

ics for Learning and Teaching


kou · Dimitra Vinatsella ·
mpson ·
apamitsiou ·
nakos · Dirk Ifenthaler

ditional Analytics achers and l Leaders

 Springer

Advances in Analytics for Learning and Teaching

Series Editors

Dirk Ifenthaler , Learning, Design and Technology
University of Mannheim
Mannheim, Baden-Württemberg, Germany

David Gibson, Teaching and Learning
Curtin University
Bentley, WA, Australia

This book series highlights the latest developments of analytics for learning and teaching as well as providing an arena for the further development of this rapidly developing field. It provides insight into the emerging paradigms, frameworks, methods, and processes of managing change to better facilitate organizational transformation toward implementation of educational data mining and learning analytics. The series accepts monographs and edited volumes focusing on the above-mentioned scope, and covers a number of subjects. Titles in the series *Advances in Analytics for Learning and Teaching* look at education from K-12 through higher education, as well as vocational, business, and health education. The series also is interested in teaching, learning, and instructional design and organization as well as data analytics and technology adoption.

Sofia Mougiakou • Dimitra Vinatsella
Demetrios Sampson • Zacharoula Papamitsiou
Michail Giannakos • Dirk Ifenthaler

Educational Data Analytics for Teachers and School Leaders

 Springer

Sofia Mougiakou
University of Piraeus
Piraeus, Greece

Dimitra Vinatsella
University of Piraeus
Piraeus, Greece

Demetrios Sampson 
Department of Digital Systems
University of Piraeus
Piraeus, Greece

Zacharoula Papamitsiou
Department of Technology Management
SINTEF Digital
Trondheim, Norway

Michail Giannakos
Dept of Computer Science
Norwegian Univ of Science & Technology
Trondheim, Norway

Dirk Ifenthaler 
Learning, Design and Technology
University of Mannheim
Mannheim, Baden-Württemberg, Germany



This book is an open access publication.

ISSN 2662-2122 ISSN 2662-2130 (electronic)
Advances in Analytics for Learning and Teaching
ISBN 978-3-031-15265-8 ISBN 978-3-031-15266-5 (eBook)
<https://doi.org/10.1007/978-3-031-15266-5>

© The Editor(s) (if applicable) and The Author(s) 2023

Open Access This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Educational Data Analytics (EDA) have been attributed with significant benefits for enhancing on-demand personalised educational support of individual learners as well as reflective course (re)design for achieving more authentic teaching, learning and assessment experiences integrated into real work-oriented tasks. As a result, most Course Management Systems are now incorporating Educational Data Analytics tools. However, these tools are still not widely used because of the low **Educational Data Literacy (EDL)** competences of the education professionals that could be using them (i.e. K12 teachers adopting the flipped classroom model in their teaching and school leaders leveraging educational data to support decision-making). Furthermore, online learning environments and education data-driven practice and assessment raise challenges such as ethical issues and implications, especially in terms of privacy, security of data and informed consent that should be addressed via transparent and well-defined ethical policies and codes of practices.

Particularly nowadays, as the Covid-19 pandemic continues to unfold around the world, **emergency remote teaching** has become the new reality for school education around the world. Subsequently, **educational data**, which is the rich data footprint that students generate through their interactions in digital learning environments, has increased exponentially. This unprecedented crisis has brought to the forefront the urgent demand for all education professionals, including schoolteachers and leaders, to reinvent their teaching and learning environments. Educational Data Analytics (EDA) has been identified as a key enabler to seize the opportunities – through the use of educational data generated during **teaching and learning** (including **assessment**) – for better supporting learners in online and blended courses. To this end, the “upskilling imperative” of comprehensive EDL competences has been recognised as an immediate need to support **creative, flexible and inclusive education and training** in the long term, highlighting the importance for educators to ground decisions based on data and evidence aiming to boost the **effectiveness** and the **efficiency** of the **education systems**, even beyond the periods of emergency education such as public health crises or natural disasters.

This book aims to support the development of core competences for Educational Data Analytics of online and blended teaching and learning.

It combines:

- **Theoretical knowledge** on core issues related to collecting, analysing, interpreting and using educational data, including ethics and privacy
- **Questions and teaching materials/learning activities** as quiz tests of multiple types of questions, added after each section, related to the topic studied or the video(s) referenced. These activities reproduce real-life contexts by using a suitable use case scenario (storytelling), encouraging learners to link theory with practice.
- **Self-assessed assignments** enabling learners to apply their attained knowledge and acquired competences on EDL. Each self-assessed assignment is a real-life scenario activity (e.g. based on a use case), using a rubric across three proficiency levels and an exemplary solution rating. The evaluation of the outcomes is done by the learners as self-assessment, using a rubric which includes the criteria that each response should meet and guidelines to assess themselves.
- **Activities/practice questions** as reflective tasks to offer deeper understanding of the educational data field.

It targets:

- **E-learning professionals** (such as *instructional designers* and *e-tutors*) of online and blended courses
- **School leaders** and **teachers** engaged in blended (using the flipped classroom model) and online (during COVID-19 crisis and beyond) teaching and learning
- **Higher education students** (undergraduates and postgraduates)

The content of this book has been developed within the action **Learn2Analyze – An Academia-Industry Knowledge Alliance for enhancing Online Training Professionals’ (Instructional Designers and e-Trainers) Competences in Educational Data Analytics**, which is co-funded by the European Commission through the **Erasmus+ Programme** of the European Union (Cooperation for innovation and the exchange of good practices – **Knowledge Alliances**, Agreement n. 2017-2733 / 001-001, Project No 588067-EPP-1-2017-1-EL-EPPKA2-KA). The European Commission’s support for the production of this publication does not constitute an endorsement of the contents, which reflects the views only of the authors, and the Commission will not be held responsible for any use which may be made of the information contained therein. More information about the project is available at www.learn2analyze.eu.

By studying this book, you will:

- *Know* where to locate useful educational data in different data sources and *understand* their limitations
- *Know* the basics for managing educational data to make them useful, *understand* relevant methods and *be able to use* relevant tools
- *Know* the basics for organising, analysing, interpreting and presenting learner-generated data within their learning context, *understand* relevant learning analytics methods, and *be able to use* relevant learning analytics tools

- *Know* the basics for analysing and interpreting educational data to facilitate educational decision-making, including course and curricula design, *understand* relevant teaching analytics methods, and *be able to use* relevant teaching analytics tools
- *Understand* issues related with educational data ethics and privacy.

The learning objectives of this text book cover the set of competences anticipated by the **Learn2Analyse Educational Data Literacy competence framework (L2A--EDL-CP)** – see Section 1.2.5. Each chapter is developed to support a given set of L2A-EDL-CP related learning objectives which are clearly stated at the beginning of the chapter.

Piraeus, Greece

Trondheim, Norway

Mannheim, Germany

Sofia Mougiakou
Dimitra Vinatsella
Demetrios Sampson
Zacharoula Papamitsiou
Michail Giannakos
Dirk Ifenthaler

Contents

1	Online and Blended Teaching and Learning Supported by Educational Data	1
1.1	Introduction and Scope	1
1.1.1	Scope	1
1.1.2	Chapter Learning Objectives	1
1.1.3	Introduction	2
1.2	Educational Data as a Key Success Factor for Online and Blended Teaching and Learning.	3
1.2.1	Educational Data for Data-Driven Decision Making.	3
1.2.2	Why Educational Data Is Important for Online and Blended Teaching and Learning?.	9
1.2.3	How Educational Data Can Help Instructional Designers and e-Tutors of Online Courses?	12
1.2.4	How Educational Data Can Help School Teachers of Blended Courses?	19
1.2.5	The Learn2Analyze Educational Data Literacy Competence Framework.	26
1.3	Data Is Everywhere (Educational Data Collection).	33
1.3.1	Posing Questions and Identifying Appropriate Educational Data	33
1.3.2	Matching Appropriate Educational Data with Data Sources	41
1.3.3	Combining Data from Different Educational Data Sources	46
1.4	Concluding Self-Assessed Assignment	49
1.4.1	Introduction	49
1.4.2	Step 1. Real Life Scenario	50
1.4.3	Step 2. Getting Familiar with the Assessment Rubric	50
1.4.4	Step 3. Prepare Your Answer	52

1.4.5	Step 4. Review a Sample Solution	52
1.4.6	Step 5. Self-Evaluate Your Answer	53
	References.	54
2	Adding Value and Ethical Principles to Educational Data	59
2.1	Introduction and Scope	59
2.1.1	Scope	59
2.1.2	Chapter Learning Objectives	59
2.1.3	Introduction	60
2.2	Adding Value to Educational Datasets (Educational Data Management)	61
2.2.1	Making Data Tidy (Data Cleaning)	61
2.2.2	Data to Describe Data (Metadata)	72
2.2.3	The Significance of Data Curation.	79
2.2.4	Storage Issues for Preserving Educational Data	83
2.3	Educational Data Ethics.	92
2.3.1	Informed Consent.	92
2.3.2	Sensitive Educational Data Protection.	99
2.4	Concluding Self-Assessed Assignment	111
2.4.1	Introduction	111
2.4.2	Step 1. Real Life Scenario	112
2.4.3	Step 2. Getting Familiar with the Assessment Rubric	112
2.4.4	Step 3. Prepare Your Answer	117
2.4.5	Step 4. Review a Sample Solution.	118
2.4.6	Step 5. Self-Evaluate Your Answer	122
	References.	124
3	Learning Analytics	131
3.1	Introduction and Scope	131
3.1.1	Scope	131
3.1.2	Chapter Learning Objectives	131
3.1.3	Introduction	132
3.2	Using Learner-Generated Data and Learning Context for Extracting Learning Analytics	133
3.2.1	Definition and Objectives of Learning Analytics.	133
3.2.2	Measurements as Indicators of Learners' Current Learning States	135
3.2.3	Limitations and Data Quality Issues of Learners' Data Measurements in Open and Blended Courses	141
3.2.4	Ethical Treatment of Learner-Generated Data and Measurements	147
3.3	Analyzing Data and Presenting Learning Analytics	154
3.3.1	Methods for Analyzing the Learner-Generated Data and the Measurements Over Them	154
3.3.2	Presentation Methods for Reporting on Learner Data Analytics	161

- 3.4 Interpreting Learning Analytics and Inferring Learning Changes 173
 - 3.4.1 Making Sense of Learners’ Data Analytics and Analysis Results 173
 - 3.4.2 Explaining the Data Analysis Results in an Educationally Meaningful Manner to Understand Learners and the Environment they Learn In 178
- 3.5 Concluding Self-Assessed Assignment 179
 - 3.5.1 Introduction 179
 - 3.5.2 Step 1. Real Life Scenario 179
 - 3.5.3 Step 2. Getting Familiar with the Assessment Rubric 179
 - 3.5.4 Step 3. Prepare Your Answer 182
 - 3.5.5 Step 4. Review a Sample Solution 182
 - 3.5.6 Step 5. Self-Evaluate Your Answer 184
- References 185
- 4 Teaching Analytics 189**
 - 4.1 Introduction and Scope 189
 - 4.1.1 Scope 189
 - 4.1.2 Chapter Learning Objectives 189
 - 4.1.3 Introduction 190
 - 4.2 Data Sources for Supporting Teaching Analytics 191
 - 4.2.1 Learning and Teaching 191
 - 4.2.2 Design of Learning Environments 191
 - 4.2.3 Learning Design 193
 - 4.2.4 TPACK Model 193
 - 4.3 Data Sources Within the Instructional Design Process 197
 - 4.3.1 Broadening the Perspective for Data-Driven Education ... 197
 - 4.3.2 Data Sources Within a Holistic Analytics Framework. 198
 - 4.3.3 Sources of Learner Data. 199
 - 4.3.4 Sources of Online Learning Data. 199
 - 4.4 Key Concepts of Data Quality and Limitations of Data Meaningfulness 202
 - 4.4.1 Data Quality in Educational Contexts 202
 - 4.4.2 Core Dimensions of Data Quality 202
 - 4.4.3 Dimensions of Educational Data Quality 203
 - 4.4.4 Data Quality Problems 203
 - 4.5 Data Ethics and Privacy Principles for Teaching Analytics 206
 - 4.5.1 Ethical and Privacy Challenges Associated with the Application of Educational Data Analytics 206
 - 4.5.2 Privacy in the Digital World 207
 - 4.5.3 Ethical Principles 207
 - 4.6 Identify Issues of Authorship, Ownership, Data Access and Data-Sharing 210
 - 4.6.1 Privacy Calculus 210
 - 4.6.2 Educational Data Analytics Benefits 210

- 4.6.3 Data for Instructional Support 213
- 4.6.4 Data for Instructional Support 213
- 4.6.5 Data Privacy in Productive Systems 214
- 4.6.6 Case Study: Curtin Challenge I 216
- 4.6.7 Case Study: Curtin Challenge II 217
- 4.7 Applying and Communicating Educational Data and Analytics
Findings 219
 - 4.7.1 Adaptive Learning Technologies 219
 - 4.7.2 Automated and Semi-Automated Interventions. 220
 - 4.7.3 Instructional Design Principles for Adaptivity 222
- 4.8 Methodologies for Improving Learning and Teaching Processes
as Well as Curricula 223
 - 4.8.1 Creating Interventions in Classroom Settings 223
 - 4.8.2 Educational Design Research at a Glance 224
 - 4.8.3 Designing Model-Based Learning Environments 226
- 4.9 Concluding Self-Assessed Assignment 228
 - 4.9.1 Introduction 228
 - 4.9.2 Step 1. Real Life Scenario 228
 - 4.9.3 Step 2. Prepare Your Answer 228
 - 4.9.4 Step 3. Exemplary Sample Solution 229
 - 4.9.5 Step 4. Rubrics for Assessing Your Work 230
 - 4.9.6 Step 5. Self-Evaluate Your Answer 230
- References. 231
- Appendix 237**

Chapter 1

Online and Blended Teaching and Learning Supported by Educational Data



1.1 Introduction and Scope

1.1.1 Scope

The goals on this chapter are to:

- introduce the concept of **educational data as a key success factor for online and blended teaching and learning**,
- present the **Learn2Analyze framework for educational data literacy competences**, and
- discuss the fundamentals of **educational data collection**.

1.1.2 Chapter Learning Objectives

This chapter learning objectives	Learn2Analyze Educational data literacy Competence profile
Learn how educational data can support successful online and blended courses	1.1
Understand the importance of data-driven decision making to continuously improve the online and blended teaching and learning	5.1
Recognise the value of educational data literacy to make data-informed reflections on the design and delivery of instruction	5.2
Know the different types of educational data in online and blended courses	1.1
Know the different educational data sources related to core elements of e-learning environments	1.1

1.1.3 Introduction

Data is identified as one of the **key enablers for driving change** in the twenty-first century.

In the context of online education, **learners are leaving behind a rich data footprint** throughout the course of their study. As a result, the existing **educational data about learners, their learning and the environments in which they learn**, has exponentially increased.

Educators can grasp the great opportunities offered by educational data and the potential provided by data analytics technologies, to gain powerful insights and develop new ways of achieving excellence in both teaching and learning.

Educational data can reveal insights about the course design and teaching practice that might not be recognised otherwise. Moreover, through educational data analysis, tutors can have a **holistic view of their learners' past, present and likely future**, develop a **deep understanding of their learners' activities, behaviour and preferences**. As a result, they can target accordingly their teaching and learning interventions to provide the learners with a **personalised learning experience** and **better feedback**, and help them meet their educational goals.

Educational Data-Driven Decision Making (DDDM) can be a useful tool for reflecting on the teaching practices and improving the teaching and learning outcomes. For effective DDDM, educators need to be able to **identify, collect, combine, analyse, interpret and effectively act upon all types of educational data from diverse sources**.

Educational Data Literacy for all Education Professionals (such as instructional designers, teachers and tutors of online and blended courses) is now recognized internationally as a **key set of competences** and a **strong competitive advantage** to get the **best results in online and blended teaching and learning**.

However, emerging advancements related to the use of data-driven design and delivery of online and blended learning courses, exploiting **Educational Data Analytics are not yet thoroughly addressed by existing competence frameworks for education professionals**.

To this end, the **Learn2Analyze project has developed a comprehensive proposal for an Educational Data Literacy Competence Framework to enhance existing competence frameworks with new Educational Data Literacy competences**. The Learn2Analyze Educational Data Literacy Competence Framework comprises of **6 competence dimensions and 17 competence statements**.

The first competence dimension of this framework refers to **Data Collection**. Since educational data comes from a variety of sources in diverse formats, the effective Data Collection is considered as an essential competence that educators need to acquire and a prerequisite for this continuous process of evaluation, reflection and improvement.

It is fundamental for educators to distinguish the different types of Educational Data in Online and Blended courses, to identify the Educational Data Sources related to core elements of e-learning environments and to access and gather the

appropriate educational data by combining data from different sources and in different formats, avoiding systematic errors induced from the data collection process employed.

1.2 Educational Data as a Key Success Factor for Online and Blended Teaching and Learning

1.2.1 Educational Data for Data-Driven Decision Making

As described in the “[What is Big Data and how does it work?](#)” video (in the useful video resources), data is identified as one of the key factors driving change in the twenty-first century. Commonly referred to as the ‘*data revolution*’, the ‘*era of big data*’, or more simply ‘*big data*’, the term is used to describe the tremendous increase in the amounts of data we generate in all aspects of our lives. Big Data can bring big possibilities and thus create big expectations (Shacklock, 2016).

“Big Data gives you the ability to achieve superior value from analytics on data at higher volumes, velocities, varieties or veracities”. This claim is summarized in Fig. 1.1, based on the infographic “Extracting business value from the 4 V’s of big data” by IBM (2019).

Volume	The size of available data has been growing at an exponential rate. “With higher data volumes, you can take a more holistic view of your subject’s past, present and likely future”.
Velocity	Data streams are created at an unprecedented speed. “At higher data velocities, you can ground your decisions in continuously updated, real-time data”.
Variety	Data comes in all types of formats. “With broader varieties of data, you can have a more nuanced view of the matter at hand”.
Veracity	Data veracity is not only how accurate or truthful a data set may be, but also how trustworthy the data source, type, and processing of it is. “As data veracity improves, you can be confident that you’re working with the truest, cleanest, most consistent data”.

This video by [intel](#) (see the useful video resources “[Big Data's Making Education Smarter](#)”) explains further how big data can make education smarter.

In the context of online education, learners are leaving behind a rich data footprint throughout the course of their study. Educational data comprises a wide range of datasets about learners, their learning and the environments in which they learn, stored in various sources. We will focus on and discuss in detail different types of educational data in the next topic of this chapter (Shacklock, 2016).

Educational data and data analytics technologies can support us in developing a better understanding of our learners’ activities, behaviour and preferences, by identifying patterns and trends in the data that, in turn, can help us predict possible future outcomes and take actions for improving the learners’ experience in our courses.

Fig. 1.1 The 4 V's of big data

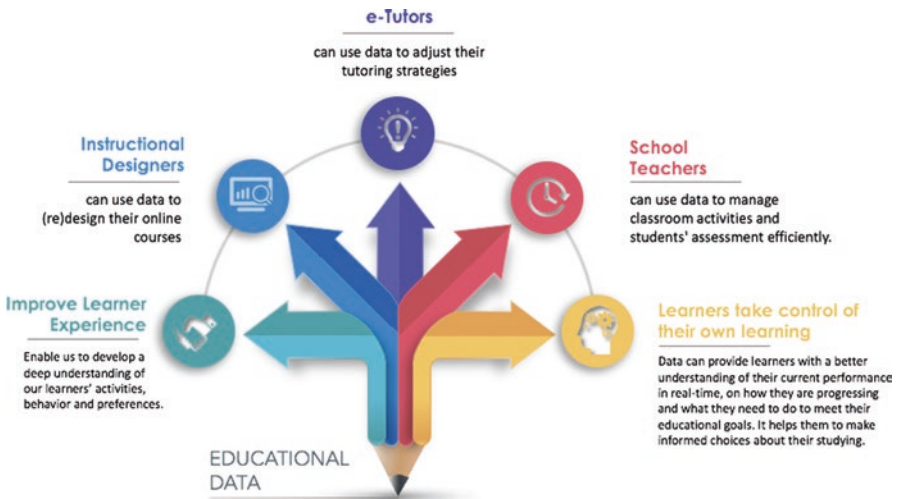
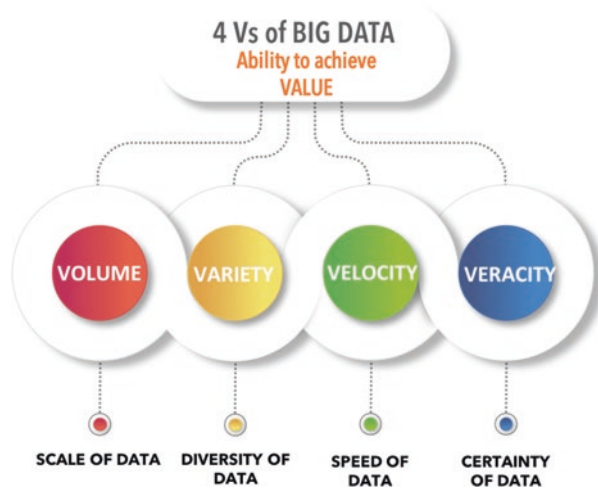


Fig. 1.2 Educational data opportunities

Therefore, in both online and blended courses (Fig. 1.2),

- **instructional designers can use data to (re)design their courses,**
- **tutors can use data to adjust their tutoring and learners' support strategies,**
- **school teachers can use data to better plan inside and outside classroom activities and assess students' learning.**

On the other hand, data could potentially enable learners to take control of their own learning. When appropriately delivered, data can provide learners with better

insights about their current academic performance in real-time, about their progress (also in comparison to their peers) and recommendations about what they need to do for meeting their learning goals and help them to make informed, data-driven choices about their studying (Sclater et al., 2016).

The Data Quality Campaign, in the video “**Data is Power**” (in the useful video resources), highlights the importance of collecting and using quality data to transform education. Nevertheless, the provision of educational data by itself does not automatically lead to improved teaching and learning. Appropriate analyses and sensemaking of educational data allow us to identify actionable insights and inform decision making.

Data-Driven Decision Making is about that.

Data-driven decision making (DDDM) is defined as

the systematic collection, analysis, examination, and interpretation of data to inform practice and policy in educational settings (Mandinach, 2012).

Data-driven decision making has become an essential component of educational practice in order to ground decisions based on data and evidence.

Data-Driven Decision Making (DDDM) crosses all levels of the educational system and uses a variety of data from which decisions can be made. Therefore, it can be challenging to engage in DDDM due to data being siloed in different sources and at different levels.

Developing **competences for effective DDDM** is essential for education professionals. Such competences require “to effectively transform information into actionable knowledge and practices by collecting, analyzing, and interpreting all types of data” (Ridsdale et al., 2015).

Decisions fall into two categories (Marsh et al., 2006):

- Using data as a **diagnostic tool** to identify, inform, or clarify issues both at learners’ level (e.g. identifying needs) and at institution level (e.g. informing the design of courses or curricula), and
- Using data **to act** (e.g. assessing and acting upon differential outcomes among the learners’ population, personalised interventions for at-risk learners).

Data is not a static entity and therefore decisions based on data should not be static either. Data usage and evaluation should be continuous and integrated into existing decision-making processes (Fig. 1.3).

Don’t approach data analysis as a cool “science experiment” or an exercise in amassing data for data’s sake. The fundamental objective in collecting, analyzing, and deploying data is to make better decisions (Díaz et al., 2018).

As per Marsh et al. (2006), “*Once the decision to act has been made, new data can be collected to begin assessing the effectiveness of those actions, leading to a continuous cycle of collection, organization, and synthesis of data in support of decision making.*”

Data analytics refers to methods and tools for analysing large sets of different types of data from diverse sources, to support and improve decision-making. Data



Fig. 1.3 From problem identification to informed decision making

analytics are mature technologies that are currently applied in real-life financial, business and health systems.

However, it is only recently (Johnson et al., 2011, p.28-30), that data analytics have been considered in education - first in higher education, and more recently in school education (Bienkowski et al., 2012).

The “[Engaging with students to build a better digital environment](#)” video (in the useful video resources) shows a real-life case study of the implementation of [Jisc digital experience insights service](#) (Jisc, 2018) aiming to improve the student experience of blended learning at Canterbury Christ Church University based on educational data analysis.

As the [Project lead Duncan MacIver](#) concludes “*The data we have from the insights service makes a significant difference to where we are moving digitally as an institution. This lends a credible voice to decisions being made and provides us with a level of confirmation that we are taking actions that are of direct benefit to students.*”

Questions and Teaching Materials

1. Please select the right answer.

As described in the video “[What is Big Data and how does it work?](#)”, in order to process big data and elaborate the endless possibilities offered, we need to use huge computers to analyze the million pieces of data we generate on a daily basis.

Is this sentence True or False?

- True
- False

Correct answer: False

2. Please match the appropriate definition (from the right column), to the respective “V of Big Data” in the left column.

1. Veracity	A. To take a more holistic view of your subject’s past, present and likely future
2. Variety	B. To be confident that you’re working with the truest, cleanest, most consistent data
3. Value	C. To have a more nuanced view of the matter at hand

4. Volume	<i>D. To ground your decisions in continuously updated, real-time data</i>
5. Velocity	<i>E. To get useful insights from superior analytics</i>

Correct answers: 1-B, 2-C, 3-E, 4-A, 5-D

3. Please select the right answer.

According to the “[Big Data's Making Education Smarter](#)” video by Intel, that presents how education technology companies are leveraging big data to make learning more effective, analytics can

1. **be a powerful tool only for assessing learners’ performance**
2. **enable teachers to see what students learn but not how students learn.**
3. **enable schools to decrease drop-out rates.**

Correct answer: C.

4. Please match how educational data and data analytics technologies (from the right column) can support each professional role (in the left column)

1. School Teachers	A. to take control of their own learning
2. Instructional designers	B to adjust their tutoring and learners’ support strategies.
3. Learners	C. To (re)design their courses
4.Tutors	D. To better plan inside and outside classroom activities and assess students’ learning

Correct answers: 1-D, 2-C, 3-A, 4-B

5. Please select the right answer(s). You may select more than one answer.

Data-Driven Decision Making (DDDM) is about:

- A. collecting a huge amount of data**
- B. grounding decisions based on evidence**
- C. informing practice and policy in educational settings**
- D. identifying actionable insights from the educational data**
- E. dealing with overwhelming statistics**

Correct answers: B, C, and D

6. Please match the steps we need to proceed with, for Data-Driven Decision Making (DDDM), in the right order:

1st step	<i>A. Transform Information into Decisions</i>
2nd step	<i>B. Collect required data</i>
3rd step	<i>C. Identify problems</i>
4th step	<i>D. Evaluate decisions</i>
5th step	<i>E. Transform data into information</i>
6th step	<i>F. Frame questions</i>

Correct answers: 1-C, 2-F, 3-B, 4-E, 5-A, 6-D

7. Please select the right answer(s). You may select more than one answer.

As per the video “[Engaging with students to build a better digital environment](#)”, the educational data analysis provided Canterbury Christ Church University with a deep understanding of the

- A. **staff dissatisfaction on having to handle the burden of the inclusion of technology within learning and teaching.**
- B. **students’ expectations when it comes to the inclusion of technology within learning and teaching.**
- C. **parents’ complaints for collecting and using students’ personal data.**
- D. **policy makers’ demand for effective learning and teaching, minimizing the drop-out rates.**
- E. **areas that the university needed to develop to best support their students**

Correct answers: B and E.

8. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to elaborate on your response about the implementation of personalised learning in the following reflective task.

You may reflect on:

1. **What data are you currently collecting? How are you using this data to make decisions and take actions?**
2. **How are you currently using your data to inform the design of your courses?**

1.2.2 Why Educational Data Is Important for Online and Blended Teaching and Learning?

Personalised learning is identified as one of the major educational challenges of the twenty-first century (2017 Horizon Report, Freeman et al. (2017)). Personalised learning refers to the supporting of individual student learning in a pedagogically effective and practically efficient personalised manner, based on their individual short, mid and long-term needs.

