.1431143

Noroozi, O., & Mulder, M. (2017). Design and evaluation of a digital module with guided peer feedback for student learning biotechnology and molecular life sciences, attitudinal change, and satisfaction. *Biochemistry and Molecular Biology Education, 45*(1), 31–39. https://doi.org/10.1002/bmb.20981

Noroozi, O., Kirschner, P. A., Biemans, H. J. A., & Mulder, M. (2018). Promoting argumentation competence: Extending from first- to second-order scaffolding through adaptive fading. *Educational Psychology Review, 30*(1), 153–176. https://doi.org/10.1007/s10648-017-9400-z

Noroozi, O., Weinberger, A., Biemans, H.J.A., Mulder, M., & Chizari, M. (2012). Argumentation-based computer supported collaborative learning (ABCSCL). A systematic review and synthesis of fifteen years of research. *Educational Research Review, 7*(2), 79–106. https://doi.org/10.1016/j.edurev.2011.11.006.

Novakovich, J. (2016). Fostering critical thinking and reflection through blog-mediated peer feedback. *Journal of Computer Assisted Learning, 32*(1), 16–30. https://doi.org/10.1111/jcal.12114

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. *Handbook of Human and Social Conditions in Assessment, 247–266*,. https://doi.org/10.4324/9781315749136-22

Panadero, E., & Alonso-Tapia, J. (2013). Self-assessment: Theoretical and practical connotations. When it happens, how is it acquired and what to do to develop it in our students. *Electronic Journal of Research in Educational Psychology*, *11*(2), 551–576. https://doi.org/10.14204/ejrep.30.12200.

Paré, D. E., & Joordens, S. (2008). Peering into large lectures: Examining peer and expert mark agreement using peerScholar, an online peer assessment tool. *Journal of Computer Assisted Learning, 24*(6), 526–540. https://doi.org/10.1111/J.1365-2729.2008.00290.X

Patchan, M. M., Schunn, C. D., & Correnti, R. J. (2016). The nature of feedback: How peer feedback features affect students' implementation rate and quality of revisions. *Journal of Educational Psychology, 108*(8), 1098–1120. https://doi.org/10.1037/edu0000103

Prins, F. J., Sluijsmans, D. M. A., Kirschner, P. A., & Strijbos, J. W. (2010). Formative peer assessment in a CSCL environment: A case study. *Assessment & Education in Higher Education, 30*(4), 417–444. https://doi.org/10.1080/02602930500099219

Rahmany, R., Sadeghi, B., & Faramarzi, S. (2013). The effect of blogging on vocabulary enhancement and structural accuracy in an EFL context. *Theory and Practice in Language Studies*, 3(7). https://doi.org/10.4304/tpls.3.7.1288-1298.

Schreiber, J. B., Nora, A., Stage, F. S., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research., 99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338

Shang, H.-F. (2019). Exploring online peer feedback and automated corrective feedback on EFL writing performance. *Interactive Learning Environments, 1–13,.* https://doi.org/10.1080/10494820.2019.1629601

Sluijsmans, D. M. A., Brand-Gruwel, S., van Merriënboer, J. J. G., & Bastiaens, T. J. (2002). The training of peer assessment skills to promote the development of reflection skills in teacher education. *Studies in Educational Evaluation, 29*(1), 23–42. https://doi.org/10.1016/S0191-491X(03)90003-4

Strijbos, J. W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction, 20*(4), 291–303. https://doi.org/10.1016/j.learninstruc.2009.08.008

Taghizadeh Kerman, N., Noroozi, O., Banihashem, S. K., Karami, M. & Biemans, Harm. H. J. A. (2022). Online peer feedback patterns of success and failure in argumentative essay writing. *Interactive Learning Environments,* 1–10. https://doi.org/10.1080/10494820.2022.2093914

Tian, L., & Zhou, Y. (2020). Learner engagement with automated feedback, peer feedback and teacher feedback in an online EFL writing context. *System, 91*, 102247. https://doi.org/10.1016/j.system.2020.102247

Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing, 9*(2), 147–170. https://doi.org/10.1016/S1060-3743(00)00022-9

Topping, K. (2017). Peer assessment: Learning by judging and discussing the work of other learners. *Interdisciplinary Education and Psychology*, 1(1), 1–17. https://doi.org/10.31532/INTERDISCIPEDUCPSYCHOL.1.1.007.

Vu, T. T., & Dall'Alba, G. (2007). Students' experience of peer assessment in a professional course. *Assessment & Evaluation in Higher Education*, 32(5), 541–556. https://doi.org/10.1080/02602930601116896.

Valero-Haro, A., Noroozi, O., Biemans, H. J. A., & Mulder, M. (2019a). First-and second-order scaffolding of argumentation competence and domain-specific knowledge acquisition: a systematic review. *Technology, Pedagogy and Education, 28*(3), 329–345. https://doi.org/10.1080/1475939X.2019.1612772

Valero-Haro, A., Noroozi, O., Biemans, H. J. A., & Mulder, M. (2019b). The effects of an online learning environment with worked examples and peer feedback on students' argumentative essay writing and domain-specific knowledge acquisition in the field of biotechnology. *Journal of Biological Education, 53*(4), 390–398. https://doi.org/10.1080/00219266.2018.1472132

Valero-Haro, A, Noroozi, O., Biemans, H. J. A., & Mulder, M. (2022). Argumentation Competence: Students' argumentation knowledge, behavior and attitude and their relationships with domain-specific knowledge acquisition. *Journal of Constructivist Psychology, 35*(1), 123–145. https://doi.org/10.1080/10720537.2020.1734995

Wang, S. L., & Wu, P. Y. (2008). The role of feedback and self-efficacy on web-based learning: The social cognitive perspective. *Computers & Education, 51*(4), 1589–1598. https://doi.org/10.1016/J.COMPEDU.2008.03.004

Wang, J., Gao, R., Guo, X., & Liu, J. (2019). Factors associated with students' attitude change in online peer assessment—a mixed methods study in a graduate-level course. *Assessment & Evaluation in Higher Education, 45*(5), 714–727. https://doi.org/10.1080/02602938.2019.1693493

Wen, M. L., & Tsai, C.-C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education 2006 51:1, 51*(1), 27–44. https://doi.org/10.1007/S10734-004-6375-8.

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2016). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist, 52*(1), 17–37. https://doi.org/10.1080/00461520.2016.1207538

Wu, Y., & Schunn, C. D. (2020). When peers agree, do students listen? The central role of feedback quality and feedback frequency in determining uptake of feedback. *Contemporary Educational Psychology, 62*, 101897. https://doi.org/10.1016/j.cedpsych.2020.101897

Wu, Y., & Schunn, C. D. (2021). From plans to actions: A process model for why feedback features influence feedback implementation. *Instructional Science 2021 49:3, 49*(3), 365–394. https://doi.org/10.1007/S11251-021-09546-5.

Wu, Z. (2019). Lower English proficiency means poorer feedback performance? A mixed-methods study. *Assessing Writing, 41*, 14–24. https://doi.org/10.1016/J.ASW.2019.05.001

Yang, Y. F. (2016). Transforming and constructing academic knowledge through online peer feedback in summary writing. *Computer Assisted Language Learning, 29*(4), 683–702. https://doi.org/10.1080/09588221.2015.1016440

Yuan, J., & Kim, C. (2015). Effective feedback design using free technologies. *Journal of Educational Computing Research, 52*(3), 408–434. https://doi.org/10.1177/0735633115571929

Zhang, H., Song, W., Shen, S., & Huang, R. (2014). The effects of blog-mediated peer feedback on learners' motivation, collaboration, and course satisfaction in a second language writing course. *Australasian Journal of Educational Technology*, *30*(6), 670–685. https://doi.org/10.14742/AJET.860.

Zhao, H. (2018). Exploring tertiary English as a Foreign Language writing tutors' perceptions of the appropriateness of peer assessment for writing. *Assessment & Evaluation in Higher Education, 43*(7), 1133–1145. https://doi.org/10.1080/02602938.2018.1434610

Zheng, L., Cui, P., Li, X., & Huang, R. (2017). Synchronous discussion between assessors and assessees in web-based peer assessment: Impact on writing performance, feedback quality, meta-cognitive awareness and self-efficacy. *Assessment & Evaluation in Higher Education, 43*(3), 500–514. https://doi.org/10.1080/02602938.2017.1370533

Zhu, Q., & Carless, D. (2018). Dialogue within peer feedback processes: Clarification and negotiation of meaning. *Higher Education Research and Development, 37*(4), 883–897. https://doi.org/10.1080/07294360.2018.1446417

Zou, Y., Schunn, C. D., Wang, Y., & Zhang, F. (2017). Student attitudes that predict participation in peer assessment. *Assessment & Evaluation in Higher Education, 43*(5), 800–811. https://doi.org/10.1080/02602938.2017.1409872

How Do Lower-Secondary Students Exercise Agency During Formative Peer Assessment?

17

Laura Ketonen, Pasi Nieminen, and Markus Hähkiöniemi

17.1 Introduction

Assessment and feedback have traditionally been the provinces of teachers, but that approach is changing (Boud, 2014). In higher education, researchers have emphasized the need for students to actively participate in feedback processes (Carless & Boud, 2018; Dawson et al., 2019; Winstone et al., 2017). In secondary education, same trend can be seen in the schools' and researchers' interest in self-assessment and peer assessment. However, the research on peer assessment has significant gaps. Studies have been largely concerned with its cognitive side (Panadero et al., 2018) and have paid little attention to sociocultural perspectives (Panadero, 2016; van Gennip et al., 2009). This is a serious gap, given that the social dimension is an elementary part of peer assessment (Panadero, 2016).

This study considers the social dimension of peer assessment by employing the notion of students' agency. Agency can be defined as a "socioculturally mediated capacity to act" (Ahearn, 2001), which signifies that agency is considered as an interplay between individuals and their environment. Peer assessment promotes student agency by giving students the formal roles of assessor and assessee. However, assigning formal roles is only the beginning, since agency is coproduced in classroom environments as interplays between the teacher and students and among students themselves (Charteris & Thomas, 2016). According to research, students will not necessarily embrace their active roles. They may question their ability as assessors (Mok, 2011) or feel uncomfortable criticizing their peers' work, even

---

L. Ketonen (✉) · P. Nieminen · M. Hähkiöniemi
Department of Teacher Education, University of Jyväskylä, Jyväskylän, Finland
e-mail: laura.k.ketonen@jyu.fi

though criticism is officially sanctioned (Foley, 2013; Harris & Brown, 2013). Additionally, students may resist their peers' feedback (Foley, 2013; Panadero, 2016), worry about the effects of peer assessment on their social relationships (Harris & Brown, 2013), and let relationships influence the feedback they provide (Panadero & Jonsson, 2013). The findings reveal that social and cultural features play roles in peer assessment and that students' agency does not always take constructive forms but can be practiced in harmful ways (Harris et al., 2018).

In general, the literature on assessment and agency is in its infancy (Nieminen & Tuohilampi, 2020), particularly that focused on peer assessment and agency. Even though student agency is considered a necessary ingredient in formative assessment (Harris et al., 2018) and a rationale for using it includes the fact that it increases students' active role in assessment and learning (Boud, 2014; Braund & DeLuca, 2018; Panadero, 2016; Topping, 2009), little is known about the forms of agency that students exercise during peer assessment. In the present study, we advance the understanding of the topic by exploring lower-secondary students' forms of agency when formative peer assessment was repeatedly used in their science studies.

17.1.1 Formative Peer Assessment

Peer assessment has many variations. It can be used for summative or formative purposes, and it can be operationalized face to face or at a distance between individuals, pairs, or groups (Topping, 2013). This study only considered the formative purpose, which is the advancement of students' learning; it did not focus on measurements of student learning, which is the purpose of the summative approach. According to Black and Wiliam (2009), the same assessment instruments (e.g., tests, projects, self-assessment, and peer assessment) can be used formatively and summatively, meaning the function of the assessment defines its type, not the assessment itself. Peer assessment is formative when its goal is helping students understand intentions and the criteria for success as well as activating them as instructional resources for one another (Black & William, 2009). Teachers are responsible for creating a learning environment, articulating that the aim of peer assessment is to advance learning, and delivering instructions that support that intention (Black & William, 2009). Topping (2013) defined peer assessment as "an arrangement for classmates to consider the level, value, or worth of the products or outcomes of learning of their equal-status peers" (p. 395), and argued that both receiving and providing feedback are beneficial. Hence, the strategy of activating students as instructional resources for each other (Black & Wiliam, 2009) entails two separate goals: guiding them to be instructional resources for others (assessor's objective) and guiding them to use others as instructional resources (assessee's objective).

Researchers have reached the consensus that peer assessment requires training (Sluijsmans, 2002; Topping, 2009; van Zundert et al., 2010). Peer assessment

comprises several phases: developing original work, providing feedback, receiving feedback, and revising one's own work (). Acting as an assessor or assessee requires diverse skills that vary depending on the form of peer assessment. Assessors need to understand their responsible position as providers of feedback (Panadero, 2016), understand the assessment criteria, judge the performance of a peer, and formulate constructive feedback (Sluijsmans, 2002). Assessees need to be able to judge feedback, manage affect, and act on feedback (Carless & Boud, 2018). These skills are needed in peer assessment, and they can be further developed by practicing it (Ketonen et al., 2020a, 2020b).