Education Elements (2018) states that personalised learning is increasingly recognized as a promising strategy to

- **help students connect with their needs and aspirations,**
- **close achievement gaps,**
- **increase student engagement, and**
- **prepare students to become self-directed, lifelong learners**

by meeting their individual needs, customizing their learning experiences to indulge their interests, using customised lessons, units and projects at their own pace.

Personalised learning has become easier with the leverage of learners' performance, engagement and behaviour data, captured in online and blended learning environments and analysed with the help of data science.

The video published by Educause "[Educause: What Is Personalized Learning?](#)" (in the useful video resources), explains aspects of personalised learning emphasizing the variety of tools and technologies that can support each learner's individual needs. The importance of a personalised learning experience, that is tailored to the learners' unique needs, skills, and interests, is also illustrated in the following infographic "[You Need Data to Personalize Learning](#)" from Data Quality Campaign.

A wide range of data is generated by the learners and stored in online and blended teaching and learning environments. Data is collected from explicit learners' activities, such as completing assignments and taking exams, and from tacit actions, including online social interactions, extracurricular activities, posts on discussion forums, and other activities that are not directly assessed as part of the learner's educational progress (U.S. Department of Education, 2012) (Bienkowski et al., 2012).

Such learner-generated data is used to assess learning progress, to predict learning performance, to detect and identify potentially harming behaviours and to act upon the findings.

Nevertheless, as stated in the [2011 Horizon report](#), we should not solely focus on learners' performance. Deeper analysis of the educational data can be used to improve understanding of teaching and learning taking place online and/or in blended courses.

As it can be seen at the video in Arizona State University "[Using big data to customize learning](#)" (in the useful video resources), online learner generated data is used to customise teaching and learning in subjects like maths, by tailoring the content to the detected needs of the learners.

Every drop-off, click or share is a learner shouting their likes and dislikes. These actions are the eye-rolls, smiles and crossed arms from the classroom, simply in digital format (Greany & Niles-Hofmann (2018), [An Everyday Guide to Learning Analytics](#)).

Questions and Teaching Materials

1. Please select the right answer(s). You may select more than one answer.

According to Anthony Kim, CEO and Founder of Education Elements “We don’t need a model of superhuman superhero teachers. We need to use the power of technology and educational design—combined with the high aspirations we all begin with—in order to create innovative learning environments that foster personalised learning for everyone.” Why does personalised learning matter, as per Chap. 2 “Why personalized Learning” in [Education Elements](#)?

Personalised learning

- A. creates lifelong learners.
- B. decreases student engagement.
- C. increases student achievement.
- D. limits the time for teachers to focus on each student.
- E. is the future of learning.

Correct answers: A, C, E

2. Please select the right answer(s). You may select more than one answer.

As stated by Data Quality Campaign (in the infographic “[You need Data to Personalize Learning](#)”) “For all students to be college and career ready, they need a learning experience that is tailored to their unique needs, skills, and interests. Data is a critical tool that makes this personalised learning possible”. You are requested to explain the reasons.

With Data:

- A. Learning is individualized.
- B. Learning continues outside of the classroom.
- C. Teachers have the total control of their students learning.
- D. Learning is about mastery.
- E. Learning is about time spent in class.

Correct answers: A, B, D

3. Please select the right answer.

As described in the [Executive Summary of the Issue Brief](#) published by [U.S. Department of Education in 2012](#), K–12 schools and school districts are starting to adopt applications of educational data mining and learning analytics techniques in order to proceed with institution-level analyses

- A. **for recognizing the costs and challenges associated with collecting and storing logged educational data.**
- B. **for detecting areas for instructional improvement, setting policies, and measuring results.**
- C. **for detecting ethical obligations associated with knowing and acting on student data.**

Correct answer: B

4. Please select the right answer

At the Arizona State University (as per the video “[Using big data to customize learning](#)”), appropriate software adapts to each individual student needs by analyzing students’ every keystroke to figure out their learning styles. The software harvests information from the devices the students use and collates grades, learning skills, strong and weak points and even hesitation patterns when using the computer mouse.

This is achieved using:

- A. **Predictive Algorithms**
- B. **Prescriptive Algorithms**

Correct answer: A.

5. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to elaborate on your response about the implementation of personalised learning in the following reflective task.

You may reflect on:

1. **How can you as an instructional designer, tutor or school teacher develop an evaluation plan for a personalised learning intervention?**
2. **Apart from personalised learning, how can educational data be important to your role as instructional designer, tutor or school teacher?**

1.2.3 *How Educational Data Can Help Instructional Designers and e-Tutors of Online Courses?*

Instructional design - also referred to Learning Design or Educational Design - is a systematic and iterative process for any educational challenge (including professional training and human performance improvement) that requires an educational intervention.

Reiser (2001) states that: “The field of instructional design and technology encompasses the analysis of learning and performance problems, and the design, development, implementation, evaluation and management of instructional and non-instructional processes and resources intended to improve learning and performance in a variety of settings, particularly educational institutions and the workplace” (p.53).

The widely used **ADDIE model**, illustrated in the below infographic from Obsidian Learning (2018) (Fig. 1.4), is a five-phase approach to analyse, design, develop, implement and evaluate any teaching and learning product and process in an effective and efficient way.

Within the context of the ADDIE approach

- **Instructional designers (ID)** are the professionals in the field of instructional design, mainly engaged in the analysis, the design, the development and the evaluation phases, whereas,
- **Trainers or tutors** are the professionals engaged mainly in the implementation phase and they can also inform the evaluation phase.

The roles of instructional designers and trainers/tutors in online and blended courses require new competences compared to those in traditional face-to-face education and training programs.

Instructional Designers are mainly engaged in the analysis, the design, the development and the evaluation phases of the ADDIE process.

Analysis Phase

During this phase, the instructional designer identifies an instructional (educational or learning) problem and analyses the parameters of the context in which teaching and learning will take place, as well as the learners’ characteristics and their existing competences (knowledge, skills and attitudes). As a result, the key elements of this phase can be codified as follows:

- A.1. **Instructional/educational/learning problem Identification:** aims to address why a teaching/learning process (broadly referred to as *educational intervention*) is needed for the identified problem. **Contextual Analysis:** aims to capture where the educational intervention will be implemented, namely the learning environment.
- A.2. **Learner Analysis:** aims to analyse for whom the educational intervention will be designed.

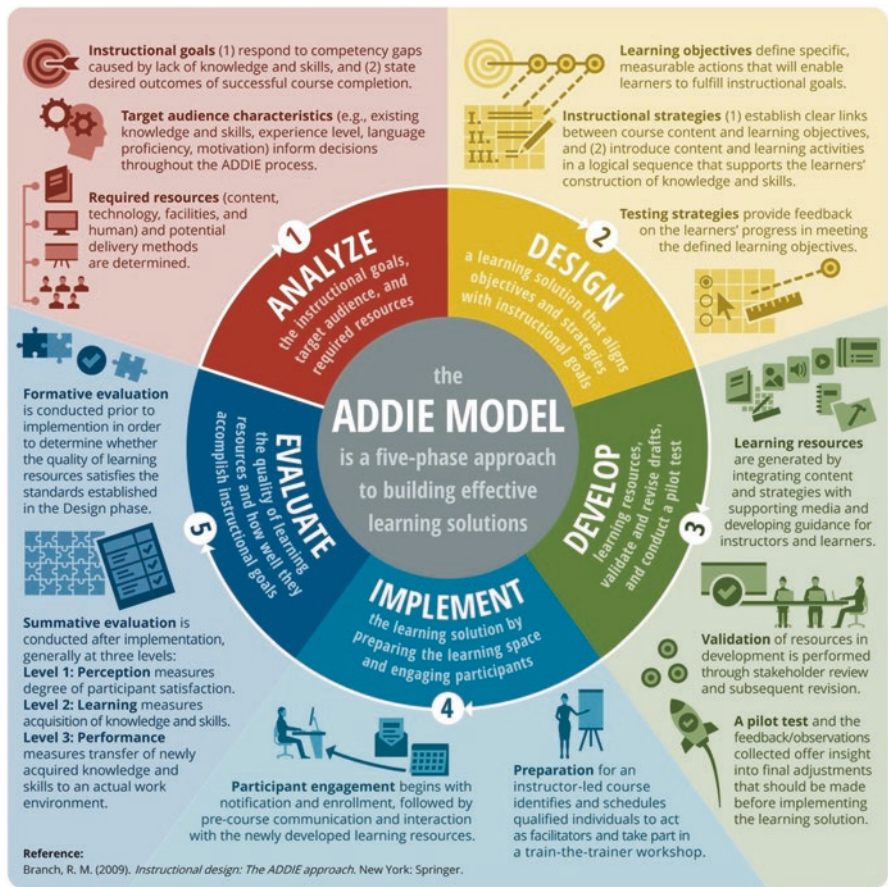


Fig. 1.4 ADDIE model representation (Krisna kristiandi hartono [CC BY-SA 4.0])

The major outcome of the Analysis phase is a granulated overview of the contextual and learner conditions that will be used to configure and formulate the upcoming Design phase.

Design Phase

During this phase, the instructional designer defines the educational objectives to be achieved, selects an appropriate teaching approach for attaining these objectives, as well as, appropriate assessment methods for evaluating whether and to what extent the educational objectives have been met. As a result, the key elements of this phase can be codified as follows:

DES.1. Definition of Educational Objectives: this includes the definition of general educational objectives, as well as the development of specific subject matter objectives, aligned to the general objectives.

- DES.2. **Selection of Teaching Approach/Strategy:** this includes the selection of an appropriate teaching approach/strategy for supporting learners in attaining the educational objectives. Additionally, based on the selected strategy, this phase also includes the formulation of the specific learning activities and their appropriate sequencing in order to attain the expected educational objectives. Finally, a direct mapping between each learning activity to the educational objectives that they aim to cultivate is also performed.
- DES.3. **Selection of Assessment Method(s):** this includes the selection of appropriate assessment methods for evaluating the level of achievement of the educational objectives. This includes also sequencing and description of assessment activities according to the selected teaching approach/strategy and assessment method(s).

The main outcome of the Design phase is a detailed blueprint of the flow and description of the learning and assessment activities, which also accommodates the contextual and learner considerations from the Analysis phase.

Develop Phase

During this phase, the development or selection of appropriate educational materials and the development/arrangement of the appropriate delivery setting is performed for the outcome of the Design Phase. This phase can involve except from the instructional designer, other individuals such as subject matter experts or technical and media experts. As a result, the key elements of this phase can be codified as follows:

- DEV.1. Development or selection of **educational resources** for supporting learning and/or assessment activities of the Design Phase
- DEV.2. Development or selection of **educational tools and/or services** for supporting learning and/or assessment activities of the Design Phase
- DEV.3. Development/arrangement of the **appropriate delivery setting** where learning will take place. For example, development/selection of a digital delivery system or appropriate arrangement of a physical delivery setting such as a classroom.

The main outcome of the Develop phase is the selection or production of educational materials/tools that can appropriately support the outcome of the Design Phase.

Evaluate Phase

During this phase, an evaluation of both the entire teaching and learning process, as well as each phase, is performed towards identifying whether the desired results have been achieved. As a result, the key elements of this phase can be codified as follows:

- E.1. **Formative Evaluation:** this includes an ongoing evaluation process during design, development and implementation phases and aims to maximize pedagogical/ andragogical effectiveness (e.g. achievement of educational objectives) and/or implementation efficiency (e.g. time/cost reduction)

E.2. Summative Evaluation: this is performed after completion of the Implement phase and aims to measure pedagogical/ andragogical effectiveness (e.g. achievement of educational objectives) and/or implementation efficiency (e.g. time/cost reduction).

The main outcome of the Evaluate phase is to identify issues or changes needed, so as to refine the design, development and implementation phases of future designs and to assess whether the desired results have been achieved.

Trainers or tutors, as part of their role, they are mainly engaged in the implementation phase of the ADDIE process, whereas they can also inform the evaluation phase.

Implement Phase

During this phase, the outcome of the previous phases is delivered to the learners. Although delivery is typically addressing groups of learners, still emphasis should be given to providing individual learning experiences, including *scaffolding* and *feedback*. To this end, it is important that learners' (and teachers'/tutors') actions are tracked and meaningful educational data is collected (to be analysed and inform reflection and decision making). As a result, the key elements of this phase can be codified as follows:

- I.1. **Delivery:** this includes the delivery of the product from Analysis, Design and Develop phases to the learners.
- I.2. **Monitoring:** this includes tracking of learners' (and teacher/tutors') actions and collecting meaningful educational data based on which teachers can form evidence-based run-time adaptations/revisions (and also specify which and why these adaptations or revisions were performed).

The main outcome of the Implement phase is to support learners in attaining the educational objectives by appropriately monitoring them so, if needed, changes and adaptations can be made.

As presented in the figure below (Fig. 1.5), instructional designers and trainers/tutors, as part of their role, leverage educational data at all phases of the ADDIE process they are engaged in.

Questions and Teaching Materials

1. **ADDIE model is a five-phase approach to analyse, design, develop, implement and evaluate any teaching and learning product and process in an effective and efficient way (as presented in the infographic from Obsidian Learning (2018) of Fig. 1.4).**

Please match the sentences (from the right column) corresponding to the respective phase of ADDIE Model (in the left column).

1. Analyze	A. <i>The quality of learning resources</i>
	B. <i>Target audience</i>
2. Design	C. <i>Learning resources</i>
	D. <i>The learning solution by preparing the learning space and engaging participants</i>

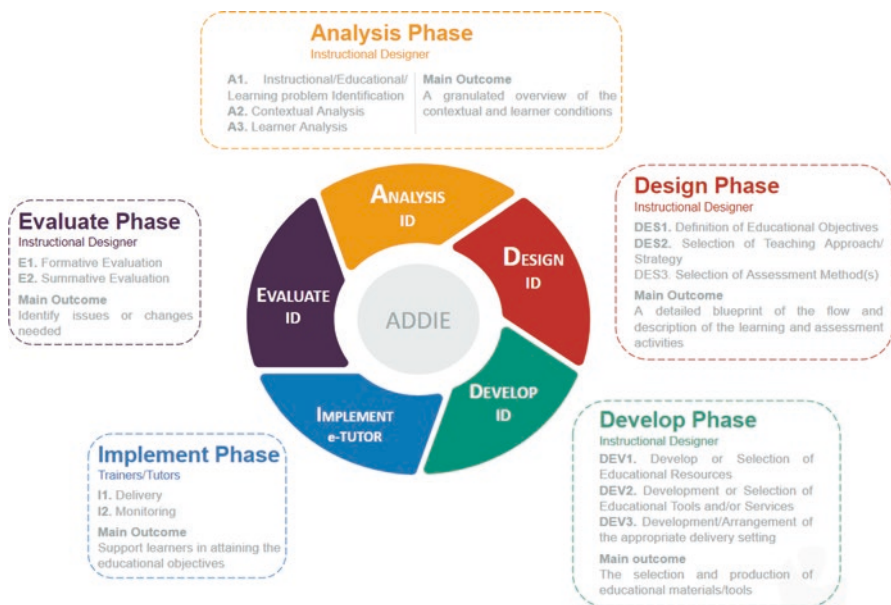


Fig. 1.5 Instructional designers and trainers/tutors leverage educational data in all phases of the ADDIE process

3. Develop	<i>E. Instructional goals</i>
	<i>F. How well the learning resources accomplish instructional goals</i>
4. Implement	<i>G. Validate and revise drafts</i>
	<i>H. a learning solution that aligns objectives and strategies with instructional goals</i>
5. Evaluate	<i>I. Required resources</i>
	<i>J. Conduct a pilot test</i>

Correct answers: 1: B, E, I – 2: H – 3: C, G, J – 4: D – 5: A, F

2. Instructional Designers are mainly engaged in the analysis, the design, the development and the evaluation phases of the ADDIE process. Please mark the correct key elements corresponding to each phase of the ADDIE process.

	Analysis phase	Design phase	Develop phase	Evaluate phase
Formative evaluation				X
Capture learners’ characteristics and their existing competences	X			

	Analysis phase	Design phase	Develop phase	Evaluate phase
Selection of assessment method(s)		X		
Arrangement of the appropriate delivery setting			X	
Identify instructional (educational or / learning) problem	X			
Summative evaluation				X
Definition of educational objectives		X		
Assess whether the desired results have been achieved				X
Selection of educational resources			X	
Selection of teaching approach/strategy		X		

Correct answers: as marked with X above

3. The key elements in the Implement Phase of the ADDIE process are Delivery and Monitoring. Please match the sentences (from the right column) corresponding to the respective key element of the Implement Phase of the ADDIE process (in the left column).

1. Delivery	A. Of meaningful educational data based on which teachers can form evidence-based run-time adaptations/revisions
	B. Of the product from develop phase to the learners
2. Monitoring	C. Of the product from analysis and design phases to the learners
	D. Of learners' (and teacher/tutors') actions

Correct answers: 1: B, C - 2: A, D.

4. Please mark the correct professional role and outcome to the respective phase of the ADDIE process

Correct answers: as marked with X above

	Instructional Designers	Trainers/ Tutors	Support learners in attaining the educational objectives	Description of the learning and assessment activities	Identify issues or changes needed	Granulated overview of the contextual and learner conditions	Detailed blueprint of the flow	Selection and production of educational materials/tools
ADDIE Phase	X					X		
Analysis phase								
Design phase	X			X			X	
Develop phase	X							X
Implement phase		X	X					
Evaluate phase	X				X			

5. Please select the right answer(s). You may select more than one answer.

As described in the video “[Data Driven Learning Design](#)”, “Data-Driven Learning Design very simply is all about just looking at data at the start before you begin any design, put all this data together and build a picture for how you want to design your content to respond to what insights they’re telling you.”

Can you give some examples of the data that you should be looking at, in order to decode the digital body language of your learners?

- A. **what type of devices learners use to log in?**
- B. **what is the political belief of the learners?**
- C. **what’s the time of day that learners are engaging with your content?**
- D. **what is the financial status of the learners?**
- E. **how long of the portion of a video a person looks?**
- F. **whether learners are married**
- G. **data that is easily available – start small**

Correct answers: A, C, E, G.

6. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to elaborate on your response about the use of educational data, in the following reflective task.

You may reflect on:

1. **How can you as an Instructional Designer or e-Tutor leverage educational data from online courses? Please share either your past experience or your thoughts for future actions.**
2. **How can you as an Instructional Designer or e-Tutor use educational data to enhance engagement in an online course? Please share either your past experience or your thoughts for future actions.**

1.2.4 How Educational Data Can Help School Teachers of Blended Courses?

Schools are using self-evaluation as an instrument to engage all key stakeholders (namely, school leaders, educators, parents and students) in reflecting and improving school activities.

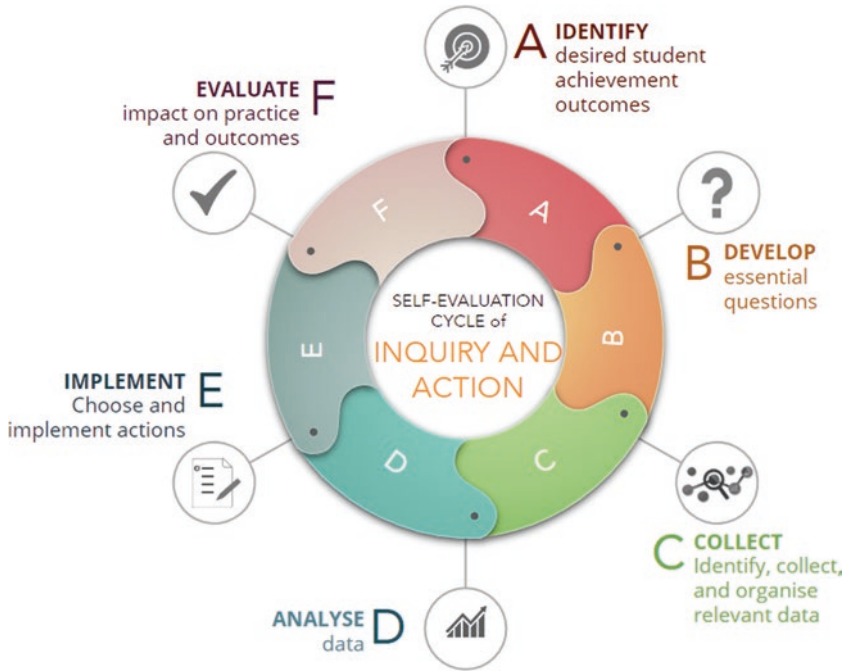


Fig. 1.6 The six essential activities for continuous, non-linear inquiry process for self-evaluation

For example, as presented in the “[School Self-evaluation Guidelines 2016-2020 Primary](#)”, the [Irish](#) Inspectorate of the Department of Education Skills (2016) defines School Self-evaluation as:

a collaborative, inclusive, and reflective process of internal school review. An evidence-based approach, it involves gathering information from a range of sources, and then making judgements. All of this with a view to bring about improvements in students’ learning.

The Annenberg Institute for School Reform (Barnes, 2004) has developed a continuous, non-linear inquiry process for self-evaluation, comprised of six essential activities, depicted in the figure below (Fig. 1.6).

In the video “[Data: It’s Just Part of Good Teaching](#)” (in the useful video resources) from the Data Quality Campaign, Sherman Elementary in Rhode Island demonstrates how the effective use of data by a school community can improve students’ performance. Moreover, the video “[How Data Help Teachers](#)” (in the useful video resources), from the Data Quality Campaign, demonstrates how data helps school teachers and their students succeed. For more details, you may also review the corresponding [infographic](#), “Ms. Bullen’s Data-Rich Year” by DQC.

This video from the Data Quality Campaign, “[Data Can Help Every Student Excel](#)” (in the useful video resources) also discusses what does it mean to use data in service of student learning, taking the stand that data is one of the most powerful tools to inform, engage, and create opportunities for students along their education journey.

Before proceeding further, let’s now discuss what flipped classroom is all about. As per Panopto (2015) “The flipped classroom is a teaching strategy in which the traditional class format is turned on its head. This inverted model “flips” the traditional order of class activities so that school work is done at home, and “homework” is done at school. In flipped classes, students review lecture materials prior to class, reserving in-class time for teacher-guided activities that allow students to put the lecture materials into practice. Activities can include in-depth discussion, labs, debates, problem-solving, or just open time for individual assignments — all with the added benefit of having the teacher nearby to help when questions arise.”

This new approach enables teachers to make the shift from teacher-driven instruction to student-centred learning and thus to reinforce deeper learning. You may also review the video “The Flipped Classroom Model” (in the useful video resources).

As Brame (2013) from the University of Vanderbilt, Center for Teaching, suggests, the flipped classroom approach yields statistically significant improvements in engagement, test scores and overall long-term learning.

The infographic “How Educational Data can help School Teachers of Blended (Flipped Classroom) Courses?” is presenting a use-case with an example for the school teacher of blended learning courses in the K-12 education context (Fig. 1.7).

The video “Why Personalized Learning: 4 Stories from 4 School Districts” (in the useful video resources) shows 4 School Districts sharing their findings on the

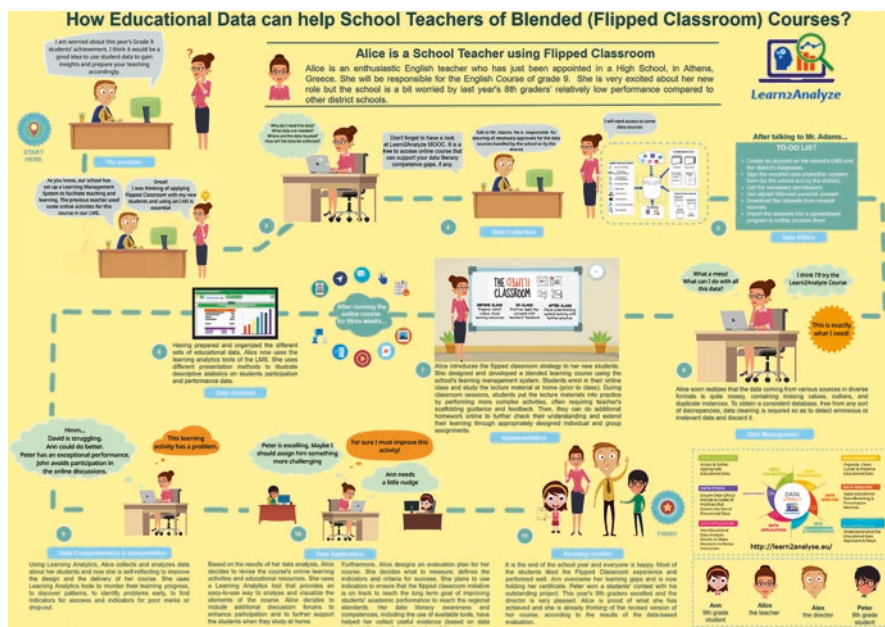


Fig. 1.7 Infographic for school teacher in K12 blended courses

implementation of blended courses aiming to provide a unique personalised learning experience to their students.

Questions and Teaching Materials

1. Please select the right answer(s). You may select more than one answer.

As described in the Chap. 2 of the “[School Self-evaluation Guidelines 2016-2020 Primary](#)” (Inspectorate, Department of Education and Skills, 2016), self-evaluation requires a school to address the following key questions with regards to an aspect or aspects of its work.

- A. **How bad are we doing?**
- B. **How do we know?**
- C. **What other schools do?**
- D. **How can we find out more?**
- E. **What are our strengths?**
- F. **How can we hide our weaknesses?**
- G. **What are our areas for improvement?**
- H. **How can we improve?**

Correct answers: B, D, E, G, H

2. Please select the correct answer.

According to Sherman Elementary in Rhode Island (please refer to the video “[Data: It's Just Part of Good Teaching](#)”), the use of data may be really beneficial for the school community and can lead to improved academic performance. Nevertheless, it creates an extra add-on and an overwhelming burden for both teachers and students.

Is this statement valid?

- **Yes**
- **No**

Correct answer: No

3. Let’s meet Alice! Alice is an enthusiastic English Language teacher who has just been appointed in an Experimental High School, in Athens, Greece. She will be responsible for the English Language Course of class1 and class2 of the ninth Grade (14 to 15 years’ students). Alice is very excited about her new role. Nevertheless, the school’s principle, Alex, is concerned about the relatively low performance of last year’s eighth graders compared to other experimental schools in the region. Alex encourages Alice to use student

data to gain insights and plan her teaching activities accordingly, so as to improve this year's Grade 9 students' academic performance.

Alice decides to watch the video by Data Quality Campaign “[How Data Help Teachers](#)”, so as to find out how she can leverage data for her students to succeed. Alice is really inspired by Ms. Bullen who empowers her student Joey to get on track meeting the educational goals.

Can you help Alice to arrange the instances of Ms. Bullen's story in the right order?

-
- A. Joey's data shows that he's on track for success.
-
- B. Ms. Bullen reviews her performance with the principal to note strengths and opportunities.
-
- C. She uses data and her experience in the classroom to see where Joey and his classmates excel or struggle.
-
- D. Ms. Bullen gets access to data on her students' past performance, behaviour and attendance.
-
- E when Ms. Bullen sees that Joey is at risk of failing, she works with Joey, his parents and his other teachers to get him on track by the end of the year.
-
- F. Ms. Bullen talks with Joey about his own data and they work together to set goals throughout the year.
-

Correct answers: 1-D, 2-C, 3-F, 4-B, 5-E, 6-A

4. **Please select the appropriate answer (s). You may select more than one answer.**

Alice is really excited about the power of using data in service of learning and for personalizing her instruction to keep every student on track to excel. Thus, she searches for further information. Alice now watches the video about [How “Data Can Help Every Student Excel”](#). She then reviews the below statements. Something seems wrong...

Which of these statements are not valid?

- A. **When students, parents and educators have the right information to make decisions, students excel.**
- B. **Data is one of the most powerful tools to inform, engage, and create opportunities for students along their education journey.**
- C. **Education data is all about test scores.**
- D. **Education data comes from many sources and in many formats.**
- E. **Policy makers know the detailed profile for each and every student.**
- F. **Making data work for students means collecting all possible data in all available formats**

Correct answers: C, E, F

5. Please select the right answer(s). You may select more than one answer.

Let's go back to Alice. The principal informs Alice about the Learning Management System used by the school to facilitate teaching and learning, pointing out that the previous teacher has already created some online activities there.

Alice decides to apply the flipped classroom strategy to her new students using the school's LMS. For this purpose, she designs and develops online teaching resources for Class1 and Class2. Students of these classes enrol in the respective group and study the lecture material at home (prior to classroom meeting). The material is in the form of video, text, small activities with automatic feedback (such as online quizzes), and forum discussions. During the classroom sessions, students are performing more complex activities, typically in small groups, with the benefit of Alice's scaffolding, guidance and feedback. Then, they can undertake some additional homework online to further check their understanding and extend their learning through appropriately designed individual and group assignments.

Alice reads the article ["Flipping the classroom" By Cynthia J. Brame](#). She is looking at the key elements of the flipped classroom.

- A. **Provide an opportunity for students to gain first exposure prior to class.**
- B. **Provide a mechanism to assess student understanding.**
- C. **Provide in-class activities that focus on lower-level cognitive activities.**
- D. **Provide an incentive for students to prepare for class.**
- E. **Provide an opportunity for teachers to keep in contact with parents.**

Correct answers: A, B, D

6. **As referred to the infographic ["How Educational Data can help School Teachers of Blended \(Flipped Classroom\) Courses?"](#), to improve this year's Grade 9 students' academic performance, Alice decides to apply the flipped classroom strategy to her new students using the school's Learning Management System. Alice starts creating a detailed plan about the needed steps to go through in order to make her flipped classroom strategy a success story for herself, her students, her principle and the parents. Can you help Alice to arrange the steps of her plan in the right order?**

Please arrange the instances in the right order.

- A. **Following data analysis, Alice plans to use Learning analytics to monitor students' learning process, to discover patterns, to identify problems early, to find indicators for success and indicators for poor marks or drop-out and self-reflect to improve the design and the delivery of her course comprehending the story that the collected data reveals.**

- B. Furthermore, Alice plans to design an evaluation plan for her course using indicators to ensure that the flipped classroom initiative is on track for reaching the long-term goal of improving students' academic performance to reach the regional standards.
- C. Alice will contact the school's data protection officer (DPO), to secure all necessary approvals for the sources handled by her school or by the corresponding district. She will also request to grant her access to the LMS used by the school and ask the DPO if signed informed parental consent are necessary for all participating students before implementing the flipped classroom initiative.
- D. After running the online course for three weeks, Alice will use different presentation methods to illustrate descriptive statistics on students' participation and performance data.
- E. Following the principal's advice, Alice plans to use student data to gain insights and plan her teaching activities accordingly. She will gather a variety of students' data, including demographics, perception data, past academic performance. To retrieve the needed data, she has to access diverse sources: school's internal data sources like the student information system as well as external data sources, like the district's databases.
- F. As data coming from various sources is quite messy, containing missing values, outliers, and duplicate instances, data cleaning is required to obtain a consistent database, free from any sort of discrepancies.

Correct answers: 1E-2C-3F-4D-5A-6B

7. Please select the right answer (s). You may select more than one answer.

Alice is unstoppable. She meets regularly with other teachers for training and for identifying promising data use practices for her flipped classroom model. She now wonders. What are the findings of other schools that have successfully implemented their Blended Courses?

She retrieves the video describing the 4 stories from 4 School Districts on personalised learning "Why Personalized Learning: 4 Stories from 4 School Districts, https://www.youtube.com/watch?v=ur2E_S1IBP0". She notes down some initial findings. She reads them again and she notices that she probably mixed it up a bit. **Can you help her identify the right findings?**

- A. Personalised learning is really about traditional notions of grade-level expectations.
- B. Personalised learning enables teachers to understand where students are and recognize the potential for moving them forward.
- C. Keeping middle school kids on track and on task academically can often be the biggest challenge for teachers.
- D. In this model the teacher is not the key education element anymore.

- E. **In this model teachers may have individual conversations with students each day and figure out where they are at and take them to the next level.**
- F. **Teachers are not interested in looking at each individual student's growth.**
- G. **There is a lot more student engagement in the digital content or with the device and less student engagement with the teacher.**
- H. **With supportive personalised learning schools create students who are knowledgeable critical thinkers, communicators, collaborators, creators and contributors.**

Correct answers: B, C, E, H.

8. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to elaborate on your response about using educational data in blended learning environment, in the following reflective task. You may reflect on your experience from implementing flipped classroom and/or express your opinion on why to use blended learning (or not!):

1. **Teachers always used data in their everyday practice. Whether you are an instructional designer, an e-tutor or a school teacher, reflect on which particular learners' data do you use and how?**
2. **How data gathered from the online component of a blended learning flipped classroom, can help school teachers improve their teaching?**

1.2.5 *The Learn2Analyze Educational Data Literacy Competence Framework*

Educational data literacy is defined as:

the ability to collect, manage, evaluate, and apply data, in a critical manner (Ridsdale et al., 2015).

the ability to accurately observe, analyse and respond to a variety of different kinds of data for the purpose of continuously improving teaching and learning in the classroom and school (Love, 2012).

the ability to understand and use data effectively to inform decisions ... composed of a specific skill set and knowledge base that enables educators to transform data into information and ultimately into actionable knowledge (Mandinach & Gummer, 2013).

[the capacity to] continuously, effectively, and ethically access, interpret, act on, and communicate multiple types of data from state, local, classroom, and other sources in order to improve outcomes for students in a manner appropriate to their professional roles and responsibilities (Data Quality Campaign, 2014a).

Thus, educational data literacy refers to the competence set which is required to identify, collect, combine, analyse, interpret and act upon educational data from



Fig. 1.8 Educational data literacy roadmap

different sources, with the aim of continuously improving the teaching, learning and assessment process.

In the “[Roadmap for Educator Licensure Policy Addressing Data Literacy](#)” report, the Data Quality Campaign recommends the following set of Data Literacy Competences for teachers, (Fig. 1.8):

1. **Locate and Collect Relevant Educational Data**
2. **Synthesise and Analyse Educational Data from Diverse Sources**
3. **Know about Educational Data beyond Grades**
4. **Understand How to Use Educational Data beyond Grades**
5. **Engage in a Data-Driven Continuing Inquiry Process**
6. **Use Data Analysis to Customise Teaching Plans to Diverse Groups**
7. **Use Own Data to Reflect on Practice**
8. **Facilitate Students to Understand their Data**
9. **Communicate Insights from Data Analysis to Diverse Internal and External Stakeholders**
10. **Monitor this process in a continuous manner**

As already discussed, Educational Data Analytics are attributed with significant benefits for enhancing **personalised educational support** of the learners as well as **reflective course (re)design** for achieving improved **teaching, learning and assessment**.

However, emerging advancements related to the use of data-driven design and delivery of online and blended learning courses, exploiting Educational Data Analytics are not yet thoroughly addressed by existing competence frameworks for education professionals (instructional designers, trainers, educators, teachers). Existing professional competence frameworks for instructional designers and trainers almost ignore the dimension of Educational Data Literacy.



Fig. 1.9 Learn2Analyze educational data literacy dimensions

To this end, the **Learn2Analyze project** has developed a comprehensive proposal for an **Educational Data Literacy Competence Framework** to enhance existing competence frameworks for instructional designers and e-trainers of online courses with new Educational Data Literacy competences.

The Learn2Analyze Educational Data Literacy Competence Framework comprises of 6 competence dimensions and 17 competence statements, as captured in Fig. 1.9.

In addition, the following table (Table 1.1.) provides a brief overview of the Learn2Analyze Educational Data Literacy Competence Framework.

Questions and Teaching Materials

1. Please select the right answer.

Alice is back! She has realized that, for effective data use, Educational Data Literacy is a prerequisite. So, she is trying to understand the meaning of this much-discussed key component.

Alice believes that educational data literacy refers to the specific skill of using assessment data effectively in order to improve outcomes for students.

Is this definition by Alice valid?

- Yes
- No

Correct answer: No

Table 1.1 Learn2Analyze educational data literacy competences

1. Data Collection	1.1 Know – understand – be able to obtain, access and gather the appropriate data and/or data sources
	1.2 Know – understand – be able to apply data limitations and quality measures (e.g., validity, reliability, biases in the data, difficulty in collection, accuracy, completeness)
2. Data management	2.1 know – understand – be able to apply data processing and handling methods (i.e., methods for cleaning and changing data to make it more organized – e.g., duplication, data structuring)
	2.2 know – understand – be able to apply data description (i.e., metadata)
	2.3 know – understand – be able to apply data curation processes (i.e., to ensure that data is reliably retrievable for future reuse, and to determine what data is worth saving and for how long)
	2.4 know – understand – be able to apply the technologies to preserve data (i.e., store, persist, maintain, backup data), e.g., storage mediums/ services, tools, mechanisms
3. Data analysis	3.1 know – understand – be able to apply data analysis and modeling methods (e.g. application of descriptive statistics, exploratory data analysis, data mining).
	3.2 know – understand – be able to apply data presentation methods (e.g., pictorial visualisation of the data by using graphs, charts, maps and other data forms like textual or tabular representations)
4. Data Comprehension & Interpretation	4.1 know – understand – be able to interpret data properties (e.g., measurement error, outliers, discrepancies within data, key take-away points, data dependencies)
	4.2 know – understand – be able to interpret statistics commonly used with educational data (e.g., randomness, central tendencies, mean, standard deviation, significance)
	4.3 know – understand – be able to interpret insights from data analysis (e.g., explanations of patterns, identification of hypotheses, connection of multiple observations, underlying trends)
	4.4 be able to elicit potential implications/links of the data analysis insights to instruction
5. Data application	5.1 know – understand – be able to use data analysis results to make decisions to revise instruction
	5.2 be able to evaluate the data-driven revision of instruction
6. Data ethics	6.1 know – understand – be able to use the informed consent
	6.2 know – understand – be able to protect individuals’ data privacy, confidentiality, integrity and security
	6.3 know – understand – be able to apply authorship, ownership, data access (governance), re-negotiation and data-sharing

2. Please select the right answer.

Alice decides to study further on Educational Data Literacy. She reads the brief intended for State Policy Makers by Data Quality Campaign “[Teacher Data Literacy: It’s About Time](#)”.

According to this brief, “One element of quality teaching for improving student outcomes is effective data use. Teacher data use is also the best way to maximize state investment in data systems.”

Consequently, to date, policies have heavily promoted the skills teachers need to be data literate. Thus, many teachers regard data as a powerful tool for improving instruction and ultimately outcomes for students.

Is this interpretation by Alice, True or False?

- **True**
- **False**

Correct answer: False.

3. **Alice is now interested in finding out the skills that teachers need to be qualified, so as to integrate the use of data into their everyday practice as one tool for improving student achievement. Are you ready to support Alice in this task?**

According to the “[Roadmap for Educator Licensure Policy Addressing Data Literacy](#)” report, the ability to effectively use data includes a set of skills that teachers (and administrators) need to use data both collaboratively and individually to inform instruction.

Please match the appropriate definition (from the right column), to the respective data use skill in the left column.

1. Synthesize and analyze diverse data	A. Ability to use one’s own performance data and other relevant data to assess and reflect on personal practice for the purpose of continuous improvement.
2. Know about and use student-level and other types of data beyond assessment data	B. Ability to engage in the ongoing process of identifying classroom and system problems, forming questions and hypotheses about each issue, collecting and analyzing relevant information and translating.
3. Engage in a data-driven and cyclical inquiry process	C. Ability to use data to communicate with students about their progress so that they can evaluate their own performance and set goals.
4. Use one’s own data	D. Ability to explore and organize many types of data.
5. Facilitate student understanding of data	E. Understanding that multiple types of data beyond student assessments can be used to inform practice.

Correct answers: 1-D, 2-E, 3-B, 4-A, 5-C.

4. **Alice is confident with the flipped classroom approach, as she has used it before with great results. However, she realises that she is lacking data literacy competences**

The principle encourages her to enrol in the Learn2Analyse MOOC before the school year starts – it is only an 8-week course and it is free. She is really excited and she immediately reviews the Educational Data Literacy Competence Profile Framework that comprises of 6 competence dimensions.

Can you assist Alice to arrange these 6 competence dimensions in the right order?

Please arrange the 6 competence dimensions in the right order.

- _____ A. *Data analysis*
- _____ B. *Data ethics*
- _____ C. *Data management*
- _____ D. *Data comprehension & interpretation*
- _____ E. *Data collection*
- _____ F. *Data application*

Correct answers: 1-E, 2-C, 3-A, 4-D, 5-F, 6-B

5. **Alice studies thoroughly the competence statements of the Learn2Analyse Educational Data Literacy Competence Profile Framework (L2A EDL-CP Framework). She is really interested in investigating the exact competences she needs to develop in order to be Educational Data Literate and use data both effectively and ethically to improve her students’ achievements.**

But she needs your assistance, again!

Please help Alice to mark the correct statements corresponding to each of the 6 competence dimensions of the EDL-CP Framework.

	Data collection	Data management	Data analysis	Data comprehension & interpretation	Data application	Data ethics
Use the informed consent						X
Apply the technologies to preserve data		X				
Apply modelling methods			X			

	Data collection	Data management	Data analysis	Data comprehension & interpretation	Data application	Data ethics
Evaluate the data-driven revision of instruction					X	
Gather the appropriate data	X					
Elicit potential implications				X		
Apply data limitations and quality measures	X					
Apply data presentation methods			X			
Interpret insights from data analysis				X		
Apply authorship & ownership						X
Use data analysis results to make decisions to revise instruction					X	
Apply data curation processes		X				

Correct answers: as marked with X above

6. Alice now has a good sense that in order to achieve her strategy to become a success story, her school needs to develop a culture of data enablement for all parties involved in the life of a student.

Thus, she watches the interview of the well-known Gartner’s Analyst Rita Sallam “[Develop a culture of data and analytics enablement at the summit](#)”, to find out what are the actions that her school needs to proceed with.

She understands how important it is to ensure that every person in her school is data literate in order to be able to leverage the information provided to each of them.

Enabling analytics to have maximum impact on her school, involves addressing key concepts like diversity in school's data, systems and most importantly diversity of the people, to make sure that the school can get diverse ideas and diverse skills to really innovate. Do you agree with the assumption of Alice?

Please select the correct answer:

- **Yes**
- **No**

Correct answer: Yes.

7. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to elaborate on your response to educational data literacy training in the following reflective task. You may reflect on:

1. **your experience from attending such a course, and/or**
2. **your thoughts on why to attend such a course (or not!)**

1.3 Data Is Everywhere (Educational Data Collection)

1.3.1 *Posing Questions and Identifying Appropriate Educational Data*

In the previous sections we reviewed the key role of educational data. In this section we will have a closer look of what educational data is.

In the school context, educational data can be broadly defined as:

information that is collected and organised to represent some aspect of schools. This can include any relevant information about students, parents, schools, and teachers derived from qualitative and quantitative methods of analysis (Lai & Schildkamp, 2013, p. 10).

As this definition suggests, educational data is not restricted to students' grades in national exams and standardised tests (although that is a common misconception). Instead, educational data comprises a wide range of data from various sources, both internal (school-wide and classroom-specific data) and external (state and/or district data) to the school.

This definition can be extended to higher education and professional training institutions, as represented in Fig. 1.10 (Long & Siemens, 2011).

We can distinguish two major categories of data, the qualitative and quantitative data. With a combination of different types of data being the most effective in generating powerful evidence to assess learning performance and improve teaching practice. Both quantitative and qualitative data is equally important in these processes (Fig. 1.11).

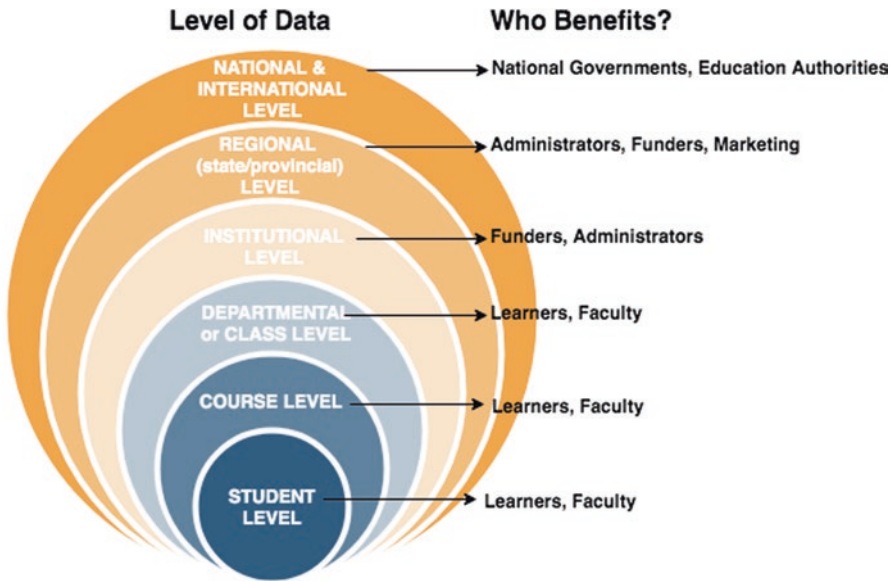


Fig. 1.10 Different levels of educational data



 qualitative data	vs	 quantitative data
qualitative data is data that is not easily reduced to numbers. In a school setting, qualitative data may come from observations, work samples, conversations, written documents and more.		quantitative data is any information that can be reduced to a set of numbers. Information from which you can create averages, differences or totals is quantitative data.
is particularly helpful for summarising large amounts of information in a snapshot, tracking trends over time and understanding patterns and differences from one group of participants to another.		is particularly helpful for exploring emerging issues, providing rich description and context about a complicated issue and building theories about what might explain trends and patterns in quantitative data.
help us answer questions about <ul style="list-style-type: none"> • 'what', • 'how' and • 'why' of a phenomenon 		help us answer questions about <ul style="list-style-type: none"> • 'how many' or • 'how much'

Fig. 1.11 Quantitative and qualitative educational data

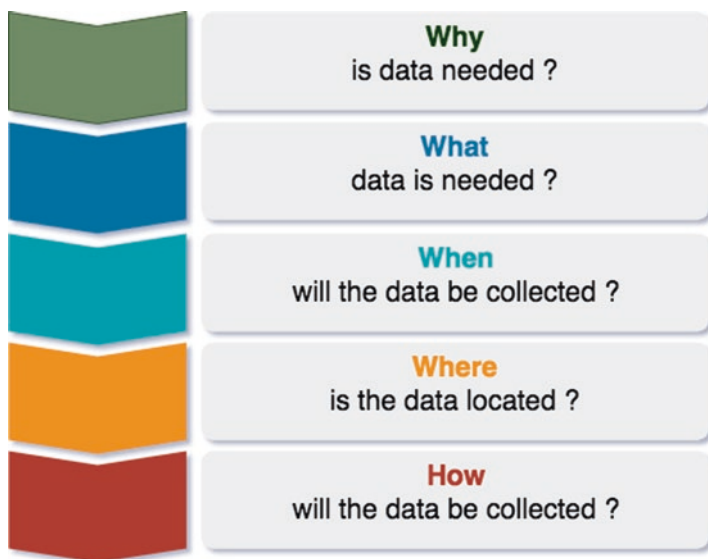


Fig. 1.12 The four Ws and one how of educational data

As discussed, for effective Data-Driven Decision Making, we need to be data literate, to **be able to understand basic data science processes (speak data)**. According to Gartner analysts Idoine, Schlegel, and Sallam (Petty, 2018), “*learning to “speak data” is like learning any language. It starts with understanding the basic terms and describing key concepts.*” In our case, the first key area of data literacy vocabulary is **Educational Data Collection**.

Educational Data is everywhere. To inform our decisions and benefit from them, we need to **collect the necessary data**. To do this, we need to answer to the “**Four Ws and One How**” questions, presented in Fig. 1.12. We should know why we collect this data, what types of data we need to collect, when and how to get it and where to find it. “*Who will collect or grant access to the needed data*”, is also a question that should be answered, since, obviously, we can only collect data to which we have access and which we have been granted permission to use.

You will probably need to utilize a variety of data types from different sources and use various methods to process and analyse them according to our goals.

Bear this strategy in mind and start posing questions that will help you identify and collect the appropriate educational data. Your ultimate goal is to improve your instructional e-learning strategy and make your online and blended course a success story for your target learners.

We’ll now guide you step by step through the effective process for collecting educational data by answering each one of these key questions.

Most things start with a question. The first question to ask ourselves is “Why is data needed? Why we need to collect the data, in the first place?”

	Instructional Designer	e-Tutor	Teacher
Navigation & click stream data	<ul style="list-style-type: none"> Do different course structures show a difference in learner navigation patterns? Where do students exit? How should the course home page be designed to make sure learners come back? Is there an event that is always triggered first? Does it lead to more events or more pages? 	<ul style="list-style-type: none"> How do students navigate through course content? Which files had the most usage? How much time do learners spend in the online learning environment? Is there a relationship between students activity time in the course and their performance? 	<ul style="list-style-type: none"> Are there paths through the course site that are more popular than others, and if so, are those the paths that you want students to follow? Did students go right from homepage to assignments/quizzes without additional navigation?
Discussion Forum Interaction Data	<ul style="list-style-type: none"> How can I generate more activity in the forums? How do discussion features affect student participation? Does a graded discussion facilitate a higher participation rate? When discussions are set to Require Initial Post (post before reading other posts), are there significant differences in student participation? 	<ul style="list-style-type: none"> Which forums generate the most posts and why? Which discussion boards generate the most traffic – have more students' views? Who are the learners actively engaged by providing many comments to peers' postings? Who are the learners whose initial thread became so popular that received quite a number of replies? 	<ul style="list-style-type: none"> How much do students use discussions as an extension of F2F classroom participation? Does discussion interaction reflect students' participation in class activities or subgroups of students with common interests in reality? Should I promote "burstiness" in asynchronous discussions, or longer, more sustained reflection?
Assignments, Quiz submission data	<ul style="list-style-type: none"> Is there a relationship between quiz performance and content access, or overall activities in a LMS? Why do students access a particular material more times than other materials? How might we alter prompts or instructions given to better ensure that course goals are met? 	<ul style="list-style-type: none"> What was the overall performance on a quiz? How well an individual student did in comparison to the entire class? Did students struggle with a specific material? Could additional help (intervention) or materials be provided to students that struggle? 	<ul style="list-style-type: none"> Are student submission activities associated with due dates? How well an individual student did in comparison to the entire class? Why is a particular student struggling?

Fig. 1.13 Key questions to help us identify the needed data

When you analyse and design any course you need to gather the questions that are related to your instructional design, your teaching and tutoring strategy and your learners' support:

- What is my Target Audience (whose instructional needs are to be addressed)?
- What are the Learning Environment Characteristics (educational context, limiting factors, affordances and constraints, technical requirements)?
- What criteria will be used to assess the achievement of the expected learning outcomes by the learners?

When the course is up and running:

- What is the learners' activity in the Online Environment?
- What are the tutors' support activities (scaffolding, feedback, answering questions, stimulating engagement)?
- How the combination of the learners' and tutors' activities relate to the academic performance, motivation and/or engagement?

Figure 1.13 summarizes some key questions to help us identify the needed data.

Now that we have the right questions in place, we can identify the type of data that may help us find the answers we are looking for. As suggested by Fig. 1.14 and this [infographic](#) by the Data Quality Campaign project, the types of educational data commonly used can be classified in two types: Static and Dynamic Data.

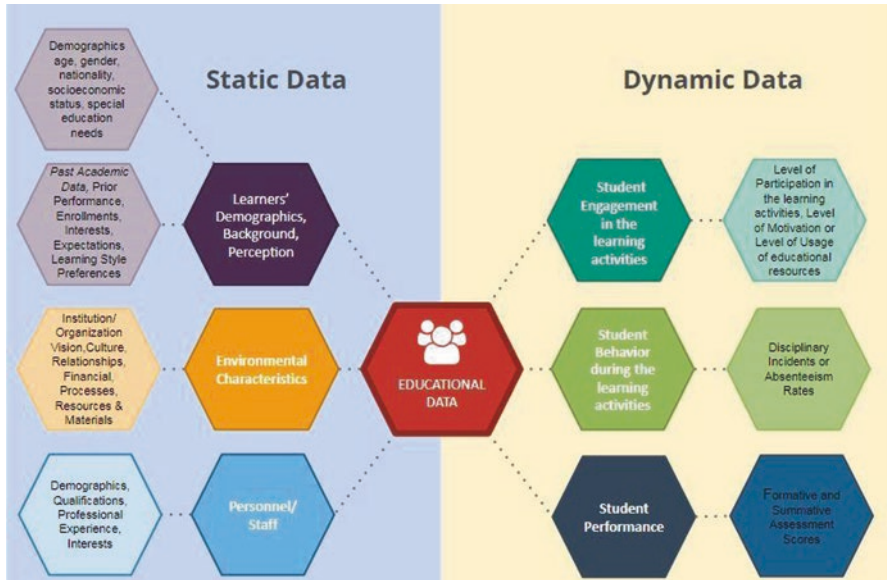


Fig. 1.14 Static and dynamic educational data

Static data, refers to data which can remain unchanged for large periods of time. According to Shacklock (2016), it is the data “which is collected, recorded and stored by institutions and traditionally includes student records, staff data, financial data and estates data”.

As Shacklock (2016), points out “Static data has always been a strategic asset for both institutions and government. It informs all operational and business decision-making and planning in an institution, and indicates to government and the public how the sector is performing as a whole.”

Dynamic data refers to data generated at a more frequent rate and they are mainly related to **learners’ activities during the learning process**. Such data is usually collected by the e-tutors, classroom teachers typically through Learning Management Systems.

If we manage to collect, link and analyse dynamic data, then we can probably get an instant, accurate view of how an individual learner or a group of learners is performing.

Lai and Schildkamp (2013, p. 11–12) have extended Ikemoto and Marsh’s (2007) categories of educational data, to input data, context data, process data and outcome data. Each category indicates when data will be collected. Figure 1.15 presents examples of educational data for each category.

To get a better understanding of the use of data to strengthen lifelong learning, you may watch the video (in the useful video resources) from UNESCO “[Data for Lifelong Learning](#)“, presenting an overview of the tools developed by the UNESCO Institute for Statistics (UIS) to measure learning and improve learning outcomes.

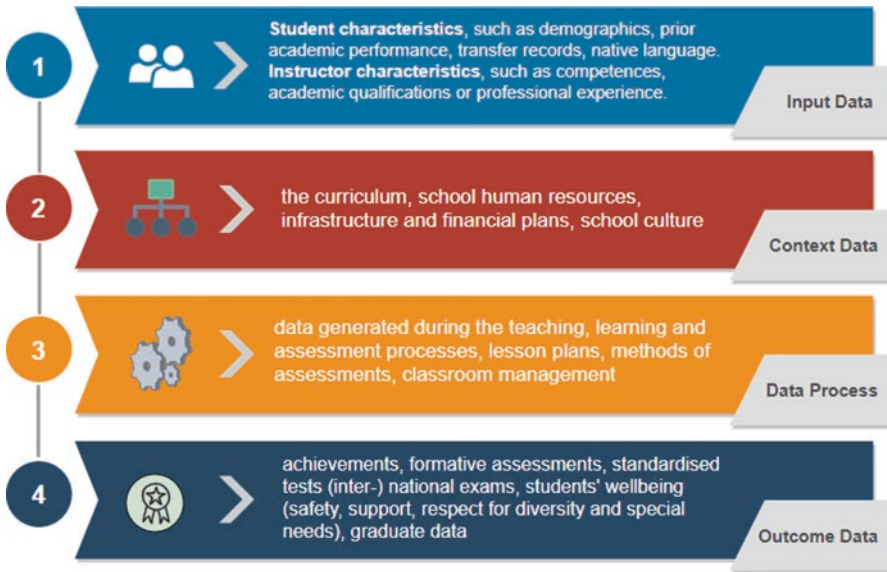


Fig. 1.15 Examples of educational data for each category

Questions and Teaching Materials

1. Please select the correct answer.

Do you remember Alice? Alice is an English teacher who has just been appointed in an Experimental High School, in Athens, Greece. She is responsible for the English Course of Class1 and Class2 of the ninth grade. Her principal has encouraged her to use student data to gain insights and prepare her instruction accordingly, so as to improve this year's Grade 9 students' achievement.