17.1.2 Agency in Peer Assessment

Depending on the research tradition, the concept of agency has different definitions and emphases (Eteläpelto et al., 2013). In this section, we discuss three aspects of agency for which researchers' views diverge, and we clarify our stance toward them. The first concerns the ontological dimension of agency—more precisely, the extent to which agency is considered an individual versus a social attribute. At one end of the spectrum, agency is construed as an individual's autonomous, rational actions; at the other, it is construed as shaped by structural factors, even to the point that the existence of agency is questioned (Eteläpelto et al., 2013). We take a middle ground in this research, following Billett's (2006) theorization of the "relational interdependence" between individual and social agency. Billett (2006) suggests that individuals practice agency by choosing which problems and social suggestions they engage in and by regulating their level of engagement when participating in these undertakings. Hence, individual agency has a social origin, but it is not socially determined. When considering schools, students' levels of agency may vary even within a single classroom because there are various microenvironments for participation (e.g., the whole class, a small group, or pairs) and social roles (e.g., colleague, peer assessor, or friend) offering different kinds of social suggestions and problems to engage in.

Temporality is another aspect in which the views of agency diverge. Some approaches do not consider the temporal element of agency, whereas others do (Eteläpelto et al., 2013). In this study, as presented by Emirbayer and Mische (1998), we construe students' agency as a composite of past, present, and future, which are all relevant when practicing peer assessment. First, students' agency in the classroom builds on experience. Even the first time they engage in peer assessment, students bring their experiences of learning, being assessed, and correcting and advising others. Their former ways of participating have developed patterns of agency that create expectations for their participation (Gresalfi et al., 2009). Second, agency is derived from imagined outcomes of action. Students visualize the consequences of complimenting and criticizing their peers, and apart from how those choices' influence learning, they weigh their influence on their relationships with their peers and teacher. Third, agency is enacted in the present, which is not necessarily a straightforward process. For example, the act of providing feedback

during formative peer assessment might demand considerations of the assessed work, the assessment criteria, one's own capacity as an assessor, the teacher's expectations, and the social norms and relationships in the classroom.

The third aspect of agency that has different emphases in the literature is the requirement of transformation. Some researchers highlight the transformative nature of agency and define it as transcendence of established patterns (Kumpulainen et al., 2018; Matusov, 2011; for transformative agency see Sannino, 2015). Others suggest that exercising agency does not require bringing about a change (Biesta & Teddler, 2007); instead, adaptive behaviors, such as seeking help, self-regulating, and setting goals, are also forms of agency. From such a perspective, students never lack agency completely; rather, they can always exercise at least a minimal amount of agency via either compliance or resistance (Gresalfi et al., 2009). Furthermore, forms of agency cannot be categorized as good or bad. For example, resisting authorship (Matusov et al., 2016) is neither unambiguously right nor wrong but rather reflective of students' interpretations of tasks, environments, and their positions within those environments. Students can use either compliance or resistance as a means to achieve their goals. For example, by working hard and utilizing feedback, students can pursue learning or good grades; conversely, by rejecting feedback and purposefully underperforming, they can protect the ego from criticism or manage an overwhelming workload (Harris et al., 2018). In this study, we take the stance that transformative behavior is not the only way of exercising agency; rather, agency can also be seen in adaptive behavior.

17.1.3 Study Objective

In this study, we explored students' actions during formative peer assessment during science studies in a lower-secondary school. The objective was to advance understandings of students' agency during peer assessment. The research questions are set out below.

1. What forms of agency do students exercise during formative peer assessment?
2. How do students exercise agency in different positions that peer assessment offers them with respect to other students?

17.2   Method

17.2.1 Participants and Procedure

This study was carried out in a standard classroom in a typical lower-secondary school in Finland; most of the students were born in Finland, and there was a roughly equal share of boys and girls. As to participants, we selected four seventh grade students (mean age: 13 years). The criteria for selection were that
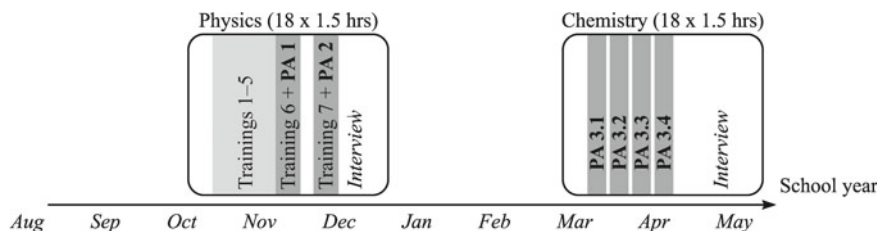
Fig. 17.1 Timeline of the training sessions, peer assessments (here abbreviated to "PA"), and interviews

they had participated in all the types of peer assessments and a majority of the peer assessment training sessions during the study and did not seem to struggle with motivation or have particular challenges with learning. All four students' attitudes toward science learning and peer assessment appeared positive. We made the choice to examine the role of agency when students were willing to participate in peer assessment. If a student struggled significantly with learning, the potential reasons for that disengagement or misbehavior were wide ranging and thus not only related to peer assessment. In this exploratory study, we sought to exclude such factors.

Two participants, Rachel and Maggie, worked in the same group of four students, while Lucas and Nathan in another group of four students. Students studied physics for half their fall semester and chemistry for half their spring semester (Fig. 17.1). These were their first physics and chemistry courses and were taught by a subject teacher. Students first received training in peer assessment and then performed assessment three different ways, twice in physics and four times in chemistry.

The training included class discussions and written tasks. Over six weeks, there were seven 10- to 45-min sessions, which are further described in Table 17.1 and in (Ketonen, 2021). The overarching message of the training sessions was that peer assessment was for learning. The assessors' goal was to help classmates progress, and the assessees' goals were to respect peers' assistance and use feedback if possible.

The peer assessments had different organizational forms and objectives, which are further explained in Table 17.2 and further in (Ketonen, 2021).