Alice is studying the categories of data and realizes that we can distinguish two major categories of data, qualitative and quantitative data.

In a school setting, quantitative data may include notes from classroom observations.

Do you agree with the assumption of Alice?

- Yes
- No

Correct answers: No

2. Please select the correct answer.

Alice realizes that the first key to be data literate is Educational Data Collection. To achieve her goals, she needs to utilize a variety of data types from different sources and use various methods to process and analyse them.

Do you agree with the assumption of Alice?

- Yes
- No

Correct answers: Yes

3. Alice is a bit confused about Navigation and click stream data. What information could this data reveal?

Help Alice select the correct answer(s):

- A. The most popular paths through the course site
- B. Time spent in the online learning environment
- C. Quiz performance
- D. Critical dropout points
- E. All the above

Correct answers: A, B, D

4. Alice starts posing questions to identify and collect the appropriate educational data. She asks herself “Why do I need the data?”, “What data is needed?” “Where are data located?” “How will data be collected?”

Alice decides to gather a variety of students’ data, including demographics, perception data, past academic performance, last year’s achievements and formative assessments for English lesson and other courses, as well as the regional (district’s) performance over the past 5 years.

Alice understands that static data remain unchanged for large periods of time, while dynamic data is generated at a more frequent rate.

Help Alice match the data from the left column to the appropriate type from the right column (static or dynamic).

1. Personnel’s professional experience	A. STATIC DATA
2. Student performance	
3. Absenteeism rates	
4. Engagement in the discussion forum	
5. Learner’s background	B. DYNAMIC DATA
6. Prior performance	
7. Level of usage of educational resources	
8. Resources and materials	

Correct answers: 1-A, 2-B, 3-B, 4-B, 5-A, 6-A, 7-B, 8-A

5. According to Lai and Schildkamp (2013) educational data can be categorized as:
- input data
 - context data
 - process data and
 - outcome data.

Alice wants to make effective instructional changes to her reading program to better cater for the boys in her class. She is considering using the following data:

- Data on student characteristics such as absenteeism rates for boys.
- Analysis of student performance on reading tests.
- Discussions with the boys about their strengths and weaknesses in reading and their love of reading.
- Examination of the school curriculum to determine whether the reading texts are engaging for boys.

Help Alice to match the data [1 to 4] with the categories of the data [a to d] mentioned above.

Correct answer: 1a – 2d – 3c – 4b

6. Please select the correct answer.

After watching the video “[Data for Lifelong Learning](#)”, from the UNESCO Institute for Statistics, Alice realizes that “robust monitoring is needed to track whether children and adults are gaining the skills they need to thrive in today’s world.”

Do you agree with the assumption of Alice?

- Yes
- No

Correct answer: Yes.

7. **ACTIVITY/PRACTICE QUESTION (Reflect on)**

We encourage you to elaborate on your response about the use of educational data in the following reflective task. You may reflect on:

- As an instructional designer or a school teacher, you want to collect data to redesign your course. Describe your evaluation plan. Define the questions you need to answer and the data you will need to collect. Please share either your past experience or your thoughts for future actions.



Fig. 1.16 Difference between reliability and validity

- As a tutor of an online course, you want to collect data to enhance your learners’ participation in the course. Define the questions you need to answer and the data you will need to collect. Please focus on either your past experience or your thoughts for future actions.

1.3.2 Matching Appropriate Educational Data with Data Sources

In this section we will discuss **where** to find the educational data you need and **how**.

WHERE applies to the location where you might have to go for the data collection, according to the data you need.

There are **numerous data sources of learners’ information available**:

- data stored in institutional student information systems, e.g. high school grades, socio-economic status, citizenship and immigration status, parents’ education and language skills,
- trace data recorded within Learning Management Systems and other online learning environments such as e-libraries and virtual labs,
- data from systems that analyse discussion in online forums,
- survey data (e.g., questionnaires) (Fig. 1.16).

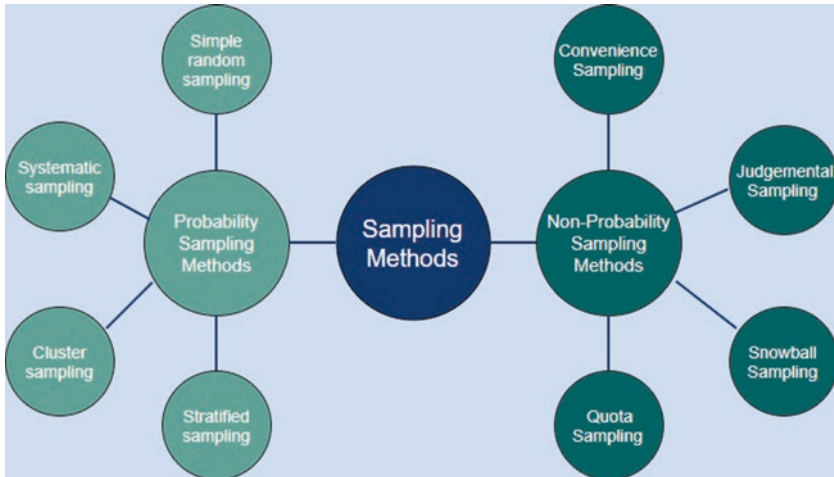


Fig. 1.17 Different sampling methods

By EGalvez (WMF) – Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=31697223>

Before proceeding further with data collection, we need to agree on a few basic concepts related to the nature of data itself. As Guerra-López (2008) points out, data must meet three basic characteristics:

- **Relevancy:** The data must directly relate to the research questions being answered.
- **Reliability:** The data must be measured, trustworthy, and consistent.
- **Validity:** The data must measure what we intend to measure.

Another important question is which methods will be used to select a representative group of people from the learners’ target audience? How can we avoid **biases** in sampling?

Rothwell et al. (2016) argue that the four types of sampling procedures commonly used are: (1) convenience or judgmental sampling, (2) simple random sampling, (3) stratified sampling, and (4) systematic sampling. To determine which one to select, we need to consider our goals and objectives, certainty needed in the conclusions, the willingness of decision makers in the organisation to allow information to be collected for our study, and the resources (time, money, and staff) available (Rothwell et al., 2016) (Fig. 1.17).

Let’s have a closer look at this important aspect affecting our data, biases. **Biases** are systematic errors induced from the data collection process employed, reducing potential biases allows us to have data that represent the population. This “**Bias when collecting data**” video (in the useful video resources) explains different kind of biases that occur when collecting data.

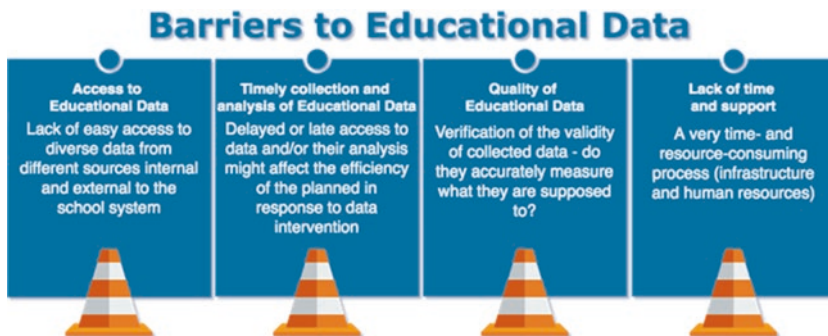


Fig. 1.18 Barriers to educational data

- **Voluntary bias** occurs when the responders choose themselves to participate.
- **Undercoverage bias** occurs when some members of the population are inadequately represented in the sample.
- **Overcoverage bias** occurs when some members of the population are overrepresented in the sample.
- **Non-response bias or participation bias** occurs when the sample is unwilling to participate.
- **Convenience sample bias or availability bias** occur when collecting the data that is easier to obtain, rather than collecting more relevant data.
- **Response bias** occurs when respondents provide dishonest or misleading answers, due to many reasons such as survey design, survey fatigue, missing answers etc (Fig. 1.18).

As seen in Fig. 1.18, access to educational data may be a really serious barrier to overcome when gathering appropriate educational data. Here are some additional questions that we need to answer. What data do we need versus what data can we access? With whom in the organisation should we interact during our data collection process? How many people? For what issues? Whose approval is necessary to collect information?

Perhaps the most common failure during the collection process is failing to receive enough—or the right—permissions to collect data. To overcome this problem, we should make sure we have secured all necessary approvals before collecting data.

Failure to complete this step successfully can create significant, and often unfortunate, barriers to cooperation within the organisation.

In their 2006 report, [Making Sense of Data-Driven Decision Making in Education](#), Julie Marsh and her colleagues (Marsh et al., 2006, p. 9) identified a number of barriers to the effective and efficient take-up of educational data use.

Questions and Teaching Materials

1. **Alice wants to collect the following data, in order to study how her students' academic achievement is related to their learning behaviour.**

Help Alice decide which of this data is stored in the institutional Student Information.

System (select all that apply):

- A. **Grades.**
- B. **Demographics.**
- C. **Clickstream data.**
- D. **Number of posts in the discussion forum.**
- E. **Parents education.**

Correct answers: A, B, E

2. **Alice wants to understand the difference between validity and reliability, so she asks Steven, a colleague. Steven uses the following example to explain validity and reliability in datasets.**

“You want to measure students' intelligence so you ask students to do as many push-ups as they can every day for a week. After one week, you find that each student did approximately the same number of push-ups on each day. What is the data you collected, in terms of validity and reliability?”

Help Alice find the right answer (you may also refer to <https://www.thegraidenetwork.com/blog-all/2018/8/1/the-two-keys-to-quality-testing-reliability-and-validity>). The data collected is:

- A. **Both valid and reliable**
- B. **Neither valid nor reliable**
- C. **Valid but not reliable**
- D. **Reliable but not valid**

Correct answer: D

3. **Alice has decided to apply the flipped classroom strategy with her students using the school's Learning Management System. This inverted model “flips” the traditional order of class activities so that school work is done at home, and “homework” is done at school.**

Before using the flipped classroom initiative, Alice wants to study students' perceptions of technology. She prepares a questionnaire and uses students from her class as a sample for her study. She wants to generalize her findings to all High-School students.

The above procedure is an example of:

- A. convenience sampling
- B. simple random sampling
- C. systematic sampling
- D. snowball sampling

Correct answer: A.

4. Alice is excited. There are so many different kinds of bias that occur when collecting data.

Alice wonders if she understood the different categories of bias. Can you help

HAGAR the Horrible



The data collection process described in the above comic strip with Hagar the Horrible is an example of:

- A. Non response bias
- B. Voluntary bias
- C. Response bias
- D. Convenience sample bias

Correct answer: C

5. Alice contacts Mr. Adams, who is appointed as school's Data Protection Officer (DPO), to secure all necessary approvals for the sources handled by her school or by the corresponding district. As soon as Alice signs the required data protection consent form, she gets permission and downloads the datasets from the various sources. Alice also requests that she be granted access to the LMS used by the school (a new teacher account is created by the LMS administrator).

Without the availability of high-quality data and perhaps technical assistance, data may become misinformation or lead to invalid inferences. Delayed or late access to data and/or its analysis might affect the efficiency of the planned activities in response to the data intervention.

This kind of data barrier is referred as:

- A. **Lack of easy access to educational data**
- B. **Lack of timely collection and analysis of educational data**
- C. **Poor quality of educational data**
- D. **Lack of time and support**

Correct answer: B.

6. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to elaborate on your response about data collection, in the following reflective task. You may reflect on:

1. **How can you as an instructional designer or tutor avoid biases in educational data collection? Please share either your past experience or your thoughts for future actions.**
2. **You are an instructional designer or tutor or school teacher and you want to collect and analyse educational data from your course discussion forum to evaluate learners' participation. Describe your evaluation plan.**

1.3.3 *Combining Data from Different Educational Data Sources*

As we have already seen, there are many different data sources that contain useful educational data. Shacklock (2016) reports that “*some institutions are beginning to explore the possibility of incorporating more types of data into their analytics systems. The University of Lancaster is considering capturing and using data on which students are accessing library PCs and for how long, NTU are also looking at capturing data on e-book usage*”.

Figure 1.19 summarizes indicative educational data sources that store data from various sources:

- **Internal Data Sources**
- **Online Learning Courses**
- **Surveys and polls**
- **Cloud applications**
- **Social media**
- **External data sources, like open repositories**

Once we decide upon the “*Ws*” of the data we need, we have to define *HOW* we will collect the data. School of Data (2013) distinguishes three basic ways of getting hold of data:

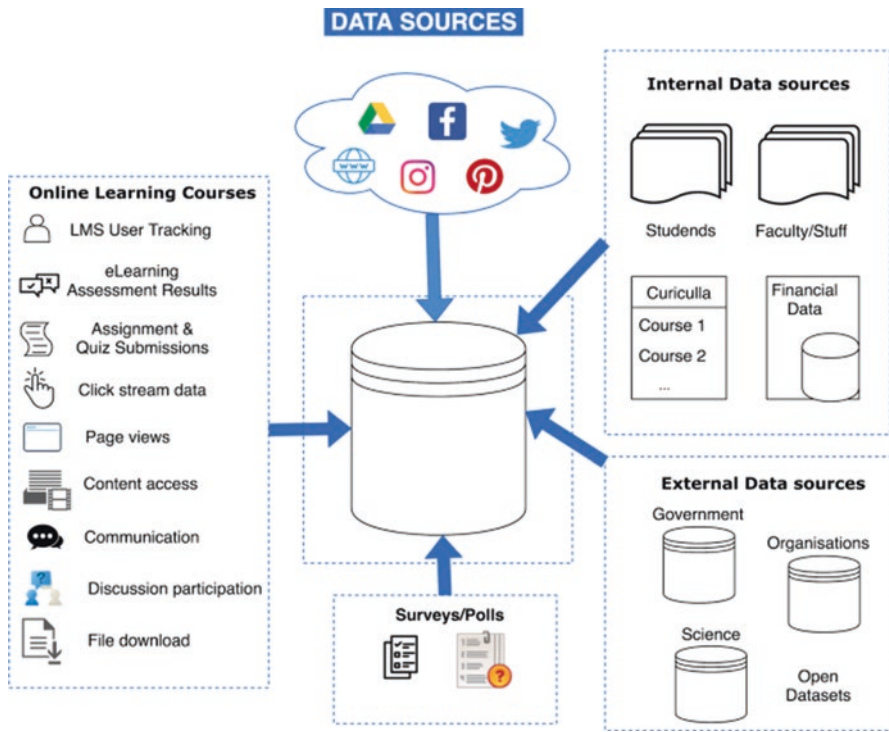


Fig. 1.19 Indicative educational data sources

1. **Finding data** – this involves searching and finding data that has already been released e.g. through [open data repositories](#).
2. **Getting hold of more data** – asking from official sources to release ‘new’ data, e.g. through [Freedom of Information requests](#).
3. **Collecting data, yourself** – This means gathering data through:
 - surveys and polls
 - internal data sources, like Institutions’ Management Information Systems and/or Students Information Systems.
 - online educational environments, such as LMSs, MOOCs, ITSs which record any learner activity involved, such as reading, writing, taking tests, performing various tasks and commenting with peers (Fig. 1.20).

There are various sources of educational data where data is stored in different formats. Romero et al. (2014) state that “*the goal of data aggregation/ integration is to group together data from multiple sources into a coherent recompilation, normally into a database*”.

Aggregation is the process of grouping together same type of data from different organisations/institutions and **integration** is the process that groups different types of data from the same organisation/institution.

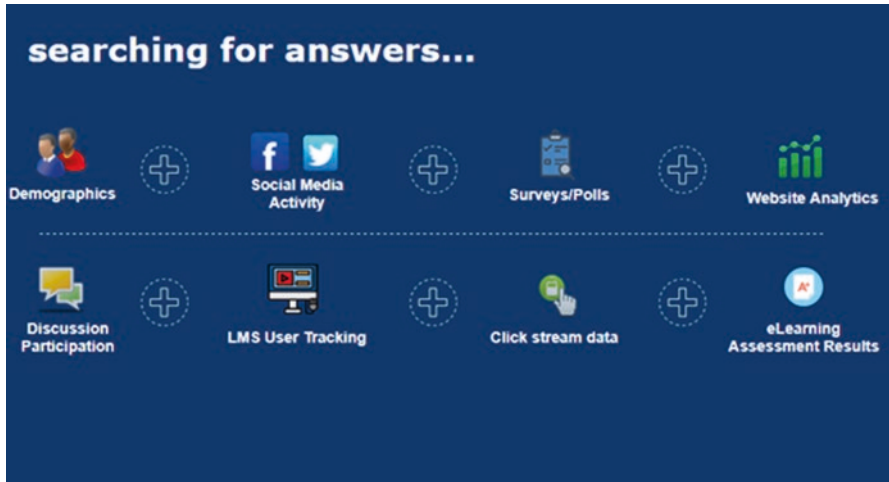


Fig. 1.20 Searching for answers

Using aggregation and integration we can combine data from different sources and in different formats, for example performance data, attendance records, past academic data and forum participation data into a single database.

Questions and Teaching Materials

- Alex, the principal of Alice’s school, is now interested in studying the effects of the free school breakfast programme on children’s attendance and academic achievement. He poses some questions and asks Alice to help him find the appropriate data.**

Match the items in column A with column B and choose the correct answer to help Alice find the appropriate data collections.

Column A: Data needed	Column B: Where to look for data
1. National/regional school breakfast program participation	A. Finding data in government or open data repositories and/or asking from official sources to release ‘new’ data
2. Institutional school breakfast program participation	
3. National/regional students’ academic achievements	B. Online educational environments
4. Institutional students’ academic achievements	
5. National/regional students’ absenteeism data	C. Surveys and polls
6. Institutional students’ absenteeism data	D. Institutions’ management information systems

Column A: Data needed	Column B: Where to look for data
7. Participation in online learning activities	E. Students information systems
8. Dietary attitudes	

Correct answer: 1A – 2D – 3A – 4E – 5A – 6E – 7B – 8C

2. **Alice wants to retrieve students’ information from the Student Information System and combine it with data from the Learning Management System, in order to study the factors that affect student’s participation in the online course she prepared. The process that groups different types of data from the same organisation/institution is called:**
- A. **Aggregation**
 - B. **Integration**

Correct answer: B.

3. **ACTIVITY/PRACTICE QUESTION (Reflect on)**

We encourage you to elaborate on your response in the following reflective task. You may reflect on:

1. **You are an instructional designer and you need to collect data in order to redesign your online course. Describe your plan to ask permission to collect the appropriate data. You can use [Freedom of Information request](#) (The Freedom of Information Act (FOIA) gives you the right to access recorded information held by [public sector organisations](#))**
2. **You are an instructional designer or tutor or school teacher and you want to improve learners’ retention. How can you benefit from combining different data sources?**

1.4 Concluding Self-Assessed Assignment

1.4.1 Introduction

Now that you have a better understanding of the power of educational data as a key success factor for online and blended teaching as well as of the fundamentals of Educational Data Collection, you are ready to link theory with practice and apply

your Educational Data Literacy Competences focusing on the creation of a Data Collection plan.

In order to proceed, you are requested to complete a concluding self-assessed assignment. This self-assessed assignment is a real life scenario activity (based on the use case of our teacher Alice), using a rubric across three proficiency levels and an exemplary solution rating. When you have completed this assignment, you will assess it yourself, following the rubric which will list the criteria required and give guidelines for the assessment.

This self-assessed assignment procedure consists of 5 steps:

- **Step 1.** Real life scenario
- **Step 2.** Getting familiar with the assessment rubric
- **Step 3.** Prepare your answer
- **Step 4.** Review a sample solution
- **Step 5.** Self-evaluate your answer

1.4.2 Step 1. Real Life Scenario

Let's go back to Alice. As introduced, Alice has decided to apply the flipped classroom strategy with her students using the school's Learning Management System. She wants to use educational data to reveal insights about her course design and students' activities, to reflect on her teaching practices and to target accordingly her teaching and learning interventions so as to help every student to excel and meet their educational goals.

To inform her decisions and benefit from them, Alice needs to **access and gather the appropriate data**. She should know why she collects this data, what types of data she needs to collect, when and how to get it and where to find it.

You need to help Alice to design her Educational Data Collection plan, in order to monitor students' performance in the online course, for the flipped classroom initiative.

1.4.3 Step 2. Getting Familiar with the Assessment Rubric

Alice has already prepared an Initial Educational Data Collection Plan and asks you to evaluate it, using the Rubric for assessing the Educational Data Collection Plan, to identify potential issues.

ACTIVITY/PRACTICE QUESTION (Reflect on) We encourage you to elaborate on your response about the evaluation of the Initial Educational Data Collection Plan created by Alice, in the following reflective task. You may reflect on:

1. *Does this Educational Data Collection Plan addresses student's performance aspects?*

- If not, what would you advise Alice to add/modify, so that the data collection plan addresses all aspects and monitor student’s performance, during the flipped classroom initiative?

1.4.3.1 Initial Educational Data Collection Plan

Evaluation questions (Why?)	Indicators (What?)	Data source (Where?)	Time of collection (When?)	Collection methods (How?)
Online presence	Connection frequency	LMS	End of year	In class survey
Performance	Final exams scores	Region database	End of year	SIS
Activities	Activities completed	LMS	End of year	Online survey

1.4.3.2 Rubric for Assessing the Educational Data Collection Plan

Criteria	1 Unacceptable	3 Good/Solid	5 Exemplary
Determine appropriate, clearly defined evaluation questions	Evaluation questions are not relevant or adequately defined and address only one or two aspects of student’s performance.	Evaluation questions are relevant, adequately defined and address several aspects of student’s performance.	Evaluation questions are relevant, well defined and address almost every aspect of student’s performance.
Determine what data must be collected in order to answer the each evaluation question (indicators)	Most of the indicators are not defined or are not relevant with the respective performance evaluation question.	Relevant indicators are defined for most of the evaluation questions.	Multiple relevant indicators are defined for each evaluation question.
Determine the data sources (where to find data)	Data sources are not defined for most of the evaluation questions.	Most of the data sources defined for The evaluation questions are relevant.	All data sources defined for the evaluation questions are relevant.
Determine when to collect the data	Collection time is not defined for most of the evaluation questions.	Collection time is defined for most of the evaluation questions, but not always correctly.	Collection time is defined correctly for each of the evaluation questions.
Determine how to collect the data (collection methods)	Collection methods are not defined for most of the evaluation questions.	Collection methods are adequately defined for most of the evaluation questions.	Appropriate collection methods are defined for all the evaluation questions.

1.4.4 Step 3. Prepare Your Answer

Please assist Alice in preparing an Educational Data Collection plan to monitor her students' performance in the online course for the flipped classroom initiative, by completing the following matrix.

Evaluation questions (Why?)	Indicators (What?)	Data source (Where?)	Time of collection (When?)	Collection methods (How?)
e.g. online presence	e.g. connection frequency, Time spent, etc.	LMS	e.g. end of each month	System reports
Add rows...				

ACTIVITY/PRACTICE QUESTION (Reflect on) We encourage you to elaborate on your response about the preparation of the Educational Data Collection plan to monitor students' performance in the online course, in the following reflective task. You may reflect on:

1. *How should the Educational Data Collection plan be formulated so that Alice can obtain useful insights for student's performance, to inform her decisions and benefit from them?*
2. *What are the main questions to ask for evaluating student's performance in online courses and what data could be useful indicators to address these questions?*

1.4.5 Step 4. Review a Sample Solution

Please review a sample of an **Exemplary solution** that follows the criteria specified in the **Rubric for assessing the Educational Data Collection Plan**.

ACTIVITY/PRACTICE QUESTION (Reflect on) We encourage you to elaborate on your response about the Exemplary solution that follows the criteria specified in the Rubric for assessing the Educational Data Collection Plan, in the following reflective task. You may reflect on:

1. *Do you identify any requirements that you did not take under consideration when creating your Educational Data Collection Plan?*

1.4.5.1 Exemplary Sample Solution

Educational Data Collection plan to monitor students' performance in the English Language online course, for the flipped classroom initiative of the ninth Grade of Athens Experimental High School.

Evaluation questions (Why?)	Indicators (What?)	Data source (Where?)	Time of collection (When?)	Collection methods (How?)
Online presence	Connection frequency, Time spent online, date of last login	LMS	End of week	System reports
Content study	Number of lessons read, latency, downloaded PDFs, viewed resources, time spent on lessons and resources	LMS	End of month	System reports
Activities done	Number of activities done for each type (quiz, assignment)	LMS	End of week	System reports
Activities result	Average grade, average last grade, average best grade, final exams result	LMS	End of week	System reports
Social activities	Topics read, answers posted, new topics created, time spent on the forums	LMS	End of month	System reports

Source: Clustering moodle data as a tool for profiling students Bovo et al. (2013)

1.4.6 Step 5. Self-Evaluate Your Answer

Now that you have seen the Exemplary Sample Solution, please rate your initial answer (evaluate the *Educational Data Collection Plan* you created), using the Rubric table below.

Criteria	1 Unacceptable	3 Good/Solid	5 Exemplary
Determine appropriate, clearly defined evaluation questions	Evaluation questions are not relevant or adequately defined and address only one or two aspects of student's performance.	Evaluation questions are relevant, adequately defined and address several aspects of student's performance.	Evaluation questions are relevant, well defined and address almost every aspect of student's performance.
Determine what data must be collected in order to answer the each evaluation question (indicators)	Most of the indicators are not defined or are not relevant with the respective performance evaluation question.	Relevant indicators are defined for most of the evaluation questions.	Multiple relevant indicators are defined for each evaluation question.

Criteria	1 Unacceptable	3 Good/Solid	5 Exemplary
Determine the data sources (where to find data)	Data sources are not defined for most of the evaluation questions.	Most of the data sources defined for The evaluation questions are relevant.	All data sources defined for the evaluation questions are relevant.
Determine when to collect the data	Collection time is not defined for most of the evaluation questions.	Collection time is defined for most of the evaluation questions, but not always correctly.	Collection time is defined correctly for each of the evaluation questions.
Determine how to collect the data (collection methods)	Collection methods are not defined for most of the evaluation questions.	Collection methods are adequately defined for most of the evaluation questions.	Appropriate collection methods are defined for all the evaluation questions.

References

- Barnes, F. D. (2004). *Inquiry and action: Making school improvement part of daily practice*. Annenberg Institute for School Reform, Brown University.
- Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and Learning through educational data mining and Learning analytics: An issue brief*. U.S. Department of Education, Office of Educational Technology.
- Bovo, A., Sanchez, S., Heguy, O., & Duthen, Y. (2013). Clustering moodle data as a tool for profiling students. In *2013 Second international conference on E-Learning and E-Technologies in Education (ICEEE)* (pp. 121–126).
- Brame, C. J. (2013). *Flipping the classroom*. Vanderbilt University Center for Teaching. Retrieved from <http://cft.vanderbilt.edu/guides-sub-pages/flipping-the-classroom/>
- Data Quality Campaign. (2014a). *Roadmap for educator licensure policy addressing data literacy*. Retrieved from <https://dataqualitycampaign.org/resource/roadmap-educator-licensure-policy-addressing-data-literacy/>
- Data Quality Campaign. (2014b). *Teacher data literacy: It's about time. A brief for state policymakers*. Retrieved from <http://2pido73em67o3eytaq1cp8au.wpengine.netdna-cdn.com/wp-content/uploads/2016/03/DQC-Data-Literacy-Brief.pdf>
- Díaz, A., Rowshankish, K., & Saleh, T. (2018, September). Why data culture matters. *McKinsey Analytics*. Retrieved from <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/why-data-culture-matters>
- Education Elements. (2018). *Personalised learning: The present and future of education*. Retrieved from <https://www.edelements.com/personalized-learning>
- Freeman, A., Adams Becker, S., Cummins, M., Davis, A., & Hall Giesinger, C. (2017). *NMC/CoSN horizon report: 2017 K–12 edition*. The New Media Consortium.
- Greany, K., & Niles-Hofmann, L. (2018). An everyday guide to learning analytics. How to use data to decode the digital body language of your audience and design elearning that works. *Elucidat*. Retrieved from <https://elearningindustry.com/free-ebooks/everyday-guide-to-learning-analytics>
- Guerra-López, I. (2008). *Performance evaluation: Proven approaches for improving program and organizational performance*. Jossey Bass.
- IBM. (2019). Extracting business value from the 4 V's of big data. IBM Big Data & Analytics Hub. Retrieved from https://www.researchgate.net/figure/The-Four-Vs-of-Big-Data-and-Its-Value-IBM-2019_fig1_344303432

- Ikemoto, G. S., & Marsh, J. A. (2007). Chapter 5 cutting through the "data-driven" mantra: Different conceptions of data-driven decision making. *Yearbook of the National Society for the Study of Education*, 106(1), 105–131. <https://doi.org/10.1111/j.1744-7984.2007.00099.x>
- Inspectorate, Department of Education and Skills. (2016). *School self-evaluation guidelines 2016–2020 primary*. Department of Education and Skills.
- Jisc. (2018, September 10). Canterbury Christ Church University. Jisc digital experience insights. Retrieved from <https://digitalinsights.jisc.ac.uk/case-study-listing/canterbury-christ-church-university/>
- Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood, K. (2011). *The 2011 horizon report*. The New Media Consortium.
- Lai, M. K., & Schildkamp, K. (2013). Data-based decision making: An overview. In K. Schildkamp, M. K. Lai, & L. Earl (Eds.), *Data-based decision making in education: Challenges and opportunities*. Springer.
- Long, P. D., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review*, 46(5), 31–40.
- Love, N. (2012). Data literacy for teachers. *Hawker Brownlow Education*. Retrieved from <http://files.hbe.com.au/samplepages/NPR8884.pdf>
- Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, 47(2), 71–85. <https://doi.org/10.1080/00461520.2012.667064>
- Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*, 42(1), 30–37. <https://doi.org/10.3102/0013189X12459803>
- Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education: Evidence from recent RAND research*. RAND Corporation.
- Obsidian Learning. (2018, October 16). ADDIE Model. Retrieved from <https://www.obsidian-learning.com/projects/addie-model/>
- Panopto. (2015, January, 12). *Infographic: The growth in flipped learning*. Retrieved from <https://www.panopto.com/blog/growth-in-flipped-learning-infographic/>
- Pettey, C. (2018, March 5). *How data and analytics leaders learn to master information as a second language*. Retrieved from <https://www.gartner.com/smarterwithgartner/gartner-keynote-do-you-speak-data/>
- Reiser, R. A. (2001). A history of instructional design and technology: Part I: A history of instructional media author(s). *Educational Technology Research and Development*, 49(1), 53–64.
- Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., & Wuetherick, B. (2015). *Strategies and best practices for data literacy education: Knowledge synthesis report*. Dalhousie University. <https://doi.org/10.13140/RG.2.1.1922.5044>
- Romero, C., Romero, J. R., & Ventura, S. (2014). A survey on pre-processing educational data. In *Educational data mining studies in computational intelligence* (pp. 29–64). Springer. https://doi.org/10.1007/978-3-319-02738-8_2
- Rothwell, W. J., Bencotter, G. M., King, M., & King, S. B. (2016). *Mastering the instructional design process: A systematic approach*. John Wiley & Sons. Retrieved from <https://www.wiley.com/en-us/Mastering+the+Instructional+Design+Process%3A+A+Systematic+Approach%2C+5th+Edition-p-9781118947142>
- School of Data. (2013). School of data handbook. Retrieved from <https://schoolofdata.org/handbook/>
- Sclater, N., Peasgood, A., & Mullan, J. (2016). *Learning Analytics in Higher Education: A review of UK and international practice. Full report*. Jisc.
- Shacklock, X. (2016). *From bricks to clicks: The potential of data and analytics in higher education*. Policy Connect – Higher Education Commission.

Useful Video Resources

- External Video: What is Big Data and how does it work? [1:33].
- External Video: Big Data's Making Education Smarter [2:16].
- External Video: Data is Power [2:32]
- External Video: Engaging with students to build a better digital environment [2:56]
- External Video: Educause: What Is Personalized Learning?[2:27].
- External Video: Using big data to customize learning[3:08].
- External Video: Data Driven Learning Design [5:35].
- External Video: Data: It's Just Part of Good Teaching [3:43].
- External Video: How Data Help Teachers[1:51].
- External Video: Data Can Help Every Student Excel [2:00].
- External Video: The Flipped Classroom Model [3:00].
- External Video: Why Personalized Learning: 4 Stories from 4 School Districts [3:41]
- External Video: Develop a culture of data and analytics enablement at the summit [1:30].
- External Video: Data for Lifelong Learning [2:49].
- External Video: Bias when collecting data [5:59].

Further Reading

- Baker, R. S., & Inventado, P. S. (2016). Educational data mining and learning analytics: Potentials and possibilities for online education. In G. Veletsianos (Ed.), *Emergence and innovation in digital learning* (pp. 83–98). <https://doi.org/10.15215/aupress/9781771991490.01>
- Brown, M. (2012). *Learning analytics: Moving from concept to practice*. EDUCAUSE Learning Initiative.
- Centre for Education Statistics and Evaluation. (2018). *Using data with confidence*. NSW Department of Education.
- Chapman, C., & Sammons, P. (2013). *School self-evaluation for school improvement: What works and why?* Project Report. CfBT Education Trust, Reading.
- Ellis, R. A., Han, F., & Pardo, A. (2017). Improving learning analytics – Combining observational and self-report data on student learning. *Educational Technology & Society*, 20(3), 158–169.
- Friesen, N. (2013). Learning Analytics: Readiness and Rewards/L'analyse de l'apprentissage: état de préparation et récompenses. *Canadian Journal of Learning and Technology/La revue canadienne de l'apprentissage et de la technologie*, 39(4), 1–12.
- Gendron, S. A., & Traub, S. (2015). *Data: The guide on the path to personalized Learning for ALL students*. International Center for Leadership in Education.
- Goldring, E. B., & Berends, M. (2009). The importance of data-driven decision making. In *Leading with data: Pathways to improve your school*. Corwin Press.
- Government Digital Service. (2017, April 05). How to make a freedom of information (FOI) request. Retrieved from <https://www.gov.uk/make-a-freedom-of-information-request>
- Grant, P., & Basye, D. (2014). *Personalized Learning: A guide for engaging students with technology*. International Society for Technology in Education.
- Hamilton, L., Halverson, R., Jackson, S. S., Mandinach, E., Supovitz, J. A., & Wayman, J. (2009). *Using student achievement data to support instructional decision making (NCEE 2009–4067)*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- HM Government. (2013). *Seizing the data opportunity: A strategy for UK data capability*. BIS.
- Hora, M. T., Bouwma-Gearhart, J., Park, H. J. (2014). Exploring data-driven decision-making in the field: How faculty use data and other forms of information to guide instructional decision-

- making (WCER working paper no. 2014-3). Retrieved from <https://files.eric.ed.gov/fulltext/ED556492.pdf>
- Houston, P. D. (2002). *Using data to improve schools: What's working*. American Association of School Administrators.
- Johnson, L., Adams, S., Cummins, M., Estrada, V., Freeman, A., & Ludgate, H. (2013). *The NMC horizon report: 2013 higher. Education Edition*. The New Media Consortium.
- Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2015). *NMC horizon report: 2015 K-12 edition*. The New Media Consortium.
- Khalil, M., & Ebner, M. (2015). *Learning analytics: Principles and constraints*. <https://doi.org/10.13140/RG.2.1.1733.2083>
- Lancy, D. (2001). *3D data management: Controlling data volume, velocity, and variety*. META Group.
- Lewis, J. (2014). Next gen tools. Strategies and Innovations for Implementing Breakthrough Models. Blended Learning Classroom Design. EDUCAUSE. Retrieved from <https://s3.amazonaws.com/nglc/resource-files/NGB1405-pdf.pdf>
- Mandinach, E. (2013, September 17). Data literacy vs. assessment literacy. [Blog post]. Retrieved from <https://www.msdf.org/blog/2013/09/ellen-mandinach-data-literacy-vs-assessment-literacy/>
- Marsh, J. A., & Farrell, C. C. (2015). How leaders can support teachers with data-driven decision making: A framework for understanding capacity building. *Educational Management Administration & Leadership*, 43(2), 269–289. <https://doi.org/10.1177/1741143214537229>
- Martin, J. (2010, August 4). Aligning data collection with school mission: Preparing for DoE ONPE Panel. [Blog post]. Retrieved from <https://21k12blog.net/2010/08/04/aligning-data-collection-with-school-mission-preparing-for-doe-onpe-panel/#more-1913>
- Matters, G. (2006). *Using data to support learning in schools: Students, teachers, systems*. ACER Press.
- McIntosh, K., Horner, R. H., & Sugai, G. M. (2009). Sustainability of systems-level evidence-based practices in schools: Current knowledge and future directions. In W. Sailor, G. Dunlap, G. M. Sugai, & R. Horner (Eds.), *Handbook of positive behavior support* (pp. 327–352). Springer. https://doi.org/10.1007/978-0-387-09632-2_14
- Means, B., Chen, E., DeBarger, A., & Padilla, C. (2011). *Teachers' ability to use data to inform instruction: Challenges and supports*. Office of Planning. Retrieved from <https://files.eric.ed.gov/fulltext/ED516494.pdf>
- MOEduSAIL – Educational Systems and Instruction for Learning. (2019). Data-based decision making for teacher teams. Retrieved from <http://www.moedu-sail.org/dbdm-materials/>
- Mor, Y., Ferguson, R., & Wasson, B. (2015). Editorial: Learning design, teacher inquiry into student learning and learning analytics: A call for action: Learning design, TISL and learning analytics. *British Journal of Educational Technology*, 46(2), 221–229. <https://doi.org/10.1111/bjet.12273>
- Next Generation Learning Challenges. (2017). New designs for school. What does next generation learning look like in practice? EDUCAUSE. Retrieved from <http://nextgenlearning.org/personalized-learning>
- Niles-Hofmann, L. (2016). Data-driven Learning design. How to decode learner digital body language. Retrieved from <http://www.loriniles.com/ebook/>
- Pane, J. F., Steiner, E. D., Baird, M. D., & Hamilton, L. S. (2015). *Continued progress: Promising evidence on personalized Learning*. RAND Corporation.
- Pane, J. F., Steiner, E. D., Baird, M. D., Hamilton, L. S., & Pane, J. D. (2017). *Informing Progress: Insights on personalized Learning implementation and effects*. RAND Corporation.
- Peersman, G. (2014). Overview: Data collection and analysis methods in impact evaluation: Methodological briefs – impact evaluation no. 10, methodological Briefs no. 10.
- Qi, J., & Coogan, P. (2015). Leveraging canvas participation data to inform course design. *Dartmouth College*. Retrieved from <https://cpb-us-e1.wpmucdn.com/sites.dartmouth.edu/dist/c/571/files/2015/04/LeveragingCanvasData1.pdf>

- Qi, J., & Esco, J. (2017). Leveraging Canvas learning data to inform MOOC design. Retrieved from <https://cpb-us-e1.wpmucdn.com/sites.dartmouth.edu/dist/c/571/files/2015/04/IC17Presentation.pdf>
- Rajbhandari, L., & Schober, J. (2013). *Data warehouse and analytics discussion paper*. Northwestern University.
- Renshaw, P., Baroutsis, A., van Kraayenoord, C., Goos, M., & Dole, S. (2013). *Teachers using classroom data well: Identifying key features of effective practices. Final report*. The University of Queensland.
- Reyes, J. A. (2015). The skinny on big data in education: Learning analytics simplified. *TechTrends*, 59(2), 75–80. <https://doi.org/10.1007/s11528-015-0842-1>
- Schildkamp, K., Poortman, C., Luyten, H., & Ebbeler, J. (2017). Factors promoting and hindering data-based decision making in schools. *School Effectiveness and School Improvement*, 28(2), 242–258. <https://doi.org/10.1080/09243453.2016.1256901>
- Spearman, M. M. (2019). *Using evidence based interventions & practices: A process guide for improvement*. State of South Carolina Department of Education.
- Treser, M. (2015). *Elearning industry: Getting to know ADDIE*. Retrieved from <https://elearningindustry.com/getting-know-addie-analysis>
- Vai, M., & Sosulski, K. (2011). *Essentials of online course design: A standards-based guide* (2nd ed.). Routledge.
- Verbert, K., Manouselis, N., Drachsler, H., & Duval, E. (2012). Dataset-driven research to support learning and knowledge analytics [electronic version]. *Educational Technology & Society*, 15(3), 133–148.
- Wagner, E., & Ice, P. (2012). Data changes everything: Delivering on the promise of learning analytics in higher education. *Educause Review*, 47(4), 32–36.
- Williams, K. C. (2018). *Top five business trends impacting K-12 education in 2018*. Gartner. Retrieved from <https://emtemp.gcom.cloud/ngw/globalassets/en/research/documents/top-five-business-trends-impacting-k-12-education-in-2018-3847265-344453.pdf>
- Yupangco, J. (2017, December 21). What sources of learning analytics should you be collecting? Retrieved from <https://elearningindustry.com/sources-of-learning-analytics-should-collecting>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 2

Adding Value and Ethical Principles to Educational Data



2.1 Introduction and Scope

2.1.1 Scope

The goals on this chapter are to:

- discuss the fundamentals of **educational data management**, including issues related with data cleaning methods, metadata, data curation and storage for preserving educational data, and
- introduce the **key Ethical Principles** that govern the use of educational data, especially in terms of **privacy**, **security of data** and **informed consent** that should be addressed via transparent and well-defined ethical policies and codes of practices.

2.1.2 Chapter Learning Objectives

Learning Objectives	Learn2Analyse Educational data literacy Competence profile
Know and Understand the most common quality issues of raw educational data	1.2
Understand data cleaning methods for educational datasets	2.1
Understand the advantages of enhancing educational data through data description	2.2
Understand the need for data curation in educational data management	2.3
Be able to identify storage issues for preserving educational data	2.4

(continued)

Learning Objectives	Learn2Analyse Educational data literacy Competence profile
Understand the importance of informed consent as a key Ethical Principle of Educational Data	6.1
Understand the significance of educational data protection policies	6.2

2.1.3 Introduction

This chapter will introduce the second key competence of educational data literacy, namely, **Educational Data Management**.

The first step in this imperative process is **Data Cleaning**. Since educational data comes from various sources, it could be really messy. It may come in diverse formats and it may contain various types of inaccuracies. Thus, it is essential to know the most common quality issues of raw educational data and understand the data cleaning methods for educational datasets.

In order to add value to the datasets, educators need to understand the advantages of **enhancing educational data through data description by using Metadata**, usually defined as “data about data”.

Data Curation is attributed with great importance in educational data management, in order to transform raw data into consistent data that can then be analysed.

Moreover, to ensure continued and reliable long-term access there are many important aspects we need to consider and manage, when it comes to an **effective digital preservation process for the educational data**.

Special focus should be given on key technical elements of digital preservation. The selected **storage solution is of prime importance for digital preservation**, since security and privacy issues are significant concerns.

Along with the emerging opportunities offered, education data-driven practice and assessment **raise challenges such as ethical issues and implications** especially in terms of **privacy, security of data and informed consent that should be addressed via transparent and well-defined ethical policies and codes of practices**.

Several frameworks, policies and guidelines have been developed to help institutions and educators to identify potential ethical issues and to apply clear ethical policies that govern the use of educational data.

New regulations, like the GDPR (General Data Protection Regulation) have raised awareness of data ethics issues that can arise from data misuse.

Informed consent is declared by most international guidelines as one of the pivotal principles in Data Ethics. The way individuals are informed is crucial for the informed consent process. Educators should ensure that individuals fully realize the expected consequences of granting or withholding consent.

With regards to the collection of personal data about children, additional protection should be granted since children are less aware of the risks and consequences of sharing data and of their rights.

As mentioned, in the light of rapid development of Educational Data Analytics on a global basis, **new challenges to privacy and data protection have also emerged.**

Do educational data analytics challenge the principles of data protection? Is privacy a show-stopper? How privacy is guaranteed/secured, especially if minors and/or sensitive data is involved?

Education professionals need to pay extra attention to **sensitive data** (special category of personal data) since an organisation can only process this data under specific conditions (explicit consent may be needed).

Moreover, the protection of the rights and freedoms of natural persons with regard to the processing of personal data require that appropriate technical and organisational measures are taken. In order to identify sensitive data, assess and respond to data risks and monitor implemented security processes, a Data Protection Impact Assessment (DPIA) may be required whenever processing is likely to result in a high risk to the rights and freedoms of individuals (IT Governance UK, 2016).

2.2 Adding Value to Educational Datasets (Educational Data Management)

2.2.1 Making Data Tidy (Data Cleaning)

We are surrounded by a sea of data. As per BrightBytes (2017) “*The widespread availability of accurate and usable data has the potential to unlock a universe of information for educators.*” We could add, that without the appropriate process of getting data ready to use (whether you call it wrangling, cleansing or simply cleaning), “*data is simply a scatter of numbers*”. You may also review the video “[Data Wrangling for Faster, More Accurate Analysis](#)” (in the useful video resources) showing that “Data discovery is a critical step when working with complicated data”.

In this topic, we will continue studying the language of data. It is time for the second key area of data literacy vocabulary, **Educational Data Management**. The first step in this imperative process is **Data Cleaning**. Figure 2.1 depicts the framework of data cleaning as defined by Maletic and Marcus (2000) in [Data Cleansing: Beyond Integrity Analysis](#).

As mentioned, educational data comes from various sources. There is data from online learning environments, data from state tests, demographic data, data from management information systems, from open educational resources and much more. It would be really useful if we could unify all these little pieces to reveal the big picture and realize the untapped potential.

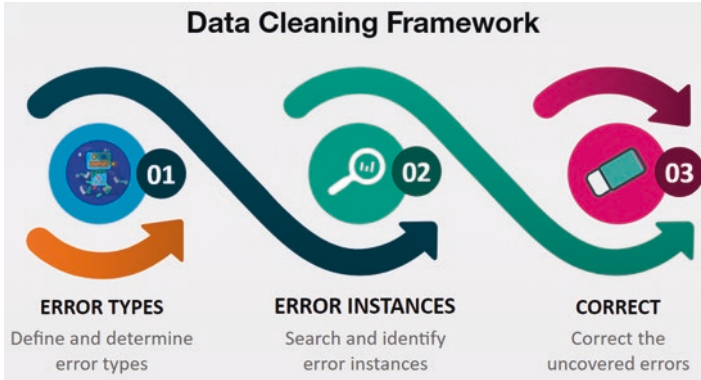


Fig. 2.1 Data cleaning framework. (based on Maletic & Marcus, 2000)

All this data could be really messy. It may come in diverse formats and it may contain various types of inaccuracies like missing values, outliers, duplicate instances. To obtain an integrated and consistent database that is free from any sort of discrepancies, data clean-up is required.

As Romero et al. (2014) describe in [A Survey on Pre-Processing Educational Data](#), the **data cleaning task** concerns the detection of erroneous or irrelevant data and how to discard it.

Let's move on and find out the most common discrepancies in data, like:

- **missing data,**
- **outliers,**
- **inconsistent data,**
- **double instances,**

and how to handle them (Fig. 2.2).

Missing values occur when no value is stored for the variable in the current observation (Little & Rubin, 2002).

When using an e-learning environment, it is very common for learners to study at their own pace, to follow their own learning path. They usually skip some activities and complete only a part of the tasks in the course. Sometimes they even drop out and never come back. Thus, missing data is very common when collecting educational data.

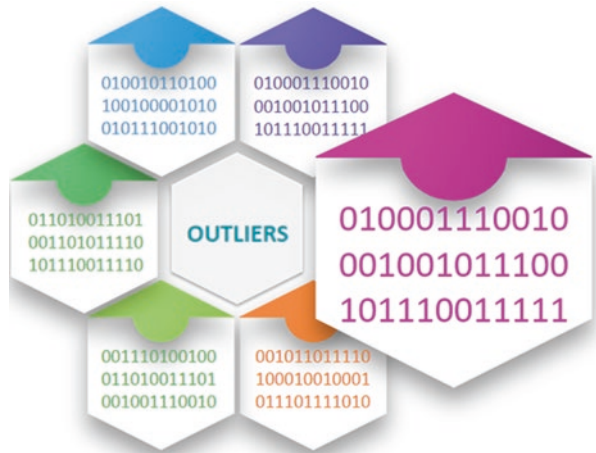
Romero et al. (2014) suggest several ways to handle missing data:

- Use a **label**, like "null" (unspecified), or "?" (missing)
- Use a substitute value like the attribute **mean** or the mode
- By determining what is the **most probable value** to fill the missing value, using **regression**.
- In some extreme cases, in order to clean data and ensure their completeness, learners who have all or almost all their values missed can be **removed** from data.

Fig. 2.2 Missing data



Fig. 2.3 Outliers



An **outlier** is an observation that has values which deviate from the expected, either too large or too small from most other observations (Fig. 2.3). They may be caused by typographical errors or errors in measurement. Remember when NASA lost a Spacecraft due to a Metric Math mistake (Harish, 2019)?

In datasets, different scales of numerical values are often used to make it easier for humans to read. For example, in budget datasets, the units are often in the millions. 1,500,000 often becomes 1.5 m. However, smaller amounts like 400,000 are still written in full. As a result, 1.5 m looks like it is an outlier, while it is an inconsistency in data types and formats.

However, Romero et al. (2010) indicate that “outliers may be phenomena of interest in a dataset, it could be correct and represent real variability for the given attribute.”

In the context of educational data, outliers can be often true observations (Romero et al., 2014). For example, there are always exceptions among learners, who succeed with little effort or fail against all expectations. In another example, very high values are often recorded for *time-spent* because the learner had not signed-out before leaving the digital learning environment.

It is clear that not all outliers are errors. It depends on the aims of the analysis, whether these outliers should be eliminated or not, and requires knowledge of the context in which the data was produced and collected.

Inconsistent data (Fig. 2.4) appears when a data set or group of data is dramatically different from a similar data set (conflicting data set) for no apparent reason (Romero et al., 2014).

For example, imagine negative values for the age of a person or height data measured either in meters or in centimetres. In fact, some incorrect data may also result from inconsistencies in naming conventions or data codes in use, or inconsistent formats for input fields, such as a date (Chakrabarti et al., 2009). The most common error is the mixed use of American (MM/DD/YYYY) and European (DD/MM/YYYY) formats (see [Date formats around the world](#)).

People often try to save time when entering data by abbreviating terms. If these abbreviations are not consistent, it can cause errors in the dataset. Differences in capitalisation, spacing, and genders of adjectives can all cause errors. There can be numerous inconsistencies. We have to deliberately deal with them. At the same time, it is in every case better to log the details of our procedure cautiously for future reference.

Fig. 2.4 Inconsistent data



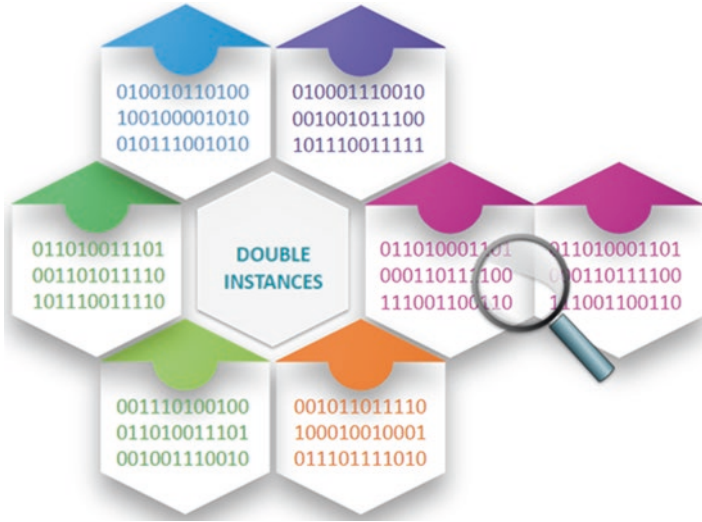


Fig. 2.5 Double instances

Data deduplication is a process that reduces storage overhead by eliminating redundant copies of data and, ensuring that storage media retain only unique instances of data. A duplicate record is where the same piece of data has been entered more than once (Fig. 2.5). Duplicate records often occur when datasets have been combined or because it was not known there was already an entry.

In educational organisations, **data integration and correlation** are essential activities related to data collection. Information obtained from multiple sources usually leads to duplicated data observations and inaccurate data. This duplicate elimination is one of the most important steps in the data cleaning process. The procedure of detecting and eliminating duplicates from a particular data set is called Deduplication.

According to [Crowdfunder Data Science Report 2016](#), scientists spend the most time collecting and cleaning data (Fig. 2.6). Messy data is by far the most time-consuming aspect of the typical data scientist’s workflow.

The point with data is that it needs to be regularly maintained to ensure that data remains clean and crystal clear Ronald van Loon (2018).

Much of the data may be unstructured, noisy and in need of thorough cleansing and preparation before it is ready to yield working insights Big Data expert, Bernard Marr (2017).

Questions and Teaching Materials

1. **Finally, after Alice collected the necessary parental consent for her intervention, the flipped classroom course is up and running.**

After running the online course for three weeks, Alice tracks her students’ activity in the online learning environment. Thus, she also collects data related to

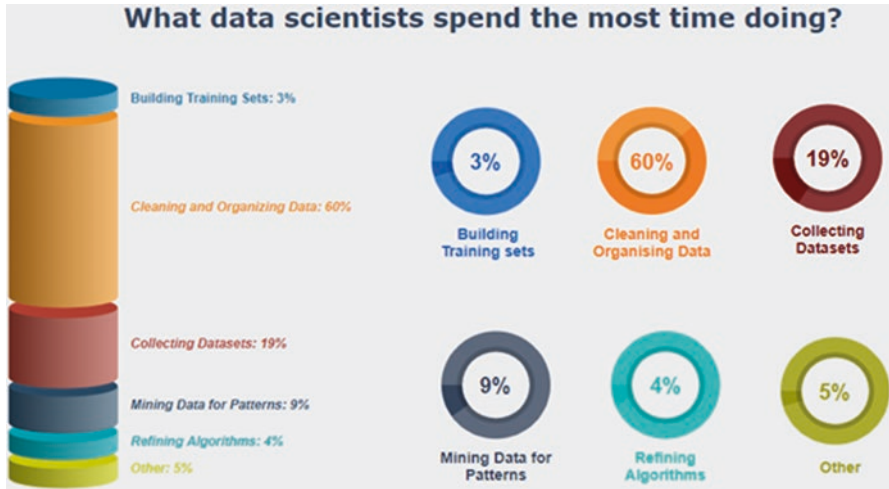


Fig. 2.6 What data scientists spend the most time doing

students’ engagement, behaviour and performance in the LMS e.g. time spent in the platform, the videos her students watched, their progress in the online course, downloaded files, their online quiz scores, their participation in the forum as well as interaction among them.

Before proceeding further, Alice confirms that the collected data meets basic quality characteristics. She watches the video “[Data Wrangling for Faster, More Accurate Analysis](#)”. Thus, she examines and verifies the educational data against different quality measures. Inconsistencies in data, like missing pieces, errors, even differences in how the same value is expressed, produce inaccurate results.