17.2.2 Research Design and Data

Since the goal of the study was to explore what happens in a classroom during peer assessment, a naturalistic study setting and a qualitative case study design were chosen. The data consisted of audio recordings of students' classroom discussions, written peer feedback, written work, student interviews, and the researcher's field notes. The first author observed the participants and made field notes during most of the 36 lessons of 1.5 h each. At the beginning of each lesson, she placed audio

Table 17.1 Peer assessment training activities

Type of task	Task	Goal
1. Individual reflection and class discussion	Understand the role of assessment at school	Understand and distinguish between different assessment aims
2. Individual reflection and class discussion	Understand what kind of assessments are helpful	Understand that (a) feedback is for learning and (b) the quality of feedback matters
3. Written task and class discussion	Create assessment criteria	Understand what constitutes good assessment criteria for inquiry tasks
4. Written task and class discussion	Assess the work of a fictional student using previously created assessment criteria	(a) Further understand what makes good assessment criteria and (b) practice comparing work to criteria
5. Self-assessment	Assess one's own inquiry task	Practice comparing work to criteria
6. Class discussion and PA1	Understand the qualities of good feedback and peer assessment principles	(a) Learn what kind of feedback is helpful for others, (b) understand that peer assessment is for helping each other move forward, and (c) practice peer assessment
7. Class discussion	Understand how to react to feedback	(a) Learn to evaluate feedback and (b) acquire strategies to deal with it

recorders on the tables of each student pair. The recorders captured students' conversations during the lessons. Students' written work included original and revised versions of their peer-assessed work and written peer feedback. All students were individually interviewed after PA2 and PA3. In semi-structured interviews that took from 6 to 11 min, their original work, revised work, and received feedback were used as bases for the conversations. An average interview followed the chronology of the peer assessment: it started with questions about the student's perception of their original work, turned to their consideration of the assessed work and the feedback they provided to others, continued to the feedback they had received, and the changes they were considering as a result of the peer assessment. If a student led the conversation to other topics, these were discussed, and this sometimes changed the order of the interview elements.

17.2.3 Analysis

The interviews and class discussions during peer assessments were transcribed, while written feedback and work were scanned, and each student's data were compiled in chronological order. Peer feedback sheets described what kind of feedback

**Table 17.2** The tasks and implementation of peer assessments

	Arrangement	Assessed task	Assessment criteria	After the peer assessment	Time used
PA1	Groups circulated through-out the class-room and assessed the other groups' plans	Technology project plan made by groups: planning and modeling a rover that moved on its own	Assessment criteria presented on a whiteboard. Feedback written on different color Post-it Notes	Groups were able to modify their plans right after the peer assessment and during the building of the rover	Task: 40 min Assessment: 30 min Revisions: 10 min
PA2	Each student assessed anonymously another's lab report (pairing planned by teacher and researcher)	Individually made inquiry report: defining the speed of a rover	Assessment criteria with a three-choice rubric and an opportunity to provide written comments for each criterion	Students had an opportunity to revise their report before returning it to the teacher for summative assessment	Task: 3 h Assessment: 45 min Revisions: 45 min
PA3.1 PA3.2 PA3.3 PA3.4	Working as "group members" in pairs or trios, students assessed each other's lab work	Chemistry inquiry conducted in pairs (four different inquiries and peer assessments, e.g., examining which substances dissolve in water)	Assessment criteria with a three-choice rubric and a requirement to provide at least one positive comment at the end	Students marked their agreement with the feedback by circling the most suitable of four options. The feedback sheets were returned to the teacher or the researcher	Task: 4 × 15–30 min Assessment: 4 × 2–10 min

students had provided and received, and their preliminary and final work provided information on how they went about their revisions. Students' conversations in their groups and working pairs provided additional information related to providing and using feedback. Students' written work, classroom discussions, and written feedback were used as primary data sources, and interviews and observations were

used to complement and explain the findings. The first researcher, who had taught at the school for some time, was responsible for the coding. She read the files carefully multiple times. Then she analyzed the data using a thematic analysis (Braun & Clarke, 2006). She marked data extracts containing information about students' agency during peer assessment and labelled them with descriptive codes. Gresalfi et al.'s (2009) description of agency was used to identify extracts relevant to our study purpose: "An individual's agency refers to the way in which he or she acts, or refrains from acting, and the way in which her or his action contributes to the joint action of the group in which he or she is participating" (p. 53). A unit of analysis was one student's data in one peer assessment in one role, for example, all of Student 1's data while they were an assessor during PA1. Since individual students' ways of participating in certain peer assessments were intertwined and partly explained each other, the researcher first coded all students' data from PA1 and proceeded chronologically through the remaining assessments.

After coding the whole data set, the researcher retrieved and examined data extracts and codes, developed preliminary categories of student forms of agency, and wrote descriptions for each. When developing the categories, she compared the codes to data extracts in each one to consider their internal consistency, and then she compared the categories with each other to examine their distinctiveness and coherence, which led to changes to the codes. After, she recoded the data set with new codes. To test, discuss, and develop the coding and to support the entire process, we used peer debriefing (Onwuegsbuzie & Leech, 2007). The second and third researchers, who were not involved in the field work, asked critical questions and explained their views of the first researcher's codes and categories. The iterative process of coding, comparing codes and categories, and revising them continued until it did not produce any changes. Then the researcher named the categories and wrote the final category descriptions. The categories' relationships were elaborated with a thematic map (Braun & Clarke, 2006). We noticed that the forms of agency were related to the positions of assessor, assessee, and group member (Fig. 17.2). Given that agency is a relational and context-dependent construct, this finding was significant. In the last phase, we examined individual students' ways of exercising agency in each of these three positions, thus answering research question 2.

Fig. 17.2 Students' ways of exercising agency as group members
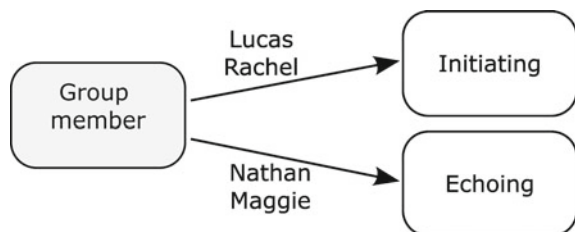
17.3 Results

In this study, we explored the forms of agency that students exercised during formative peer assessment in different positions with respect to other students. We found 12 forms of agency that related to three positions. These are presented in Table 17.3. As group members, students were on an equal footing with their peers; as assessors, they were in an advisory position; and as assessees, they were in receiving position. In some cases, students worked in several positions concurrently, such as when they acted as assessors in a group. The finding revealed that students conducting peer assessment act in various positions in relation to each other and the way their agency presents itself depends on that position.

In the following three sections, we introduce and compare the forms of agency within the position in which each form was exercised.

17.3.1 Exercising Agency as a Group Member

As group members, students exercised agency by initiating or echoing ideas. In their respective groups, Nathan and Maggie echoed others' ideas, while Lucas and Rachel were active in introducing original ideas, whether providing or receiving peer feedback (Fig. 17.2). Lucas and Rachel expressed their ideas without difficulty, whereas Nathan and Maggie hesitated to make suggestions even when they built on others' ideas.