- True
- False

Correct answer: True

2. Alice has collected educational data from various sources (data from online learning environments, data from state tests, demographic data, data from management information systems, from open educational resources and much more) and she wants to unify the datasets in order to reveal the big picture.

Alice soon realizes that the data coming from various sources in diverse formats, is quite messy, containing missing values, outliers, and duplicate instances.

To obtain a consistent database, free from any sort of discrepancies, data cleaning is required so as to detect erroneous or irrelevant data and discard it.

In the framework of data cleaning, as defined by Maletic and Marcus (2000) and presented in fig. 2.1, the following three phases define a data cleansing process.

Help Alice to arrange the phases in the right order:

- A. **Correct the uncovered errors**
- B. **Define and determine error types**
- C. **Search and identify error instances**

Correct answer: B – C – A

3. **Alice has collected data from the Learning Management System and she realizes that some users accessed her course just once (in error or in order to see one specific resource or to do an activity) but never returned to the course later.**

What would you suggest Alice to do in order to handle the missing values?

- A. **to use a label, like “null” (unspecified), or “?” (missing)**
- B. **to use a substitute value like the attribute mean or the mode**
- C. **by determining what is the most probable value to fill the missing value, using regression.**
- D. **by removing these learners from the dataset.**

Correct answer: D

4. **Alice has extracted the following dataset containing file downloads data from the school’s Learning Management System.**

She can easily identify two outliers (Student4 and Student11). Help Alice to decide what to do with these outliers, in order to proceed with the data analysis. These outliers:

- A. **are errors and should be eliminated in order to proceed.**
- B. **are true observations and should not be eliminated.**

Correct answer: B

	File1.pdf	File2.pdf	File3.pdf	File4.pdf	File5.pdf	File6.pdf	File7.pdf	File8.pdf	File9.pdf
Student1	2	1	0	2	1	1	0	1	2
Student2	1	3	2	1	1	1	2	1	1
Student3	1	1	2	1	1	0	1	2	3
Student4	12	14	18	20	16	15	14	12	9
Student5	1	0	1	2	1	2	1	0	2
Student6	1	2	1	1	1	1	3	2	1
Student7	0	1	2	3	1	1	1	2	1
Student8	1	1	0	1	2	2	1	0	2
Student9	1	1	2	1	1	1	3	2	1
Student10	1	0	1	2	3	1	1	2	1
Student11	16	15	14	12	9	12	11	10	8
Student12	1	2	1	0	2	1	0	1	2
Student13	1	1	3	2	1	1	2	1	1
Student14	1	1	1	2	1	0	1	2	3
Student15	1	0	1	2	1	2	1	0	2
Student16	1	2	1	1	1	1	3	2	1
Student17	0	1	2	3	1	1	1	2	1
Student18	1	0	1	2	1	2	1	0	2
Student19	1	2	1	1	1	1	3	2	1
Student20	0	1	2	3	1	1	1	2	1
Student21	2	1	0	2	1	0	1	2	1
Student22	1	3	2	1	1	2	1	1	1
Student23	1	1	2	1	0	1	2	3	1

5. Alice participates in an International Conference on Teaching and Learning. Therefore, she must prepare a review of students' performance from 6 different countries in three main subjects, namely Maths, English, and Science.

Students' performance data from 6 different countries are collected in the following table.

	Date of Birth	Student	Maths	English	Science	Country
1	4/9/2008	Richard	95	68	96	USA
2	9/10/2007	David	65	78	70	UK
3	12/12/2009	Mary	59	55	53	USA
4	6/12/2010	Ann	97	99	98	France
5	8/13/2011	Elen	100	97	98	Greece
6	11/14/2010	Catherine	67	59	70	UK
7	9/14/2005	James	54	67	63	USA
8	5/17/2006	Martha	79	83	88	Italy
9	4/17/2007	Bill	84	78	90	UK
10	8/18/2007	Phil	45	78	55	USA
11	9/18/2008	James	75	83	88	Italy
12	10/19/2009	Tom	85	89	92	Greece
13	6/19/2010	Joe	9,4	9,7	9,1	UK
14	9/20/2029	Jill	49	60	53	Canada
15	5/17/2006	Martha	79	83	88	Italy
16	12/12/2009	Mary	59	55	53	USA
17	24/10/2010	Tony	96	79	100	Italy
18	8/24/2006	Lisa	79	-75	69	UK
19	5/25/2004	Robert	97	83	90	USA
20	4/25/2029	Michael	100	89	55	Italy
21	25/6/2007	Rose	67	97	88	Greece
22	8/26/2008	Sofia	54	60	92	UK
23	9/26/2009	Jim	97	88	67	Greece
24	4/26/2006	Betty	60	92	54	France

Alice soon realises that the key to finding the inconsistencies is to **create a filter**. The filter will allow her to see all of the unique values in the column, making it easier to isolate the incorrect values. (Source: <https://edu.gcfglobal.org/en/excel-tips/a-trick-for-finding-inconsistent-data/1/>).

After examining carefully this table, please help Alice to select the inconsistencies you have identified

- A. negative values for students' grades
- B. different data formats
- C. typos in dates

- D. **differences in spaces**
- E. **different grades' scale**
- F. **typos in country data**
- G. **differences in capitalisation**

Correct answers: A, B, C, E, F. In our example, we can identify the following inconsistencies: In row 21 Greece is misspelled and in row 11 Italy has double l; In row 18 there is a negative value for the grade in English; In row 13 grades are in different scale; In rows 14 and 20 dates are out of range; and In rows 17 and 21 dates are in different format (DD/MM instead of MM/DD).

6. **Alice participates in an International Conference on Teaching and Learning. Therefore, she must prepare a review of students' performance from 6 different countries in three main subjects, namely Maths, English, and Science.**

Students' performance data from 6 different countries are collected in the following table.

	Date of Birth	Student	Maths	English	Science	Country
1	4/9/2008	Richard	95	68	96	USA
2	9/10/2007	David	65	78	70	UK
3	12/12/2009	Mary	59	55	53	USA
4	6/12/2010	Ann	97	99	98	France
5	8/13/2011	Elen	100	97	98	Greece
6	11/14/2010	Catherine	67	59	70	UK
7	9/14/2005	James	54	67	63	USA
8	5/17/2006	Martha	79	83	88	Italy
9	4/17/2007	Bill	84	78	90	UK
10	8/18/2007	Phil	45	78	55	USA
11	9/18/2008	James	75	83	88	Italy
12	10/19/2009	Tom	85	89	92	Greece
13	6/19/2010	Joe	94	97	91	UK
14	9/20/2009	Jill	49	60	53	Canada
15	5/17/2006	Martha	79	83	88	Italy
16	12/12/2009	Mary	59	55	53	USA
17	10/24/2010	Tony	96	79	100	Italy
18	8/24/2006	Lisa	79	75	69	UK
19	5/25/2004	Robert	97	83	90	USA

(continued)

	Date of Birth	Student	Maths	English	Science	Country
20	4/25/2009	Michael	100	89	55	Italy
21	6/25/2007	Rose	67	97	88	Greece
22	8/26/2008	Sofia	54	60	92	UK
23	9/26/2009	Jim	97	88	67	Greece
24	4/26/2006	Betty	60	92	54	France

After searching the web for answers, Alice finds out that she can identify duplicate rows by selecting Home-Conditional Formatting-Highlight Cell Rules-Duplicate Values in MS Excel.

Help Alice identify the duplicates. How many duplicates can you identify?

- A. None
- B. One pair of rows
- C. One triplet of rows
- D. Two pairs of rows

Correct answer: D

7. After reading the [Crowdfunder Data Science Report](#), Alice realises that mining data for patterns and refining algorithms are the two most time-consuming tasks of a data-scientist's workflow.

- True
- False

Correct answer: False.

8. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to elaborate on your response about data cleaning in the following reflective task. You may reflect on:

1. Identify factors that contribute to inconsistencies to educational datasets generated from online courses
2. How can we explain the existence of outliers in educational data?

2.2.2 Data to Describe Data (Metadata)

Metadata is usually defined as “data about data”. Johnson et al. (2018) provide the following definition about metadata “It is information about a data set that is structured (often in machine-readable format) for purposes of search and retrieval. Metadata elements may include basic information (e.g., title, author, date created) and/or specific elements inherent to data sets (e.g., spatial coverage, time periods).”

However, in the context of education, metadata can more aptly be defined as tags used to describe educational assets.

Metadata helps:

- to organize,
- find and
- understand data

Metadata answers the following questions about data:

- Who created it?
- What is it?
- When was it created?
- How was it generated?
- Where was it created?
- How may it be used?
- Are there restrictions on it?

Practical examples of metadata: <https://dataedo.com/kb/data-glossary/what-is-metadata> Kononow (2018), Fig. 2.7)

In *Understanding Metadata* 2017, from the National Information Standards Organization, Riley (2017) distinguishes the three types of metadata (see Fig. 2.8):

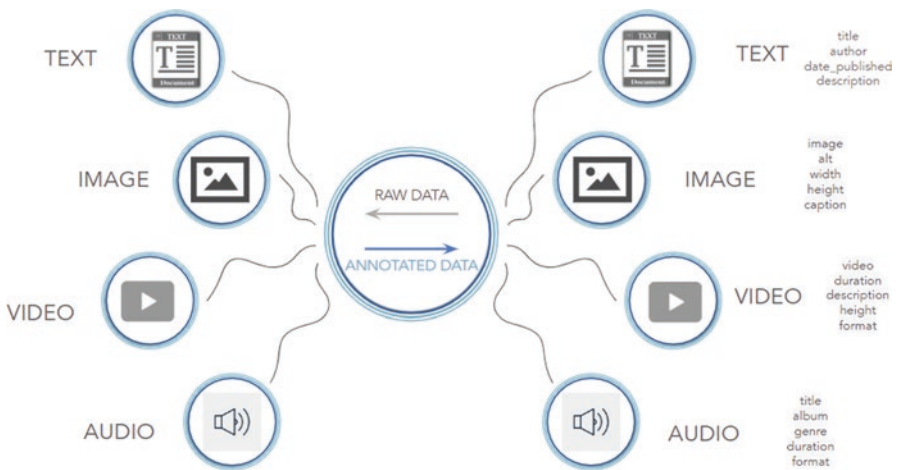


Fig. 2.7 Examples of metadata

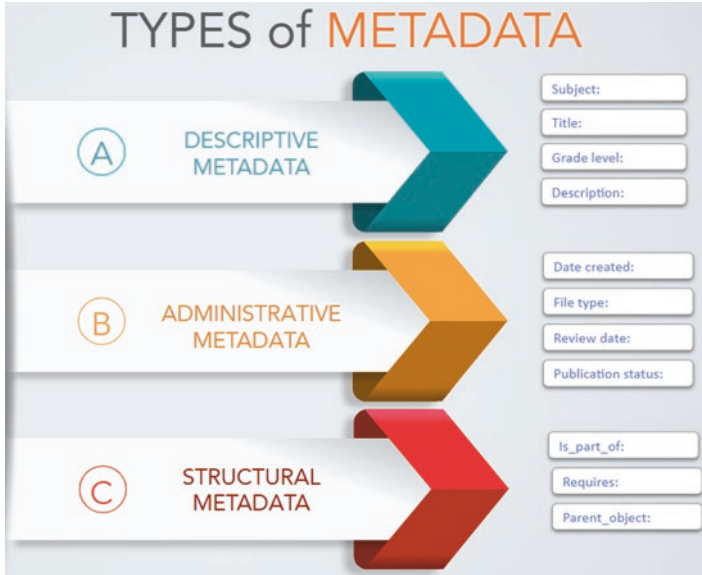


Fig. 2.8 Types of metadata

- **Descriptive metadata**
- **Administrative metadata**
- **Structural metadata**

Descriptive metadata can describe a learning asset or resource related to education — including learning standards, lessons, assessment items, books, etc. — for purposes such as identification, search and discovery. Descriptive metadata can be thought of as a keyword or tag on an asset that makes it easier to find. Examples include subject, grade level, and related skills and concepts.

Administrative metadata is used to manage a learning asset. Examples of this type of metadata include status, disposition, rights and licensing.

Structural metadata describes how data is organized or formatted and is often governed by a [widely-adopted standard](#) that ensures the data is accurately represented when exchanged and presented. Structural metadata enables content to be machine readable.

Metadata are used for the **purposes** of:

- **Discovery of information**
- **Identification of a resource**
- **Interoperability, exchange of content between systems**
- **Digital-object management i.e., deliver the appropriate version.**
- **Preservation helps signalling when preservation actions should be undertaken**
- **Navigation within parts of items**

Primary uses of various metadata types are presented in the Table 2.1 below (adapted from [Understanding Metadata](#), 2017).

The video from the National Archives of Australia “[Meta... What? Metadata](#)” (in the useful video resources) helps us understand the importance of metadata in order to describe, use, find and manage content and data.

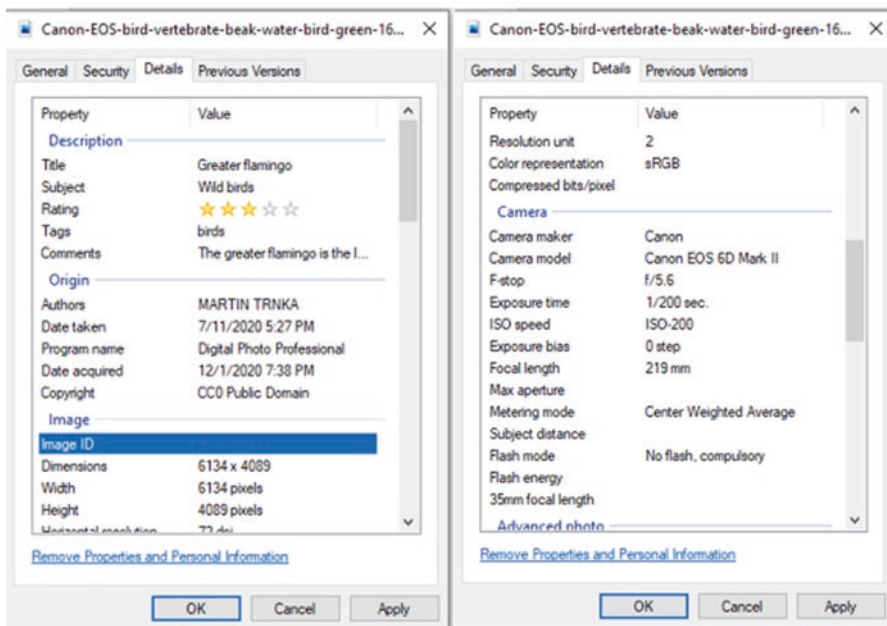
The [National Information Standards Organization](#) describes “*data interoperability, as the effective exchange of content between systems. Interoperability relies on metadata describing that content so that the systems involved can effectively profile incoming material and match it to their internal structures.*” You may also review this video “[Learn More About Data Interoperability](#)” (in the useful video resources).

Questions and Teaching Materials

1. Alice has heard of “metadata”, but she is not quite sure what it means or why she might need it. She downloaded this photo from [pxhere.com](#) an online community sharing copyright-free images.

Table 2.1 Primary uses of various metadata types

Metadata type	Example properties	Primary uses
Descriptive metadata	Title Author Subject Genre Publication date	Discovery Display Interoperability
Technical metadata	File type File size Creation date/time Compression scheme	Interoperability Digital object management Preservation
Preservation metadata	Checksum Preservation event	Interoperability Digital object management Preservation
Rights metadata	Copyright status License terms Rights holder	Interoperability Digital object management
Structural metadata	Sequence Place in hierarchy	Navigation
Markup languages	Paragraph Heading List Name Date	Navigation Interoperability



Photo's properties

What information can Alice gather from photo's metadata? Match the questions from the first column with the values in the second column.

Question	Value
A. Who created the photo?	1. Greater Flamingo 2. CC0 Public Domain 3. 12/1/2020 7:38 PM
B. What is it?	4. 7/11/2020 5:27 PM 5. Alice
C. When was it created?	6. Canon EOS 6D Mark II 7. 219 mm
D. How was it generated?	8. MARTIN TRNKA 9. sRGB
E. What are the photo's copyrights	10. ISO-200 11. Digital Photo Professional

Correct answer: A8 – B1 – C4 – D6 – E2

2. **Open educational resources (OER) are freely accessible, openly licensed text, media, and other digital assets that are useful for teaching, learning, and assessing as well as for research purposes. The term OER describes publicly accessible materials and resources for any user to use, re-mix, improve and redistribute under some licenses.**

OER Repositories are repositories of open educational resources covering most of educational disciplines. Open Repositories are websites which house open books, textbooks, lectures, tutorials, quiz/test, case studies, assessment tools, images, syllabi, simulations, online courses and other resources of educational value.

Photodentro OER repositories is the Greek National Learning Object Repository (LOR) for primary and secondary education. It hosts reusable learning objects (small, self-contained reusable units of learning). It is open to everyone, pupils, teachers, parents, as well as anybody else interested. The URL for accessing Photodentro LOR is <http://photodentro.edu.gr/lor>.

For the purpose of collecting learning material for the flipped classroom initiative, Alice has found the following Learning Object (LO) in Photodentro OER repositories:



Alice is studying the Learning Object's metadata page (<http://photodentro.edu.gr/lor/r/8521/2705?locale=en>) to find answers to the following questions:

1. What is the Subject Area of the LO?

- A. English Language > Literature – Art – Culture > Reading
- B. FOREIGN LANGUAGE
- C. B1-medium knowledge
- D. Lost in the Museum (mystery game)

Correct answer: A.

2. What are the Licence Terms of the LO?

- A. Creative Commons Attribution-NoDerivatives Greece 3.0
- B. Creative Commons Attribution-ShareAlike 3.0 International License.
- C. Creative Commons Attribution-NonCommercial-ShareAlike Greece 3.0
- D. Creative Commons Attribution-NonCommercial-NoDerivatives Greece 3.0

Correct answer: C.

3. What is the Date of Publication?

- A. 02/09/2019
- B. 03/09/2019
- C. 7/12/2020
- D. 19/05/2013

Correct answer: D.

4. What is the File Size?

- A. 4.91 MB
- B. 12–15 MB
- C. 25 MB
- D. 8125 MB

Correct answer: A.

5. After watching the video “[Meta... What? Metadata!](#)” Alice realises one of the most common uses of metadata, which is to group content, making it more efficient to retrieve it during a search.

- **True**
- **False**

Correct answer: True.

6. Alice watches the video from the League of Innovative Schools “[Learn More About Data Interoperability](#)” promoting the movement to advance data interoperability in public education.

In this video, data interoperability is defined as the seamless, safe and controlled exchange between applications, with clear standards for how to send and receive student information, privately and securely.

- **True**
- **False**

Correct answer: True

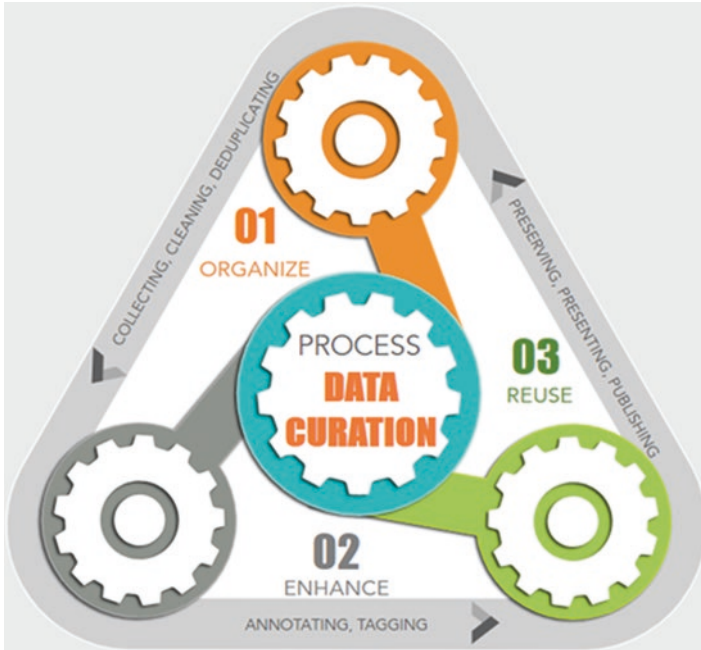


Fig. 2.9 Data curation

7. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to elaborate on your response about metadata, in the following reflective task. You may reflect on:

The advantages of enhancing educational data through data description.

2.2.3 The Significance of Data Curation

According to ICPSR (2018), “Through the curation process, data are organized, described, cleaned, enhanced, and preserved for public use, much like the work done on paintings or rare books to make the works accessible to the public now and in the future. Without curation, however, data can be difficult to find, use, and interpret” (Fig. 2.9).

Michael Stonebraker (2014), defines **data curation** as the process of turning independently created data sources (structured and semi-structured data) into unified data sets ready for analytics, using domain experts to guide the process. It involves:

- **Identifying** data sources of interest (whether from inside or outside the enterprise)
- **Verifying** the data (to ascertain its composition)
- **Cleaning** the incoming data (for example, 99,999 is not a legal zip code)

- **Transforming** the data (for example, from European date format to US date format)
- **Integrating** it with other data sources of interest (into a composite whole)
- **Deduplicating** the resulting composite data set.

Castanedo (2015), on the other hand, describes **data curation** as the process that involves data cleaning, schema definition/mapping, and entity matching to transform raw data into consistent data that can then be analysed. Schema definition/mapping is making associations among data attributes and features. Entity matching is finding data in different data sources that refer to the same entity. Entity matching is essential to remove duplicate records.

In this video, “[ICPSR 101: What is Data Curation?](#)” (in the useful video resources), ICPSR explains the intricacies of the work data processors do every day to find and fix issues in the data, ensuring their long-term availability and value to the research community.

According to The Digital Curation Centre (DCC) Fig. 2.10 provides a graphical, high-level overview of the stages required for successful curation and preservation of data from initial conceptualisation or receipt through the iterative curation cycle.

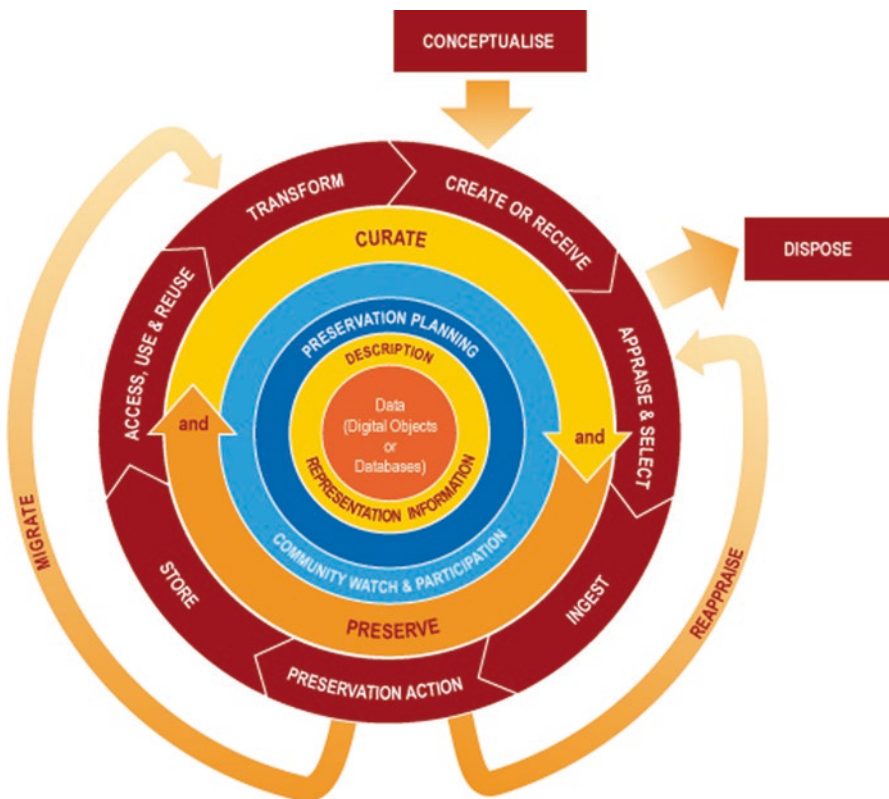


Fig. 2.10 The DCC curation lifecycle model. (Source: diagram from Higgins, 2008)

We can identify **four full life cycle actions**:

- **Description and Representation**
- **Preservation Planning**
- **Community Watch and Participation**
- **Curate and Preserve**

The outer cycle represents the **sequential actions** of the data curation process:

- **Conceptualise**
- **Create or Receive**
- **Appraise and Select**
- **Ingest**
- **Preservation Action**
- **Store**
- **Access, Use and Reuse**
- **Transform**

Digital curation is all about maintaining and adding value to a trusted body of digital information for future and current use; specifically, the active management and appraisal of data over the entire life cycle (Jisc, 2006).

You may also review the video “[Data Curation @UCSB](#)”, (in the useful video resources) to watch how UCSB Library eyes digital curation service to help preserve research data created across campus.

Now that we have completed the hard work to make our data tidy and meaningful, we will put in a little extra effort to preserve our valuable results.

Thus, we will discuss **Digital Educational Data Preservation** which is considered a key task in the data curation process, to safeguard our unique educational data from getting stolen, destroyed or simply lost.

Questions and Teaching Materials

1. **Alice is studying the Data Curation Process to ensure that data is reliably retrievable for future reuse, and to determine what data is worth saving and for how long.**

Help Alice match the following Data Curation processes to the appropriate Data Curation Phase.

Data curation process	Data curation phase
A. Cleaning	Phase 1: Organize
B. Presenting	
C. Annotating	
D. Preserving	Phase 2: Enhance
E. Collecting	
F. Tagging	Phase 3: Reuse
G. Deduplicating	
H. Publishing	

Correct answer: A1-B3-C2-D3-E1-F2-G1-H3.

2. **Data Curation is not quite clear to Alice, so she watches the [video from ICPSR \(“ICPSR 101: What is Data Curation?”\)](#) explaining what data curation is all about. According to this video, the purpose of data curation is to ensure that people can find data now and in the future. This can be achieved by following the 5 steps of data curation.**

Please help Alice to arrange the following steps in the right order:

- A. Find and fix issues with data
- B. Identify data in the scope of the archive
- C. Ensure that data will last forever (or at least for a very long time)
- D. Make data findable and usable
- E. Get data (convince the data owners to share it)

Correct answer: B-E-A-D-C

3. **Alice studies the [Digital Curation Centre’s \(DCC\) Curation Lifecycle Model](#). According to this complex diagram, there are four full lifecycle actions and eight sequential actions of the data curation process.**

Please help Alice to select only the full lifecycle data curation actions from the following list.

- A. Create or Receive
- B. Description and Representation
- C. Access, Use and Reuse
- D. Appraise and Select
- E. Preservation Planning
- F. Curate and Preserve
- G. Transform
- H. Community Watch and Participation

Correct answers: B, E, F, H

4. **The last step of Data Curation Cycle is to ensure that data will last forever (or at least for a very long time). Alice is anxious, how can digital records last “forever”? What if the technology becomes obsolete?**

Thankfully, in the “[Data Curation @UCSB](#)” video Alice just watched Greg Janeé, a Digital Library Research Specialist claims that digital information is far more robust than paper.

Is Alice’s understanding correct?

- Yes
- No

Correct answer: No.

5. ACTIVITY/PRACTICE QUESTION (Short answer)

Name some of the data curation actions described in this session.

6. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to elaborate on your response in the following reflective task. You may reflect on:

The significance of data curation in educational data management.

2.2.4 Storage Issues for Preserving Educational Data

As explained in the short Library of Congress video “[Why Digital Preservation is Important for Everyone](#)” (in the useful video resources), traditional information sources such as books, photos and sculptures can easily survive for years, decades or even centuries but digital items are fragile and require special care to keep them useable. Rapid technological changes also affect digital preservation. As new technologies appear, older ones become obsolete, making it difficult to access older content.

This video explores the complex nature of the problem, how digital content, unlike content on traditional media, depends on technology to make it available and requires active management to ensure its ongoing accessibility.

Preservation is no longer simply a concern for memory institutions in the long term but for everyone interested in using and accessing digital materials. The greater the importance of digital materials, the greater the need for their preservation: digital preservation protects investment, captures potential and transmits opportunities to future generations and our own. Digital materials – and the opportunities they create – are fragile ([Digital Preservation Handbook](#)), Digital Preservation Coalition (2015).

Jisc, 2006 defines **Digital Preservation** as “*the series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as*



Fig. 2.11 The most important aspects we need to consider and manage, so as to ensure an effective digital preservation process for our educational data

long as they are deemed to be of value. This encompasses not just technical activities, but also all of the strategic and organisational considerations that relate to the survival and management of digital material”.

According to [Principles and Good Practice for Preserving Data](#), “A sustainable preservation programme addresses organisational issues, technological concerns and funding questions” (Interuniversity Consortium for Political and Social Research (ICPSR), 2009). The simple questions to be answered:

- **Organisational Issues:** “What are the requirements and parameters for the organisation’s digital preservation programme?”
- **Technological Issues:** “How will the organisation meet defined digital preservation requirements?”
- **Resources Issues:** “What resources will be needed to develop and maintain the digital preservation programme?”

Figure 2.11 is based on Digital Preservation Handbook (Digital Preservation Coalition, 2015), and presents the most important aspects we need to consider and manage, so as to ensure an effective digital preservation process for our educational data.

Even though our main focus is not to drill down deep into technical details and aspects of digital preservation issues, which are not part of educators’ main role,



Fig. 2.12 Digital preservation activities

however it is essential to get an overview and understanding so as to be able to collaborate effectively with the responsible technical team, using a common language. Thus, next we will discuss briefly such issues for the effective educational data digital preservation.

The first steps that need to be undertaken in order to begin to build or enhance the needed digital preservation activities are summarized in Fig. 2.12. You may further review detailed information in Digital Preservation Handbook (Digital Preservation Coalition, 2015).

Special focus should be given on these key technical elements of digital preservation, as specified under USGS Guidelines, 2014:

- **Storage & Geographic Location** – Storage systems, locations, and multiple copies to prevent loss of data.
- **Data Integrity** – Procedures to prevent, detect, and recover from unexpected or deliberate changes to data.
- **Information Security** – Procedures to prevent human-caused corruption of data, deletion and unauthorized access.
- **Metadata** – Documentation of the data to enable contextual understanding and long-term usability.
- **File Formats** – File types, data structures, and naming conventions to aid long-term preservation and reuse.
- **Physical Media** – Reduce obsolescence risks that can threaten the readability of physical media.

To assess an organization's readiness, it is recommended that these components are checked against the National Digital Stewardship Alliance (NDSA) 'Levels of Digital Preservation' (Phillips et al., 2013):

- **Level 1 – protect your data**
- **Level 2 – know your data**

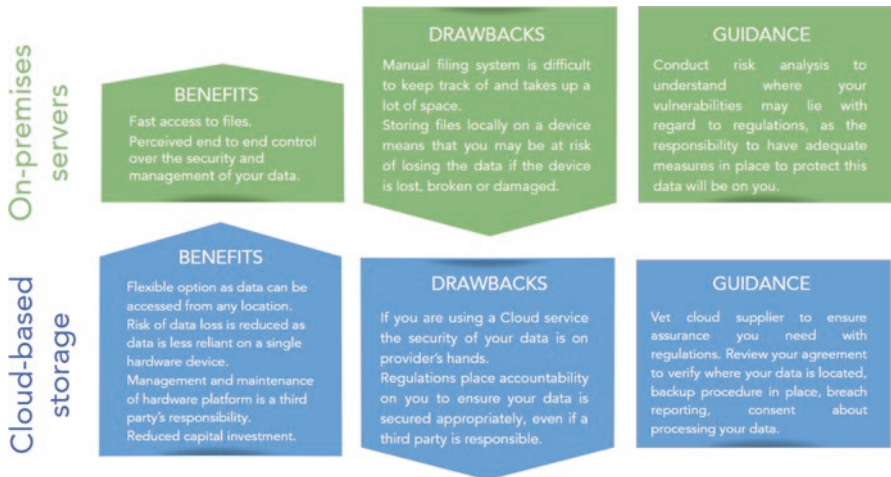


Fig. 2.13 Two storage methods

- **Level 3 – monitor your data**
- **Level 4 – repair your data**

With regards to the storage technology, it has changed dramatically over the last twenty years. Initially, the norm was storing data using discrete media items, such as CDs/DVDs and hard-disk drives. Today, it has become common practice to use IT storage systems for the increasingly large volumes of digital material that needs to be preserved and to be easily and quickly retrievable (Digital Preservation Coalition, 2015).

At this point it is important to clarify the difference between backup and digital preservation process. **Backup** refers to “short-term data recovery solutions following loss or corruption” (Jisc, 2006). **Preservation storage systems** “require a higher level of geographic redundancy, stronger disaster recovery, longer-term planning, and most importantly active monitoring of data integrity in order to detect unwanted changes such as file corruption or loss” (Digital Preservation Handbook).

The selected storage solution is of prime importance for digital preservation. When selecting the storage strategy there are several options we need to consider, such as Cost and Scalability, required Capacity, Security, Remote Access, Collaboration and Disaster Recovery. Legal provisions due to privacy or confidentiality may also influence our decision. Figure 2.13 summarizes the pros and cons of each of the two basic storage methods, on-premises servers (local infrastructure/ data centres) and Cloud-based storage, as well as recommended actions to comply with the latest regulations (COMPARE THE CLOUD, 2018). You may also review the video “Public Cloud vs Private Cloud vs Hybrid Cloud” (in the useful video resources), which compares and contrasts public, private and hybrid clouds: the basic elements of each, the features and benefits that each delivers, and how each type meets specific business needs.

In their 2018 report, [Data Management Life Cycle Final report](#), Miller and his colleagues recognise the demand for cost-effective storage technologies. *“More and more organizations are considering outsourcing storage services or cloud storage options because the availability of cloud computing resources opens up possibilities for users to purchasing access to computing power and storage space as a service instead of maintaining it themselves. This way, providers are responsible for the performance, reliability, and scalability of the computing environment, while users can concentrate on data analysis and production”.*

Nevertheless, security and privacy are significant concerns holding back use of the cloud, particularly for confidential, sensitive, or personally identifiable information. Let’s not forget what happened at [Code Space](#), which led to data deletion and the eventual shutdown of the company.

The most common risks we need to consider include: Downtime and service outages since cloud computing systems are internet based, vulnerability to external cyber-security attacks, compliance and legal issues depending on the applied regulation, lifetime costs that could end up being higher than you expected as well as limited control and flexibility since the cloud infrastructure is owned, managed and monitored by the service provider.

Despite these concerns, the potential of cloud storage seems to be more promising than the associated risks which are expected to diminish over time. As per Gartner *“Through 2025, 99% of cloud security failures will be the customer’s fault”* (Panetta, 2019). and *“Organizations that do not have a high-level cloud computing strategy driven by their business strategy will significantly increase their risk of failure and wasted investment”* (Cearley, 2017).

Whichever is our choice, even a hybrid storage solution, we need to realize that storage technologies present several risks to long-term preservation of data. Moreover, *“Many cases of content loss are not necessarily due to technical faults but can come from human error, lack of budget, or a failure to regularly monitor the integrity of the stored data”* (Digital Preservation Coalition, 2015) (Fig. 2.14).

Let’s now take a closer look at security issues and particularly cybersecurity.

According to [Digital Preservation Handbook](#), security issues relate to:

- **system security** (e.g., protecting digital preservation and networked systems / services from exposure to external / internal threats),

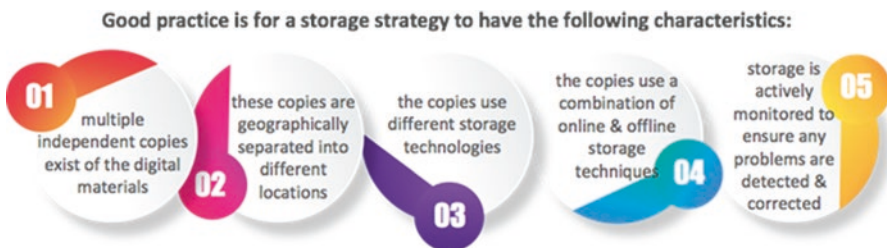


Fig. 2.14 Characteristics of good practice for storage strategy



Fig. 2.15 Countermeasures against cyber-attacks

- **collection security** (e.g., protecting content from loss or change, the authorisation and audit of repository processes), and
- **the legal and regulatory aspects** (e.g. personal or confidential information in the digital material, secure access, redaction).

When it comes to **cybersecurity**, protecting educational data requires both administrative and technological security measures, in order to prevent unauthorized parties from accessing it. In the below Fig. 2.15, you may review some of these countermeasures to create an effective defence against cyber-attacks.

In order to help school protect against cyberthreats and develop effective security programs, there is also a really useful [Report about K-12 Security Risk Methodology](#) (Woody, 2004), emphasizing that while technology “is broadly used in the K-12 environment by many participants including administrators, teachers, parents, students, school board members, etc.” “while this enables a wide range of useful activities, the risk for inappropriate and illegal behaviour that violates privacy, regulations, and common courtesy is increasing exponentially”.

The thing that kept me awake at night (as NATO military commander) was cybersecurity. Cybersecurity proceeds from the highest levels of our national interest ... through our medical, our educational, to our personal finance (systems). (Admiral James Stavridis, Ret. Former-NATO Commander in Cybersecurity and Digital Business Risk Management, 2020).

To this point we have provided an overview of the key issues of digital preservation and realized its importance to maintain usable our educational data over time. You may also review in this video “[How Toy Story 2 Almost Got Deleted: Stories From Pixar Animation: ENTV](#)” (in the useful video resources), the (mostly) true story of how ‘Toy Story 2’ was almost deleted from Pixar Animation’s computers during the making of the film. And how the film was saved by one mom’s home computer!

Let us move forwards to identify good practices and appropriate actions to collect the needed data, as well to protect this data and safeguard its privacy, especially when it comes to sensitive educational data.

After all, “Data protection is all about protecting people – not just files and computer systems” (Moore Barlow, 2018).

Questions and Teaching Materials

1. **Following the discussion with the DPO about the school’s preservation strategy and policies, Alice starts wondering. Is digital content so fragile, after all? Should I find more about preservation issues to protect my course’s digital content?**

Alice accesses the video “[Why Digital Preservation is Important for Everyone](#)”.

She now understands that though traditional information sources can easily survive for years, decades and even centuries, digital items require special care to preserve them. More specifically, the digital items are fragile as they require special care to keep them usable, they are dependent as they depend on technology to make them available and require active management to ensure their ongoing accessibility.

Is this assumption True or False? Please select the right answer.

- True
- False

Correct answer: True

2. **Alice soon realises that she needs to seek “guidance on key issues and actions to consider when creating digital materials to ensure their longevity of active use and potential for long-term preservation” ([Digital Preservation Handbook](#)).**

Please mark the correct key elements corresponding to each category of issues that Alice needs to address for digital preservation.

	Organisational issues	Technological issues	Resources issues
Integrity of Data over time		X	
Legal Compliance	X		
Budgets and Costs			X
Balancing Security and Access	X		
Staffing and needed Skills			X
Information Security		X	
Collaboration	X		
Facilities Required			X
Metadata Standards		X	
Selection of Data to be Preserved	X		
Sustainable File Formats		X	

Correct answers: as marked with X above

3. Alice is presently at the point of investigating on the key technical elements of digital preservation.

It’s a bit hard for her to deal with such technical issues. Are you ready to help her?

You may review the definitions of the key technical elements of digital preservation, presented in page 2 of the USGS Guidelines, [2014](#).

Please match the appropriate definition (from the right column), to the respective technical element (in the left column).

1. Metadata	A. Basic recommendations to reduce obsolescence risks that can threaten the readability of physical media
2. Physical Media	B. Storage systems, locations, and multiple copies to prevent loss of data
3. Information Security	C. File types, data structures, and naming conventions to aid long-term preservation and reuse
4. File Formats	D. Procedures to prevent human-caused corruption of data, deletion, and unauthorized access
5. Storage & Geographic Location	E. Documentation of the data to enable contextual understanding and long-term usability

Correct answers: 1-E, 2-A, 3-D, 4-C, 5-B

4. Let’s go back to Alice. She gets informed by the responsible colleague about the hybrid storage solution used by the school. It’s a combination of local infrastructure/data centre and cloud-based storage. Moreover, as per her school guidelines for data storage good practice strategy, she needs to create multiple independent copies to stabilize her files. The copies are geographically separated in different locations, using different storage technologies and are actively monitored to ensure any problems are detected and corrected.

She wonders about the criteria that influenced the school’s decision making for the selected storage solution for digital preservation. Can you help her specify these selection criteria?

Please select the right answers.

- A. Collision
- B. Security

- C. **Disaster Recovery**
- D. **Redundancy**
- E. **Cost**

Correct answers: B, C, and E.

5. Alice is now interested in learning more about cost-effective storage technologies and more specifically about storing data on the cloud. What is a cloud and why there are different types of clouds? She decides to watch again the video “[Public Cloud vs Private Cloud vs Hybrid Cloud](#)”.

Can you assist Alice in getting a deeper understanding of cloud-based storage? Please select the right answer(s). You may select more than one answer.

- A. **Clouds are smart, automated and adaptive**
- B. **Clouds are less efficient and cost effective that traditional Data Centers.**
- C. **Public clouds are hosted by a cloud service provider and tenants pay for services they actually use.**
- D. **Private Clouds provide higher scalability and lower control.**
- E. **Hybrid clouds are a combination of both private and public clouds enabling the creation of new innovative apps with uncertain demand.**

Correct answers: A, C, E

6. After reading the article “[Murder in the Amazon cloud](#)”, Vadali (2017), presenting the story of Code Space, which led to data deletion and the eventual shutdown of the company, Alice is more concerned about storage security.

What are the needed tasks for the school and herself personally, to keep the students ‘data safe’?

You may review again Fig. 2.15, as well as the Techniques for protecting information according to [Digital Preservation Handbook](#).

Please select the right answer(s). You may select more than one answer.

- A. **Strengthen software and operating systems.**
- B. **Do not abandon software when it becomes obsolete, you may need to reuse it.**
- C. **Use access controls to specify who is allowed to access digital material and the type of access that is permitted**
- D. **Train only the people whose security awareness is part of their duties.**
- E. **Build a short-term plan for security**

- F. **Use Encryption, a cryptographic technique which protects digital material by converting it into a scrambled form.**

Correct answers: A, C, F

- 7. **Alice watches the video “[How Toy Story 2 Almost Got Deleted: Stories From Pixar Animation: ENTV](#)” and thinks “What an unbelievable story!”**

She then starts laughing. The director could have avoided this “almost disaster” if he.

Please select the right answer.

- A. **had not typed the command RM***
- B. **had multiple independent copies of the digital material of the movie**
- C. **had used a combination of online and offline storage techniques for the copies of the digital material of the movie**
- D. **had kept the copies of the digital material of the movie geographically separated into different locations**
- E. **All the above.**

Correct answer: E.

- 8. **ACTIVITY/PRACTICE QUESTION (Short answer)**

Name some types of educational data that need long term preservation.

- 9. **ACTIVITY/PRACTICE QUESTION (Reflect on)**

We encourage you to elaborate on your response in the following reflective tasks. You may reflect on:

1. **Storage issues for preserving educational data**
2. **Good practices when preserving educational data**

2.3 Educational Data Ethics

2.3.1 Informed Consent

The video “[Introduction to data ethics](#)” (in useful video resources) introduces the basic principles of data ethics.

As Pentland states when describing Big Data, “*the ability to track, predict and even control the behaviour of individuals and groups of people is a classic example of Promethean fire: it can be used for good or ill*” (Pentland, 2013).

New regulations, like the **GDPR (General Data Protection Regulation)** (Regulation (EU), 2016) that we will discuss later on, along with recent events such as [the Cambridge Analytica and Facebook scandal](#), **have raised awareness of data ethics issues that can arise from data misuse (Open Data Institute, 2018a).**

Open Data Institute (ODI) (Broad et al., 2017), defines Data Ethics as.

a branch of ethics that evaluates data practices with the potential to adversely impact on people and society – in data collection, sharing and use.

Several frameworks, policies and guidelines have been developed to address data ethics issues, including [JISC’s code of practice](#) (Shacklett, 2016), updated in 2018, the [LACE](#) (Learning Analytics Community Exchange) framework in 2016 and the ICDE (International Council for Open and Distance Education) [Global guidelines](#) (Slade & Tait, 2019). To help identify potential ethical issues associated with a data project or activity and the steps needed to act ethically, Open Data Institute has also designed the [Data Ethics Canvas](#) in 2018 (Open Data Institute, 2018b).

We will further discuss the basic common principles of these practices in Chap. 3.

As emphasized by Shacklock (2016) “*Institutions should put in place clear ethical policies and codes of practices that govern the use of educational data. These policies should, at a minimum, address privacy, security of data and consent.*”

Before proceeding further, the brief video “[What is the GDPR?](#)” (in useful video resources) provides an overview of the European Union data protection rules, also known as the [EU General Data Protection Regulation \(or GDPR\)](#), that apply since 25 May 2018 to all entities who collect, store and process any personal data belonging to EU citizens and residents (even organisations that are not EU-based). GDPR has strengthened the conditions for consent (GDPR.eu, 2019).

We will soon discuss this new regulation and how should be applied by the various entities. First, let’s see what informed consent is all about.

Informed consent is declared by most international guidelines as one of the pivotal principles in Data Ethics and “is explicitly mentioned as a principle in article 7 of the International Covenant on Civil and Political Rights (1966), a United Nations Treaty” (European Commission, 2013).

According to Griffiths et al. (2016) “*Informed consent refers to the requirement for an individual to give consent for the collection and analysis of the data which they generate.*” While “*Transparency refers to the degree to which users can observe the ways in which the data they generate is used.*”

As per European Commission’s report (2013) regarding Ethics for Researchers “*Informed consent consists of three components: **adequate information, voluntariness and competence.***”

Thus, prior to consenting, individuals should be clearly informed of the data collection goals, possible adverse impacts and the means available to them to refuse or withdraw consent, without consequences, at any time.

Moreover, individuals must be competent to understand the information and should be fully aware of the consequences of their consent. Greater attention is required for **some special categories of people**, such as children, vulnerable adults and people with certain cultural or traditional backgrounds.

At this point, it is important to understand the distinction between **consent** and **informed consent**. For informed consent, we need to ensure that individuals genuinely understand how we intend to use their data e.g., by running focus groups and/or publishing explanatory documents.

As per [European Commission guidelines about GDPR](#), “*when a company or organisation asks for **consent to collect or reuse personal information**, the data subjects **have to make a clear action agreeing to this**, for example by signing a consent form or selecting yes from a clear yes/no option on a webpage*”... “*It is not enough to simply opt out, for example by checking a box saying they don’t want to receive marketing emails. **They have to opt in and agree to their personal data being stored and/or re-used for this purpose.***”

[European Commission](#) emphasizes that informed consent means that before you consent, you must **be given information** about the processing of your personal data, including at least:

- the **identity** of the organisation processing data;
- the **purposes** for which the data is being processed;
- the **type** of data that will be processed;
- the possibility to **withdraw** consent;
- where applicable, the fact that the data will be used **solely for automated-based decision-making**, including profiling;
- information about whether the consent is related to an **international transfer** of your data, the possible risks of data transfers to countries outside the EU if those countries are not the subject of a Commission adequacy decision and there are no adequate safeguards.

The way individuals are informed is crucial for the informed consent process. We should ensure that they fully realize the expected consequences of granting or withholding consent (Fig. 2.16).

With regards to the collection of personal data about **children**, additional protection should be granted since children are less aware of the risks and consequences of sharing data and of their rights.

In U.S., the foundational federal law on student privacy, the **Family Educational Rights and Privacy Act (FERPA)**, establishes student privacy rights by restricting with whom and under what circumstances schools may share students’ personally identifiable information. [DQC has developed a tool](#) that summarizes some of the main provisions of FERPA and can be used as a guide to help interested parties to understand when they need to take a closer look at the law or consult an expert.

[Under GDPR](#), any information addressed specifically to a child should be adapted to be **easily accessible, using clear and plain language**.

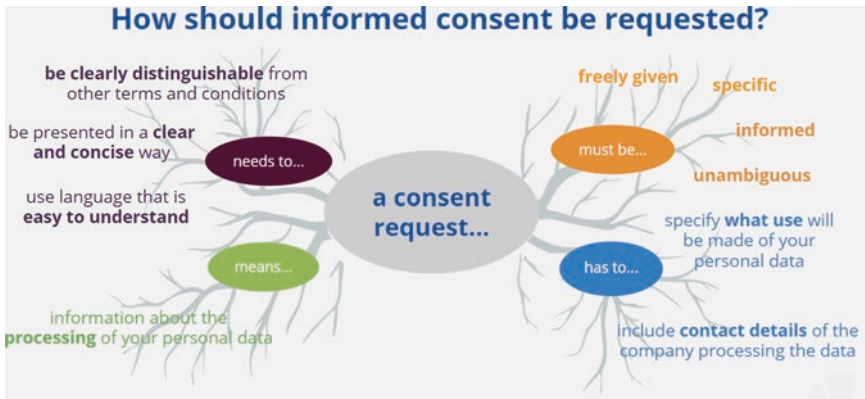


Fig. 2.16 Conditions for informed consent

For most online services (social networking sites) **the consent of the parent or guardian** is required in order to process a child’s personal data on the grounds of consent up to a certain age.

The age threshold for obtaining parental consent is established by each EU Member State and can be between 13 and 16 years, according to [National Data Protection Authority](#).

As per European Commission clarifications for the [Rights for Citizens](#), “Companies have to make reasonable efforts, taking into consideration available technology, to check that the consent given is truly in line with the law. This may involve implementing age-verification measures such as asking a question that an average child would not be able to answer or requesting that the minor provides his parents’ email to enable written consent”.

Within the context of education, there are quite different approaches relating to the consent in collecting learners’ data, according to national guidelines (when available).

Figure 2.17 depicts the main principles and challenges that should be taken under consideration to comply with GDPR. As presented, data-related activity can still be lawful, by complying with legal obligations e.g. GDPR, even though it may be considered that data is not treated ethically. Sclater (2017) also argues that “consent is required for use of sensitive data and in order to take interventions directly with students on the basis of the analytics. This implies that if the data in question are not considered ‘sensitive’, and do not form the basis for any intervention, consent is not required (on the basis that this may be considered as of legitimate interest)”.

Moreover, as per the [ICDE’s recent report \(2019\)](#), many institutions seek for consent to collect student data for additional purposes, beyond institutional reporting and basic student support, at the point of registration. As emphasized, “expectation that users should consent to uses of personal data unknown at the point of registration seems to be an unreasonable and unethical one.”

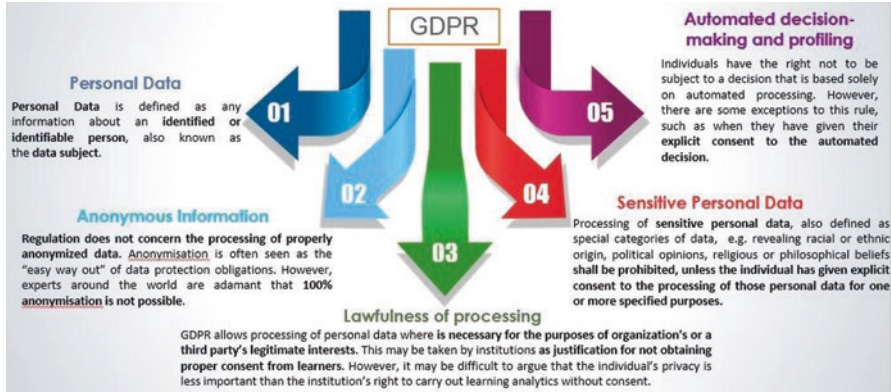


Fig. 2.17 The main principles and challenges that should be taken under consideration to comply with GDPR

An alternative approach supported by most of the existing guidelines ([Higher Education Commission](#), [JISC's code of practice](#), [ICDE Global guidelines](#)) might be to differentiate between the granting of initial consent for the collection of data and the obtaining of additional consent at the point where a specific personal intervention is proposed, or in the case where new data is incorporated into the institution's system, or existing data is used in new ways.

As concluded in [ICDE report \(2019\)](#) "national legislation will influence positions taken, but generally this principle (of consent) should be built around a minimum of informed consent (that is, transparency before registration)."

You may also review this video "[Why develop a data science code of ethics?](#)" (in useful video resources) where experts from the data science community explain why it's important to have a code of ethics.

Questions and Teaching Materials

1. After watching the video introducing Data Ethics Principles "[Introduction to Data Ethics](#)", Alice is really concerned. Companies are collecting so much data every day. According to the video, Google can track your searches on your individual devices, even if you are not logged in to your account, up to:
 - A. 7 days
 - B. 2 months
 - C. 6 months
 - D. 3 years

Correct answer: C

2. **Before using the flipped classroom initiative, Alice wants to study Grade 9 students' perceptions of technology, using an online questionnaire she made with Google Forms.**

Alice wants to prepare an informed parental consent form for her students (as they are under 15) in order to participate in the students' perceptions of technology survey, but she is a bit confused with all this information.

Can you help Alice to have a better understanding?

- A. **Prior to consenting, individuals should be clearly informed of how the data will be used**

- **True**
- **False**

Correct answer: True

- B. **When individuals give consent for the collection and analysis of the data which they generate, they cannot refuse or withdraw their consent**

- **True**
- **False**

Correct answer: False

- C. **EU General Data Protection Regulation (or GDPR), apply since 25 May 2018 even to organisations that are not EU-based, as long as they collect, store and process any personal data belonging to EU citizens and residents.**

- **True**
- **False**

Correct answer: True

3. **You give some advice to Alice in order to help her prepare the consent form for the students' perceptions of technology study. Select all that apply.**

A consent request must:

- A. **Include contact details of the company processing the data**
- B. **Be anonymized**
- C. **Include information about the possibility of withdrawing consent**

- D. **Be freely given**
- E. **Be included in the terms and conditions**
- F. **Be presented in a formal language**
- G. **Specify the purpose of the data process**
- H. **Specify the type of data that will be processed**

Correct answers: A, C, D, G, H

4. **Alice has a colleague, Betty, who has just come on board and wants to conduct an online survey with her 17-year-old students about their eating habits. Betty asks Alice if it is necessary to collect parental consent in order to process her students' personal data.**

Help Alice decide if a consent as a parent or guardian is required in order to process students' personal data

- **Yes**
- **No**

Correct answer: No.

5. **Alice's Secondary High-School relies upon the sixth lawful basis (public task basis) to justify the processing of personal data (according to GDPR) where processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller.**

Is this lawful basis (public task basis) appropriate for Alice in order to take interventions directly with students on the basis of the participation data recorded within the Learning Management System?

Help Alice find the correct answer

- **Yes**
- **No**

Correct answer: No.

6. **In the video “[Why develop a data science code of ethics?](#)”, Paula Goldman, VP/Head of Omidyar Network's Tech and Society Solutions Lab, claims that data and algorithms are neutral.**

- **True**
- **False**

Correct answer: False.

7. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to elaborate on your response in the following reflective task. You may reflect on:

1. **What information must be given to individuals, whose data is collected. You can search for additional information on the [European Commission's website](#).**
2. **Using information from the [European Commission website](#), create an infographic presenting the General Protection Data Regulations.**

2.3.2 Sensitive Educational Data Protection

Balancing digital learning with privacy and security is essential to fostering a successful digital culture (iKeepSafe, 2017).

Privacy is a fundamental human right and a core value in the functioning of democratic societies. As already discussed in the previous topics, with the exponential progress in the field of information and communication technologies and in the light of **rapid development of Educational Data Analytics** on a global basis, **new challenges to privacy and data protection have emerged**.