The following example of initiating is from PA1, in which Rachel and Maggie, and their two other groupmates, Mia and Tara, assessed another group's work. The assessed task was a plan for a mobile rover that could be built with available resources (see Ketonen, 2021 for more information). Below, the exchange begins with the group's first comment on the other group's plan.[1]

1	Tara:	*(Quoting other group's plan)* "Rubber band, catapult …
2		tail end."
3	Rachel:	It does not say what [materials] they need there
		. *(Discussion unrelated to peer assessment and physics)*
14	Tara:	Once nothing else is needed,

---

[1] Transcription notations are described immediately below.

( )   Description of context or nonverbal speech.
" "   Reading text.
—   Comment was interrupted.
…   Words were cut out.
[ ]   Clarifies the reference.

The line numbers are group specific and start from "1" after each transition (i.e., each change to new work [PA1] or a change in assessor and assessee [PA1, PA2]).

15		they can write what they need
16	Rachel:	Yeah, they can write there what they need,
17		and then they can draw it, like, from below—
18		the bottom
19		Like, from the bottom angle
20	Tara:	From the bottom
21	Maggie:	And from above,
22		not just from the side

Right after seeing the other group's plan, Rachel argued that they had not listed what material they would use to build their rover (2), and later, she proposed the need to draw the model from different angles (17, 19). At that point, she put forward two ideas that were echoed by other group members, thus practicing the initiating form of agency. Tara repeated Rachel's first (14) and second (20) ideas and Maggie elaborated on Rachel's second idea (21–22).

The difficulty of initiating new ideas became apparent when students assessed the next group's work. Rachel—the former initiator—was concentrating on another issue, and the other three group members were left with the job of providing feedback. First, they took considerable time comparing Maggie's handwriting to that of the assessees. When they turned to assessing, the conversation below took place.

37	Maggie:	There could have been… *(silence)*
38	Tara:	If …
39		nothing. *(Silence)*
40	Maggie:	This could have been better planned
41		Like, they could write what everyone brings or something
42	Mia:	But it's there
43	Maggie:	Right. *(Silence)*
44	Maggie:	This could have been drawn from several
45		different angles
46	Mia:	Yeah, right
47	Maggie:	How should I formulate it?
48	Mia:	Could you have done it from several angles?

The students tried to provide feedback, but they either did not come up with any ideas or did not feel comfortable expressing them (37–39). After a while, Maggie raised Rachel's previous idea of listing the required material (41). After Mia pointed out that the material were already listed (42), Maggie took a moment to rethink and suggested Rachel's other previous idea about drawing the rover from different angles (44, 45). This was accepted (46) and written on a Post-it Note. This excerpt demonstrates that even when assessors are willing to provide

**Table 17.3**   The forms of agency and the positions in which they were exercised

Forms of agency	Role	Description
Initiating	Group member	This form entails participating in group discussions by introducing new ideas or opinions
Echoing	Group member	This form entails participating in group discussions by repeating or elaborating on other members' ideas or opinions but not presenting ones' own
Judging work	Assessor	This form entails analyzing other students' work and performance and providing feedback that contains criticism and/or suggestions for improvement (the provision of positive comments is not included in this form, as providing superficial, positive feedback is a common way of avoiding engagement as an assessor)
Avoiding criticism	Assessor	Avoiding criticism is a concealed form of agency, as it cannot be observed by looking only at what students do but at what they do not do. Engaging in assessment but repetitively providing only positive feedback is considered as the avoidance of criticism
Seeking help	Assessor and Assessee	This form entails seeking help with assessor or assessee tasks
Appraising feedback	Assessee	This form entails examining feedback in order to judge its quality or validity and appreciating one's own judgement
Rejecting feedback	Assessee	This form entails rejecting feedback after mentioning a reason for the rejection
Revising work	Assessee	This form entails revising one's own work after assessing others and receiving feedback. It shows engagement with the task, but it does not necessarily lead to an improvement of the work
Avoiding revision	Assessee	This form entails receiving critical feedback and, while not rejecting it (i.e., mentioning a reason for not following feedback), not revising the work

feedback, new ideas may not be put forward, which constitutes a lack of initiation. Having an initiator in a group supported others in assessing their peers.

17.3.2   Exercising Agency as an Assessor

As the previous section showed that initiating ideas was challenging to some students, one may wonder how they exercised agency, when they were supposed

Fig. 17.3 Students' ways of exercising agency as assessors
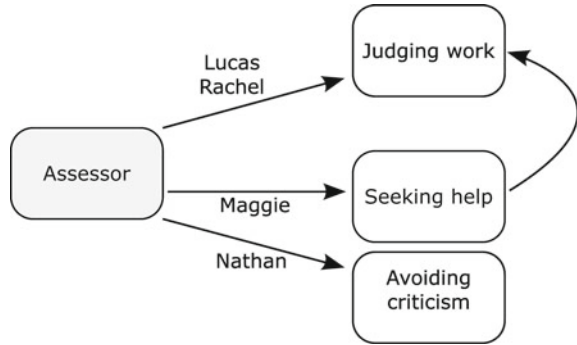
to work as individual assessors during PA2. Students' diverse ways of exercising agency are presented in Fig. 17.3. The assessed task was a lab report about determining the speed of the previously planned and built rover. The inquiry was conducted in groups, but the reports were individually written. Perhaps unsurprisingly, Rachel and Lucas—who, much like Rachel, had initiated ideas in his group—assessed their peers' work without difficulty. They concentrated on assessing for a moderate amount of time and provided both confirming and correcting comments.

Maggie, who had echoed others' ideas during PA1, accomplished the task by seeking help from peers and the teacher. At first, she spent time criticizing the assessee's handwriting. She interpreted handwriting with Tara, asked Rachel for help, and then asked the teacher for help. Since in our opinion, the handwriting looked rather clear, we interpreted her criticism of it as an excuse to avoid the task and seek help with assessing. The teacher came to Maggie, calmly read and discussed the work with her, and encouraged her to write down her thoughts. This helped Maggie complete half the criteria, after which she again criticized the handwriting and asked Rachel, the researcher, and the teacher for help. Maggie was persistent in her attempts to provide feedback, and after a considerable struggle, she provided one encouraging comment and one suggestion for improvement. Maggie's struggles became even more evident later, and this is depicted in the extract below, in which she was assessing her friend Tara's lab performance.

34	Maggie:	Tara, sorry, I can't mark
35		that you correctly used the burner
36	Tara:	But I did
37	Maggie:	You blew on it
38	Rachel:	Yes, you did *(laughs)*,
39		and you, like, blew it out
40	Tara:	I'm sorry, but my fingers almost burned
41	Rachel:	I wouldn't have *(indistinguishable)* shaken
42	Maggie:	"Your working was thoughtful and controlled."
43	Tara:	Really?
44	Rachel:	*(Laughs)*

45	Maggie:	I feel bad
46		*(Asks the teacher)* Can you mark two options
47		if it's in between?
48		Like, it seems that it's actually neither
49	Teacher:	Either/or, preferably
50	Maggie:	But I think these, like,
51		Tara did otherwise good,
52		but there was, like, one tiny thing

During the inquiry, Tara lit the gas burner and blew the match out in front of it, blowing the burner out as well. The gas kept leaking out, spreading its distinctive smell across the classroom, and this caused minor chaos. When assessing Tara's work, Maggie, quite justifiably, commented that she could not rate Tara's burner use as "excellent" but only "good" (34, 35). Notable is that even though the assessment was formative, Maggie felt uncomfortable rating Tara as "good," and in addition to explaining her decision to her (34–35, 37) and being supported by Rachel (38–39), she asked the teacher for help. For Maggie, providing criticism was laborious, but she was persistent, and with other's support, she managed to do it. It was evident that Maggie did not lack the attitude (she strove to give a solid judgement) or skills (she knew that Tara's performance was less than excellent) but rather the agency to put her knowledge into action. By seeking second and third opinions, she gained agency that enabled her to provide feedback she considered justified.

Nathan was Lucas' group member and had echoed his ideas during PA1. Nathan seemed to struggle with providing feedback too, but his solution was the opposite of Maggie's. Assessing the lab report took Nathan a substantial amount of time. On the recording, the sound of Nathan writing and erasing can be heard long after Lucas was done. He wound up marking each criterion with the best option ("Everything is ok") and provided only one written comment: "What you needed was clearly explained." It is possible that Nathan did not notice any of the several shortcomings in the lab report, but this seems unlikely, as providing trivial feedback took him such a long time. We suggest that Nathan noticed some problems and spent time thinking about how to react to them. During the year of practicing peer assessment, Nathan consistently avoided criticizing others' work and independently gave only the highest marks and compliments. During PA3, when pairs assessed each other's lab work, Lucas even corrected Nathan several times for providing him with feedback that was too positive. Apparently, providing criticism was not a satisfactory option for Nathan. Unlike Maggie, he did not seek help with assessing but kept on providing overly positive feedback.

17.4 Exercising Agency as an Assessee

Students had diverse ways of exercising agency as assessees; these are presented in Fig. 17.4 and followed by examples.

Lucas, who initiated constructive ideas during PA1, was a rapid reviser. After receiving feedback about his lab report (PA2), he read the feedback, quickly judged it, rejected part of its useful aspects, and made small-scale improvements to his lab report. Rachel, who also initiated ideas during PA1, operated in a similar way, but she was more careful and did not reject useful feedback. It seemed that both Lucas and Rachel experienced both providing and receiving feedback as appropriate and uncomplicated.

Nathan, who echoed ideas during PA1, appeared generally open to feedback and committed to using it for improvement. In PA2 (revising own lab report), Nathan's immediate reaction after receiving the feedback was to ask the teacher's opinion: "Teacher! Should I revise this?" He waited until the teacher came to see him. Nathan wanted to know whether the feedback was valid, which the teacher confirmed. They discussed the issue for a considerable amount of time, and after, Nathan revised his work independently, managing to improve it.

Maggie, who also echoed ideas during PA1, took the opposite approach to a similar situation. In the excerpt below, she reacts to corrective feedback.

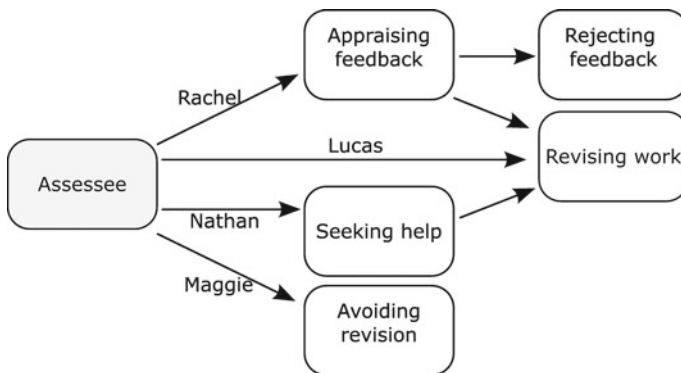| 35 | Maggie: | Look, I made a few mistakes in the text |
36		It doesn't matter. Small mistakes
37	Tara:	What mistakes do you mean?
38	Maggie:	That hypothesis was about the distance,
39		not speed
40		I guessed the distance here
41		and not the speed,
42		how fast it moved

Fig. 17.4 Students' ways of exercising agency as assessees

43	Tara:	Yeah
44	Maggie:	But it does not matter
45		I'm surprised that I was this good

The feedback Maggie received—"the hypotheses was about distance, not speed"—could have been used to improve her work. She could have changed her hypotheses, or comment the mistake in her revisions. Maggie affirmed that she had made a mistake (35) but characterized it as a small one (36) that did not matter (36, 44) and instead concentrated on her general performance (45). She bypassed the criticism by congratulating herself, did not return to the topic, and did not revise her work. One could construe that she was unresponsive, but her explanation in an interview suggested otherwise.

| Researcher: | Okay, okay. Were you motivated to make revisions since you considered that [work] was not so super, not quite superlative? |
| Maggie: | It has always been really hard for me to accomplish something *(indistinguishable)* because I always think that I'm stupid and if I do something, it always seems bad. So it's hard to begin to improve |

Maggie said that a lack of confidence in her own abilities held her back from making revisions. Under the surface of congratulating herself, she was uncertain of her skills. It appears that she did not have the agency to undertake her revisions.

17.5 Discussion

This study explored students' actions during formative peer assessment and contributed to the literature by enhancing awareness of their agency during the exercise. We identified nine forms of agency (initiating, echoing, judging work, avoiding criticism, seeking help, appraising feedback, rejecting feedback, revising work, avoiding revision) in three roles that peer assessment provided (group member, assessor, assessee).

Closer investigation of students' interaction revealed that peer assessment challenged the students unevenly. Throughout each assessment, Lucas and Rachel practiced the agencies of initiating, judging work, and appraising feedback without difficulty, while Nathan and Maggie exercised those agencies only when they received support. When working in groups, Nathan and Maggie participated only by echoing other students' suggestions. When acting individually as an assessor, Nathan consistently avoided criticizing others by providing only positive feedback. Maggie was persistent in her aspiration to provide valid critical feedback, but she needed help to do so. By asking support from other students and the teacher, she gained the agency of judging other students' work. As an individual assessee, Nathan needed help appraising feedback before he revised his work, whereas Maggie did not seek help and refrained from revising her work. The findings show that

even all students were placed in the same classroom, undertaking the same task of assessing their peers, their challenges were unequal. We explain this by referring to the notions that experience builds agency (Emirbayer & Mische, 1998) and that students' previous actions create expectations for their participation (Gresalfi et al., 2009). For students who generally initiate ideas, are active, and advise others, the assessor role is more familiar and their feedback more likely to be accepted by classmates. For them, peer assessment is a straightforward task. For others, assessing may require acting outside their accustomed role.

The operationalization of peer assessment, especially whether it was conducted individually or in groups, influenced the social suggestions that were available for students (Billett, 2006) and thereby the agencies that students exercised. When assessing and receiving feedback in a group (PA1), the agencies of initiating and echoing were practiced. Working in a group allowed struggling students to receive subtle support when assessing and receiving feedback, as they were able to echo other students' initiatives. Individual peer assessments (PA2, PA3) forced students to be responsible for themselves, which created the need to ask for and offer help and caused some students to avoid the task.

The findings are highly significant for the practice of peer assessment. The requirement of agency sheds light on the effects of students' individual attributes on peer assessment, which is thus far an unexplored area (Panadero, 2016), and it addresses the need to ensure appropriate support for students' agency when they are requested to exit their comfort zones as assessors and assessees. With an understanding of the requirements of agency, teachers can be better equipped to provide support. They can listen to, confirm, and endorse students' thoughts, guide them to discuss the issue with their friends, or open the subject to a classroom discussion. The finding also highlights the need to be careful with the use of unsupported individual peer assessment, since it can be highly stressful for students who struggle with their agency. Moreover, if teachers are not aware of the requirement of agency, they may misinterpret students' misbehavior or underperformance as stemming from a lack of skills or a negative attitude. If teachers respond by assisting students in the accomplishment of their peer assessment tasks instead of strengthening their agency, they can weaken that agency by indicating students are not capable of acting as assessors and assessees on their own.

The finding about the requirement of agency has implications for peer assessment training. Peer assessment provides a platform for students to exercise agency in assessment and learning by guiding them to act in various, and potentially new positions in relation to other students. Hence, peer assessment can advance democracy in the classroom not just between teachers and students (Gielen et al., 2010) but also by sharing among everyone the responsibility to help others. However, helping others, especially in the form of criticizing and advising, cannot be taken for granted. Nineteenth century German pedagogue Froebel (1887) argued that "the purpose of teaching and instruction is to bring ever more *out* of man rather than to put more and more *into* him" (p. 279, emphasis in the original). The quote applies to students' agency by describing a new aspect of peer assessment training. We agree with the necessity of providing students with knowledge, such as

understanding the qualities of constructive feedback (Tasker & Herrenkohl, 2016), skills, such as judging received feedback (Carless & Boud, 2018), and attitudes, such as their sense of responsibility when assessing (Panadero, 2016). However, students' agency also needs to be encouraged. As agency is seen as an interplay between an individual and their environment, training requires investing not just in individuals but also in their relationships and the culture of the classroom. We consider this a significant area for future research: how does peer assessment assist in transcending the classroom's fixed patterns and strengthening students' agency?

Technology can support the development of student agency (Marín et al., 2020). Technological environments are commonly used in peer assessment (see Fu et al., 2019). They are convenient for sharing work, matching students for peer assessment, and providing feedback, and they allow students to assess each other either anonymously or by name. The findings of this study suggest that the organization of peer assessment should be examined from the perspective of agency, which also concerns technological environments. First, how do different kinds of technological environments support students' agency? Anonymity may provide students different kinds of social suggestions, a new role in which to operate, and hence a lower threshold at which to participate actively. Interaction has been suggested as an element that deepens the learning process of peer assessment, while anonymity is a feature that diminishes that interaction (Panadero, 2016). Technology allows students to interact anonymously, and the pros and cons of such arrangements for students' agency are worth examination. Important aspect to consider is that students' agency must be supported in technological environments, one way or another. Students should not be left alone with their devices but be allowed to interact with each other and the teacher and to seek help during peer assessment. Technological environments can be interactive and allow students to seek help (e.g. Tasker & Herrenkohl, 2017). We consider the diverse ways of supporting students' agency during peer assessment—both face to face and online—to be an important topic for future research.

This was a case study of four students, two of whom appeared to struggle with their agency during peer assessment, whereas the other two did not. The finding was consistent throughout all types of peer assessment during the school year. The merit of our study is that it introduces and demonstrates the requirement of agency during peer assessment. However, by selecting students who did not have apparent cognitive or motivational challenges, we have dealt with only part of the spectrum of forms of agency during peer assessment, and further research about the topic is needed. For example, what role does students' social position in class play alongside their subject skills or confidence in mastering them, and what kinds of environments support students' agency? Potentially, different types of challenges with agency require different types of support.

Our study showed that the concept of agency is useful in unveiling and explaining peer assessment's underlying dynamics. Awareness of how students' agency plays a role in peer assessment is significant to educators and researchers. Students' reluctance or inability to help their peers or accept help do not necessarily

stem from a lack of knowledge, skills, or attitude but can be suggestive of their difficulties in exercising agency.

References

Ahearn, L. M. (2001). Language and agency. *Annual Review of Anthropology, 30*, 109–137. https://doi.org/10.1146/annurev.anthro.30.1.109

Biesta, G., & Tedder, M. (2007). Agency and learning in the lifecourse: Towards an ecological perspective. *Studies in the Education of Adults, 39*(2), 132–149. https://doi.org/10.1080/02660830.2007.11661545

Billett, S. (2006). Relational interdependence between social and individual agency in work and working life. *Mind, Culture and Activity, 13*(1), 53–69. https://doi.org/10.1207/s15327884mca1301_5

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Boud, D. (2014). Shifting views of assessment: From secret teachers' business to sustaining learning. In C. Kreber, C. Anderson, N. Entwistle, & J. MacArthur (Eds.), *Advances and innovations in university assessment and feedback* (pp. 13–31). Edinburgh University Press. https://doi.org/10.3366/edinburgh/9780748694549.003.0002

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braund, H., & DeLuca, C. (2018). Elementary students as active agents in their learning: An empirical study of the connections between assessment practices and student metacognition. *Australian Educational Researcher, 45*, 68–85. https://doi.org/10.1007/s13384-018-0265-z

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Charteris, J., & Thomas, E. (2017). Uncovering "unwelcome truths" through student voice: Teacher inquiry into agency and student assessment literacy. *Teaching Education, 28*(2), 162–177. https://doi.org/10.1080/10476210.2016.1229291

Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D., & Molloy, E. (2019). What makes for effective feedback: Staff and student perspectives. *Assessment & Evaluation in Higher Education, 44*(1), 25–36. https://doi.org/10.1080/02602938.2018.1467877

Emirbayer, M., & Mische, A. (1998). What is agency? *American Journal of Sociology, 103*(4), 962–1023. https://doi.org/10.1086/231294

Eteläpelto, A., Vähäsantanen, K., Hökkä, P., & Paloniemi, S. (2013). What is agency? Conceptualizing professional agency at work. *Educational Research Review, 10*, 45–65. https://doi.org/10.1016/j.edurev.2013.05.001

Foley, S. (2013). Student views of peer assessment at the International School of Lausanne. *Journal of Research in International Education, 12*(3), 201–213. https://doi.org/10.1177/1475240913509766

Froebel, F. (1887). *The education of man.* (W.N. Hailmann, Trans.). Appleton. (Original work published 1826).

Fu, Q.-K., Lin, C.-J., & Hwang, G.-J. (2019). Research trends and applications of technology-supported peer assessment: A review of selected journal publications from 2007 to 2016. *Journal of Computers in Education, 6*, 191–213. https://doi.org/10.1007/s40692-019-00131-x

Gielen, S., Tops, L., Dochy, F., Onghena, P., & Smeets, S. (2010). A comparative study of peer and teacher feedback and of various peer feedback forms in a secondary school writing curriculum. *British Educational Research Journal, 36*(1), 143–162. https://doi.org/10.1080/01411920902894070

Gresalfi, M., Martin, T., Hand, V., & Greeno, J. (2009). Constructing competence: An analysis of student participation in the activity systems of mathematics classrooms. *Educational Studies in Mathematics, 70*, 49–70. https://doi.org/10.1007/s10649-008-9141-5

Harris, L., & Brown, G. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies in teachers' implementation. *Teaching and Teacher Education, 36*, 101–111. https://doi.org/10.1016/j.tate.2013.07.008

Harris, L. R., Brown, G. T., & Dargusch, J. (2018). Not playing the game: Student assessment resistance as a form of agency. *The Australian Educational Researcher, 45*(1), 125–140. https://doi.org/10.1007/s13384-018-0264-0

Ketonen, L. (2021). Exploring interconnections between student peer assessment, feedback literacy and agency. [Doctoral dissertation, University of Jyväskylä].

Ketonen, L., Hähkiöniemi, M., Nieminen, P., & Viiri, J. (2020a). Pathways through peer assessment: Implementing peer assessment in a lower secondary physics classroom. *International Journal of Science and Mathematics Education, 18,* 1465–1484. https://jyx.jyu.fi/handle/123456789/76143

Ketonen, L., Nieminen, P., & Hähkiöniemi, M. (2020b). The development of secondary students' feedback literacy: Peer assessment as an intervention. *The Journal of Educational Research, 113*(6), 407–417. https://doi.org/10.1080/00220671.2020.1835794

Kumpulainen, K., Kajamaa, A., & Rajala, A. (2018). Understanding educational change: Agency-structure dynamics in a novel design and making environment. *Digital Education Review, 33*, 26–38. https://doi.org/10.1344/der.2018.33.26-38

Marín, V. I., Benito, de Benito, B., & Darder, A. (2020). Technology-enhanced learning for student agency in higher education: a systematic literature review. *Interaction Design and Architecture(s) Journal, 45*, 15–49. https://doi.org/10.55612/s-5002-045-001

Matusov, E. (2011). Authorial teaching and learning. In E. J. White & M. Peters (Eds.), *Bakhtinian pedagogy: Opportunities and challenges for research, policy and practice in education across the globe* (pp. 21–46). Peter Lang Publishers.

Matusov, E., von Duyke, K., & Kayumova, S. (2016). Mapping concepts of agency in educational contexts. *Integrative Psychological and Behavioral Science, 50*(3), 420–446. https://doi.org/10.1007/s12124-015-9336-0

Mok, J. (2010). A case study of students' perceptions of peer assessment in Hong Kong. *ELT Journal, 65*(3), 230–239. https://doi.org/10.1093/elt/ccq062

Nieminen, J. H., & Tuohilampi, L. (2020). "Finally studying for myself": Examining student agency in summative and formative self-assessment models. *Assessment & Evaluation in Higher Education, 45*(7), 1031–1045. https://doi.org/10.1080/02602938.2020.1720595

Onwuegbuzie, A. J., & Leech, N. L. (2007). Validity and qualitative research: An oxymoron? *Quality & Quantity: International Journal of Methodology, 41*(2), 233–249. https://doi.org/10.1007/s11135-006-9000-3

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review, 9*, 129–144. https://doi.org/10.1016/j.edurev.2013.01.002

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of social and human conditions in assessment* (pp. 247–266). Routledge.

Panadero, E., Jonsson, A., & Alqassab, M. (2018). Providing formative peer feedback: What do we know? In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 409–431). Cambridge University Press.

Sannino, A. (2015). The emergence of transformative agency and double stimulation: Activity-based studies in the Vygotskian tradition. *Learning, Culture and Social Interaction, 4*, 1–3. https://doi.org/10.1016/j.lcsi.2014.07.001

Sluijsmans, D. M. A. (2002). *Student involvement in assessment: The training of peer assessment skills.* (Unpublished doctoral dissertation). Open University of the Netherlands, Heerlen. https://doi.org/10.1016/s0191-491x(03)90003-4

Tasker, T., & Herrenkohl, L. (2016). Using peer feedback to improve students' scientific inquiry. *Journal of Science Teacher Education, 27*(1), 35–59. https://doi.org/10.1007/s10972-016-9454-7

Topping, K. J. (2009). Peer assessment. *Theory into Practice, 48*(1), 20–27. https://doi.org/10.1080/00405840802577569

Topping, K. J. (2013). Peers as a source of formative and summative assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 394–412). Sage Publications, Inc. https://doi.org/10.4135/9781452218649.n22

Winstone, N. E., Nash, R. A., Rowntree, J., & Parker, M. (2017). "It'd be useful, but I wouldn't use it": Barriers to university students' feedback seeking and recipience. *Studies in Higher Education, 42*(11), 2026–2041. https://doi.org/10.1080/03075079.2015.1130032

van Gennip, N., Segers, M., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review, 4*, 41–54. https://doi.org/10.1016/j.edurev.2008.11.002

van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20*(4), 270–279. https://doi.org/10.1016/j.learninstruc.2009.08.004