The “[Privacy Overview for K12 Teachers and Administrators](#)” video (in useful video resources) provides us with an overview of the privacy issues that may arise and growing concerns about educational data privacy. Is educational data privacy over in the digital age?

In the [Quantified Student infographic](#) you may see what a day in the data-driven life of most measured and monitored student in the history of education, looks like.

“*The data collection begins even before he steps into the school,*” says Khaliah Barnes, director of the Student Privacy Project at the [Electronic Privacy Information Center](#). “*The issue is that this reveals specifically sensitive information,*” says Barnes (Hill, 2014).

Moreover, as Jose Ferreira CEO at Knewton (one of the biggest actors in the field of educational technology software), points out “*We literally know everything about what you know and how you learn best, everything.*” Ferreira calls education “*the world’s most data-mineable industry by far*” (Hill, 2014).

Do educational data analytics challenge the principles of data protection? Is privacy a show-stopper? How privacy is guaranteed/secured, especially if minors and/or sensitive data is involved?

The European position has been expressed in the European Commission's report: "[New Modes of Learning and Teaching in Higher Education](#)" (European Commission, 2014). In recommendation 14, the Commission clearly stated: "*Member States should ensure that legal frameworks allow higher education institutions to collect and analyse learning data. The full and informed consent of students must be a requirement and the data should only be used for educational purposes*", and in recommendation 15: "*Online platforms should inform users about their privacy and data protection policy in a clear and understandable way. Individuals should always have the choice to anonymise their data.*" This is a **widely accepted framework mirrored** in the laws of **multiple nations and international organisations** including many U.S. states (Drachsler & Greller, 2016).

Thus, it is essential that all educators understand how learners' personal information is used and adequately protect learners' data in order to strengthen the trust of all parties involved and encourage their participation in digital learning.

In the video by the Data Quality Campaign "[Who Uses Student Data?](#)" (in useful video resources), it is emphasized that most personal student information stays local. Districts, states, and the federal government all collect data about students for important purposes like informing instruction and providing information to the public. But the type of data collected, and who can access them, is different at each point.

As clearly stated in [Foundational Principles for Using and Safeguarding Students' Personal Information](#) developed by a coalition of US national education organisations "*Everyone who uses student information has a responsibility to maintain the privacy and the security of students' data, especially when these data are personally identifiable.*"

The basic information security techniques, as specified by [Digital Preservation Handbook](#), include:

Encryption

- Encryption is a cryptographic technique which protects digital material by converting it into a scrambled form. The use of a key is required to unscramble the data and convert it back to its original form.

Access Control

- Access control enables an administrator to specify who is allowed to access digital material and the type of access that is permitted (for example read only, write).

Redaction

- Redaction refers to the process of identifying and removing or replacing confidential or sensitive information, using anonymisation or pseudonymisation.

Now that we have a better understanding of the different types of data as categorized in terms of privacy, we will further review the levels of data as specified under GDPR.

The Fig. 2.18 presents the main categories of personal data as defined by GDPR.

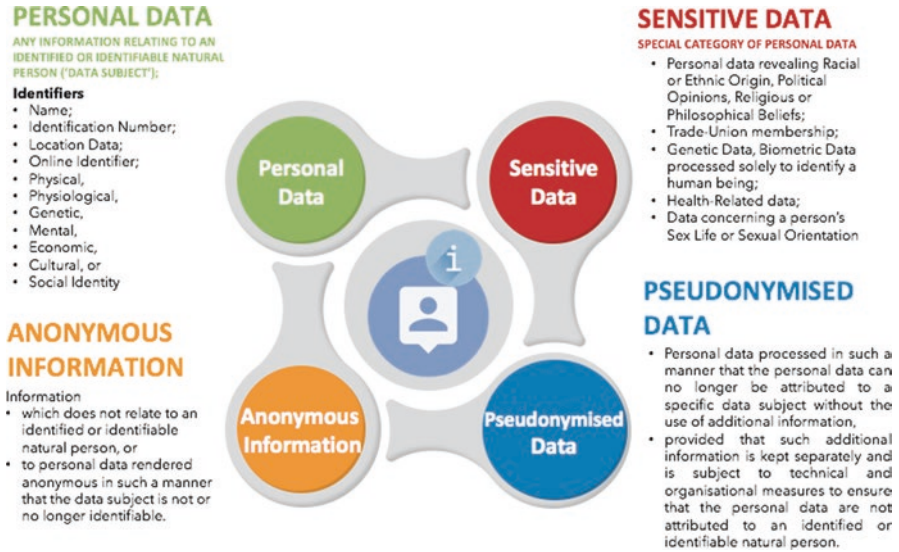


Fig. 2.18 The main categories of personal data as defined by GDPR

We need to pay extra attention to **sensitive** (special category of personal data) since an organisation can only process this data under specific conditions (explicit consent may be needed). Even **personal data**, as clarified under GDPR, “*should only be processed where it isn’t reasonably feasible to carry out the processing in another manner. Where possible, it is preferable to use anonymous data. Where personal data is needed, it should be adequate, relevant, and limited to what is necessary for the purpose* (‘data minimisation’).”

Once data is truly **anonymised** and does no longer contain any identifying elements, the anonymisation is **irreversible** and individuals are no longer identifiable, the data will not fall within the scope of the GDPR and it becomes easier to use.

Before anonymization, we should consider the purposes for which the data is to be used. Anonymisation may devalue the data, so that it is no longer useful for specific purposes.

The **ICO’s Code of Conduct on Anonymisation** provides further guidance on anonymisation techniques (UCL, 2018). Unlike anonymisation, in **pseudonymised data** personally identifiable material is replaced with artificial identifiers. Pseudonymised personal data can still fall within scope of the GDPR, depending on how difficult it is to attribute the pseudonym to a particular individual.

Whether ‘de-identified’ or pseudonymised data is in use, there is a residual risk of re-identification. For example, anonymisation is often seen as the “easy way out” of data protection obligations. However, experts around the world are adamant that 100% anonymisation is not possible. Anonymised data can rather easily be de-anonymised when they are merged with other information sources. (Drachler & Grellner, 2016).

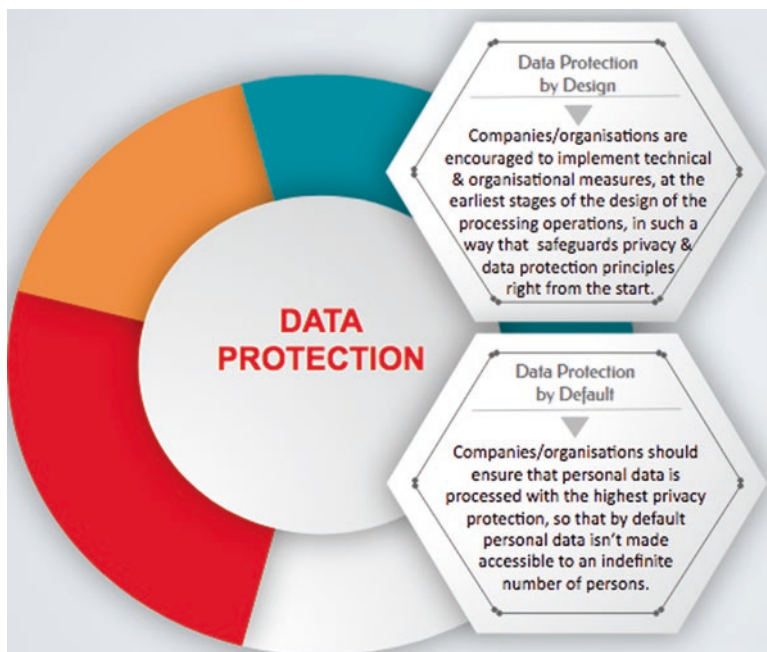


Fig. 2.19 Data protection by design and data protection by default

L. Sweeney (2000) presented that it's possible to personally identify 87% of the U.S. population based on just three data points: five-digit ZIP code, gender and date-of-birth (Wes, 2018). Later on, in 2006, the AOL release of users' search logs (Hansell, 2006) and the case of the Searcher No. 4417749, as recorded in "A Face Is Exposed for AOL Searcher No. 4417749" by M. Barbaro and T. Zeller (2006) of New York times, was one of the first widely known cases of re-identification. In 2007, the Netflix case (Narayanan & Shmatikov, 2008), followed when researchers de-anonymized some of the Netflix data by matching rankings and timestamps with public information on the Internet Movie Database. As per Hill (2012), in 2012 the retail company Target, using behavioural advertising techniques, managed to identify a pregnant teen girl from her web searches and sent her relevant vouchers at home. (D'Acquisto et al., 2015).

Thus, though de-identification techniques can reduce the risks to the data subjects concerned and help organisations to meet their data-protection obligations, we need to assess properly the adequacy of these methods so as to decide whether further steps to de-identify the data are necessary (UCL, 2018).

The GDPR introduces two new principles: **data protection by design and data protection by default**, whose definitions are presented in Fig. 2.19.

As specified in GDPR (Regulation (EU), 2016), the protection of the rights and freedoms of natural persons with regard to the processing of personal data require that appropriate technical and organisational measures be taken which meet in particular the principles of data protection by design and data protection by default.

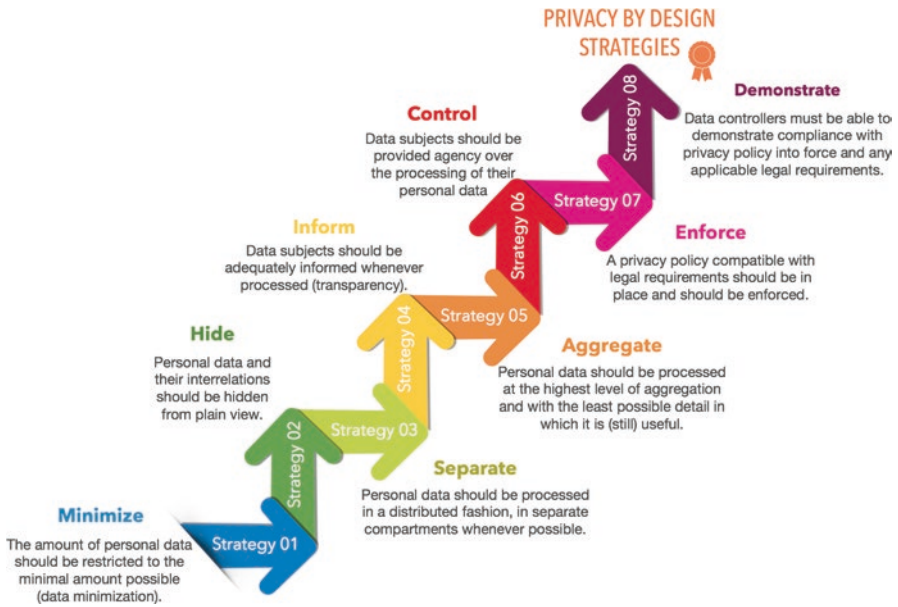


Fig. 2.20 Eight privacy by design strategies, as proposed by the European Union Agency for Network and Information Security (D’Acquisto et al., 2015)

“Data protection by design minimises privacy risks and increases trust”, while “Data protection by default entails ensuring that your company always makes the most privacy friendly setting the default setting” (European Union, 2018).

An example of Data protection by design is the use of pseudonymisation & encryption and examples for Data protection by default include “data minimisation” (only the data necessary should be processed), the limited accessibility as well as the short storage period.

Let’s now review further the privacy by design strategies and the storage privacy (Data protection by design), as well as the Storage Limitation (Data protection by default).

Figure 2.20 depicts eight Privacy By Design Strategies, as proposed by the European Union Agency for Network and Information Security (D’Acquisto et al., 2015). These strategies enable us to identify the data protection and privacy requirements early in the educational analytics value chain and subsequently to implement the necessary technical and organizational measures. One of the most significant privacy enhancing technologies that can be used for implementing such strategies, is storage privacy.

Privacy challenges should be, seen as opportunities that, if appropriately handled, can build trust in the big data ecosystem for the benefit of both users and big data industry (D’Acquisto et al., 2015).

Danezis et al. (2014), in this report “Privacy and Data Protection by Design”, defines Storage Privacy as “the ability to store data without anyone being able to read

(let alone manipulate) them, except the party having stored the data (called here the data owner) and whoever the data owner authorises.”

As specified further in the report, *“a major challenge to implement private storage is to prevent non-authorized parties from accessing the stored data. If the data owner stores data locally, then physical access control might help, but it is not sufficient if the computer equipment is connected to a network: a hacker might succeed in remotely accessing the stored data. If the data owner stores data in the cloud, then physical access control is not even feasible.”*

A straightforward option for storage privacy is storing the data, either locally or in cloud storage, in encrypted form. One can use full disk encryption (FDE) or file system-level encryption (FSE). As clarified in the report, *“encryption and decryption operations must be carried out locally, not by remote service, because both keys and data must remain in the power of the data owner if any storage privacy is to be achieved. The report specifies that outsourced data storage on remote clouds is practical and relatively safe as long as only the data owner, not the cloud service, holds the decryption keys. Such storage may be distributed for added robustness to failures.”*

When it comes to Data protection by default, **Storage limitation** is one of the key conditions for processing personal data under GDPR. It replies to a simple question *“For how long can data be kept and is it necessary to update it?”* Regulation’s answer is straightforward *“You must ensure that personal data is stored for no longer than necessary for the purposes for which it was collected”*. There are 6 basic guidelines, **specified clearly by GDPR**, which you need to take under consideration when storing personal data (Fig. 2.21).

Before closing this chapter, it is essential to analyse the **individuals’ rights**. The main reason for the introduction of GDPR is to allow European Union citizens to **better control their personal data**. More specifically is designed to:

- **Harmonize data privacy laws across Europe,**
- **Protect and empower all EU citizens’ data privacy**
- **Reshape the way organisations across the region approach data privacy.**

GDPR applies to “all companies operating in the EU, wherever they are based” (**European Commission, 2018**). The GDPR introduces stronger rights for data subjects (Intersoft Consulting, 2018), and creates new obligations for data controllers (the person or body handling the personal data).

Figure 2.22 presents **individuals’ rights** so as to have control over their personal data, **under GDPR**. To exercise individuals’ rights they should contact the company or organisation processing their personal data, also known as the controller. If the company/organisation has a **Data Protection Officer** (‘DPO’) they may address their request to the DPO. The company/organisation must respond to their requests without undue delay and **at the latest within 1 month**.

When the personal data, for which a company/organisation is responsible, is disclosed, either accidentally or unlawfully, to unauthorised recipients or is made temporarily unavailable or altered, a data breach occurs. In case a data breach occurs and the breach poses a risk to individual rights and freedoms, the company/

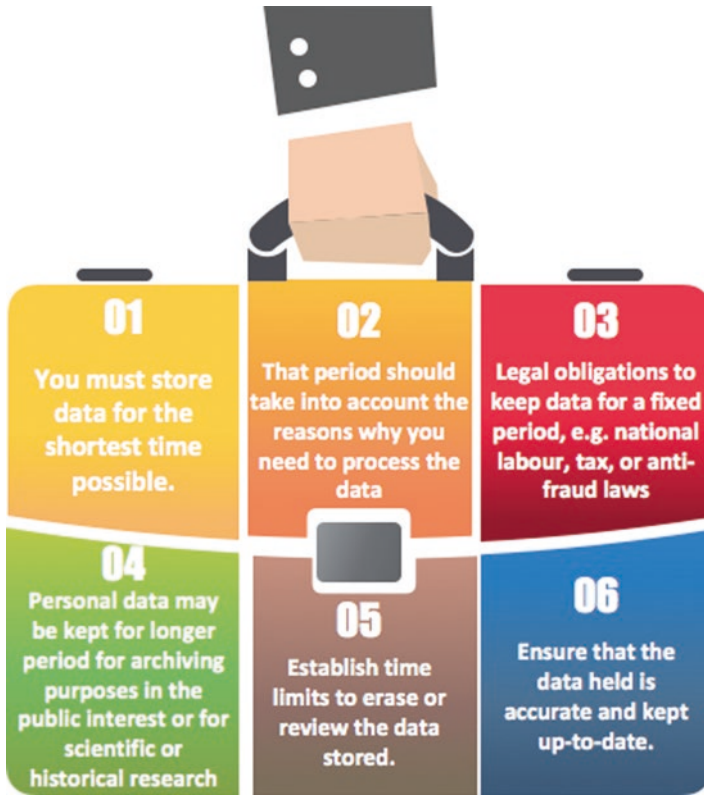


Fig. 2.21 Six basic guidelines, which you need to take under consideration when storing personal data

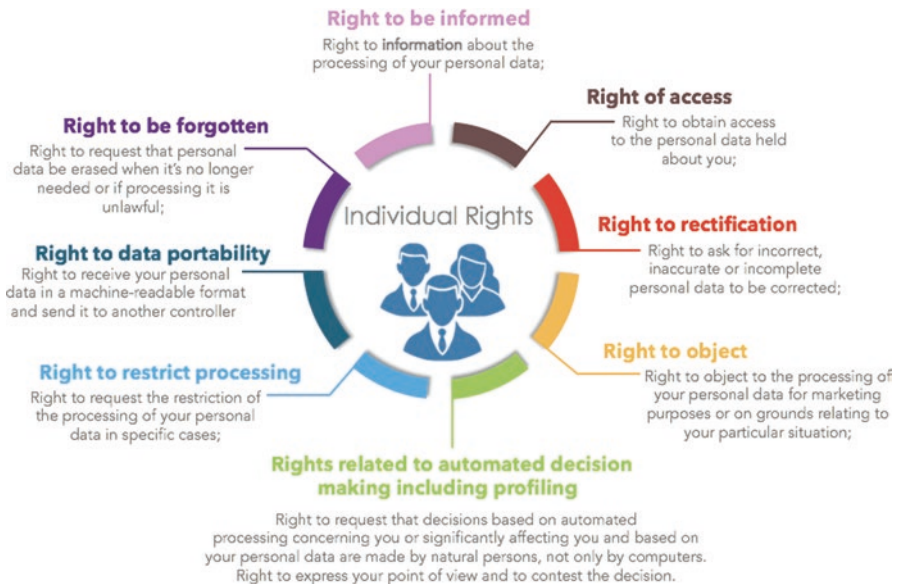


Fig. 2.22 Individuals' rights so as to have control over their personal data, under GDPR

organisation should notify its [Data Protection Authority](#) (DPA) within 72 hours after becoming aware of the breach. Depending on whether or not the data breach poses a high risk to those affected, a business may also be required to inform all individuals affected by the data breach (European Commission, 2018h).

Whenever processing is likely to result in a high risk to the rights and freedoms of individuals, as specified by GDPR, a [Data Protection Impact Assessment](#) (DPIA) is required. A DPIA is required at least in the following cases:

- a systematic and extensive evaluation of the personal aspects of an
- individual, including profiling;
- processing of sensitive data on a large scale;
- systematic monitoring of public areas on a large scale.

[National Data Protection Authorities](#), in collaboration with the [European Data Protection Board](#), may provide lists of cases where a DPIA would be required. As emphasized, *“the DPIA should be conducted before the processing and should be considered as a living tool, not merely as a one-off exercise. Where there are residual risks that can’t be mitigated by the measures put in place, the DPA must be consulted prior to the start of the processing”*.

Figure 2.23 provides the 3 Basic Steps to Identify and Protect Sensitive Data, as per Krueger (2017).

A DPIA should be conducted as early as possible in the project lifecycle, so that its findings and recommendations can be incorporated into the design of the processing operation (itgovernance).

You may also review the video [“Protecting Student-Data Privacy: An Expert’s View”](#) (see useful video resources) where Fordham University Law Professor Joel Reidenberg talks with Education Week Correspondent John Tulenko about student data and the best ways to keep it secure.



Fig. 2.23 The 3 Basic Steps to Identify and Protect Sensitive Data

Questions and Teaching Materials

1. **Alice is a bit confused. Several state and federal laws require privacy protection for students and children. In the video she just watched, “[Privacy Overview for K12 Teachers and Administrators](#)”, what laws are mentioned concerning data privacy for children?**

There is more than one correct answer. Help Alice select the right ones

- A. FERPA
- B. CIPA
- C. COPPA
- D. CAPTA

Correct answers: A, C

2. **From watching the “[Who Uses Student Data?](#)” video, Alice understands that teachers have access only to de-identified data (i.e. information about individual students but with identifying information removed).**

Is Alice’s understanding correct?

Please select the correct answer:

- Yes
- No

Correct answers: No.

3. **For the purposes of research, Alice intends to release student data.**

Alice asks to be informed by the responsible DPO on school’s policy and guidelines to protect students’ data privacy, confidentiality, integrity and security. She becomes aware of personal and sensitive data handling and the use of anonymisation and pseudonymisation to remove personally identifiable information.

As student data might be released for the purposes of research, all names, postal codes and other identifiable data are removed. Completely removing fields that could be used in any way to identify a person is considered a strong form of

- A. data pseudonymisation
- B. data anonymisation

Please select the correct term to complete the sentence.

Correct answer: B

4. **Alice has concerns about her students' records, and more specifically about medical reports related to student's learning difficulties being accessed by unauthorized third persons. She contacts the responsible DPO and is informed about the appropriate technical and organisational measures taken by the school, so as to secure data protection by design and by default.**

More specifically the DPO explains to Alice that the School Information System (SIS) has a mechanism for comprehensively logging who consulted the medical reports and preventing unauthorized access to these sensitive data. Moreover, personal and sensitive data are pseudoanonymized and "data minimization" (only the data necessary should be processed) is used.

Alice feels secure because the technical and organisational measures being taken meet in particular the principles of data protection by design and data protection by default.

Is Alice correct in feeling secure?

Please select the correct answer:

- **Yes**
- **No**

Correct answer: Yes

5. **Storage privacy is about preventing non-authorized parties from accessing the stored data. This can be achieved only when encryption and decryption operations are carried out locally, not by remote service, because both keys and data must remain in the power of the data owner.**

Alice assumes that if any storage privacy is to be achieved, then data must be stored locally and cloud storage should be avoided.

Do you agree with the assumption of Alice?

Please select the correct answer:

- **Yes**
- **No**

Correct answer: No.

6. **Alice's institution runs a recruitment office and for that purpose it collects CVs and keeps records of persons seeking employment.** They keep recruitment application forms and interview notes (for unsuccessful candidates) for 5 years in case they need them without taking any measures for updating the CVs

Alice doubts that the storage period is proportionate to the purpose of finding employment and thinks that this is not compliant with GDPR. Do you agree with Alice?

You may review “[For how long can data be kept and is it necessary to update it? | European Commission \(europa.eu\)](https://european-commission.europa.eu)”.

Please select the correct answer:

- Yes
- No

Correct answer: Yes.

7. Alice is trying to understand the rights for data subjects described in GDPR. She reviews “[Data protection and online privacy – Your Europe \(europa.eu\)](https://european-commission.europa.eu)” and “[It’s your data – take control – Data protection in the EU \(europa.eu\)](https://european-commission.europa.eu)”.

Help Alice match the cases to the appropriate individual right.

Case	Individual Right
A. You’ve bought goods from an online retailer. You can ask the company to give you the personal data they hold about you, including: your name and contact details, credit card information and dates and types of purchases.	1. Right to object
B. You bought two tickets online to see your favorite band play live. Afterwards, you’re bombarded with adverts for concerts and events that you’re not interested in. You inform the online ticketing company that you don’t want to receive further advertising material.	2. Right to rectification
C. You apply for a new insurance policy but notice the company mistakenly records you as a smoker, increasing your life insurance payments.	3. Right to be forgotten
D. When you type your name into an online search engine, the results include links to an old newspaper article about a debt you paid long ago.	4. Right of Access
E. You apply for a loan with an online bank. You are asked to insert your data and the bank’s algorithm tells you whether the bank will grant you the loan and gives the suggested interest rate.	5. Right to data portability
F. You’ve found a cheaper electricity supplier. You ask your existing supplier to transmit your data directly to the new supplier, if it’s technically feasible or to return your data to you in a commonly-used and machine readable format so that it can be used on other systems.	6. Rights related to automated decision making

Correct answer: A4 – B1 – C2 – D3 – E6 – F5.

8. Alice's institution recruitment office decides to implement an innovative recruitment procedure which includes e-recruitment tools automatically pre-selecting/excluding candidates without human intervention. Alice thinks that a **Data Protection Impact Assessment (DPIA)** is required.

Study the “[Decision of the European Data Protection Supervisor of 16 July 2019 on DPIA Lists issued under Articles 39\(4\) and \(5\) of Regulation \(EU\)](#)” and select the “**Criteria for processing ‘likely to result in high risk’**”, that will trigger DPIA in the case of Alice's institution new recruitment procedure (select 3 criteria).

Which are the criteria for processing “likely to result in high risk”?

1. Systematic and extensive evaluation of personal aspects or scoring, including profiling and predicting.
2. Automated-decision making with legal or similar significant effect: processing that aims at taking decisions on data subjects
3. Systematic monitoring: processing used to observe, monitor or control data subjects, especially in publicly accessible spaces. This may cover video-surveillance but also other monitoring, e.g. of staff internet use.
4. Sensitive data or data of a highly personal nature: data revealing ethnic or racial origin, political opinions, religious or philosophical beliefs, trade-union membership, genetic data, biometric data for uniquely identifying a natural person, data concerning health or sex life or sexual orientation, criminal convictions or offences and related security measures or data of highly personal nature.
5. Data processed on a large scale, whether based on number of people concerned and/or amount of data processed about each of them and/or permanence and/or geographical coverage
6. Datasets matched or combined from different data processing operations performed for different purposes and/or by different data controllers in a way that would exceed the reasonable expectations of the data subject.
7. Data concerning vulnerable data subjects: situations where an imbalance in the relationship between the position of the data subject and the controller can be identified.
8. Innovative use or applying technological or organisational solutions that can involve novel forms of data collection and usage. Indeed, the personal and social consequences of the deployment of a new technology may be unknown.
9. Preventing data subjects from exercising a right or using a service or a contract.

Correct answer: 1, 2, 8.

9. **According to Professor Joel Reidenberg, in the video “Protecting Student-Data Privacy: An Expert’s View”, the worst that could happen because of bad data practices is:**
- A. Students being used as guinea pigs for the development of commercial products
 - B. Educational harm to children, where they are being improperly labelled
 - C. The development of programs that assess teachers’ performance
 - D. The development of flexible mechanisms so parents can consent and opt-in to additional uses of data

Correct answer: B.

10. **ACTIVITY/PRACTICE QUESTION (Reflect on)**

We encourage you to elaborate on your response in the following reflective task. You may reflect on:

1. **Privacy issues for preserving educational data**
2. **Educational data protection**

2.4 Concluding Self-Assessed Assignment

2.4.1 Introduction

Both Alice and you have come a long way in your understanding of the power of educational data as a key success factor for online and blended teaching and learning, as well as of the fundamentals of Educational Data Collection and Management, including issues related to ethics and privacy.

You are now ready to develop further your Educational Data Literacy Competences focusing on Educational Data Analysis, Comprehension and Interpretation.

In order to proceed, you are requested to complete a concluding self-assessed assignment. This self-assessed assignment is a real life scenario activity (based on the use case of our teacher Alice), using a rubric across three proficiency levels and an exemplary solution rating. When you have completed this assignment, you will assess it yourself, following the rubric which will list the criteria required and give guidelines for the assessment.

This self-assessed assignment procedure consists of 5 steps:

- **Step 1.** Real life scenario
- **Step 2.** Getting familiar with the assessment rubric

- **Step 3.** Prepare your answer
- **Step 4.** Review a sample solution
- **Step 5.** Self-evaluate your answer

2.4.2 Step 1. Real Life Scenario

Alice is an enthusiastic English Language teacher who has just been appointed in an Experimental High School, in Athens, Greece. She wants to use student data to gain insights and plan her teaching activities accordingly, so as to improve this year's Grade 9 students' academic performance.

Alice contacts Mr. Adams, appointed as school's Data Protection Officer (DPO), to secure all necessary approvals for the sources handled by her school or by the corresponding district. As soon as Alice signs the required data protection consent form, she gets permission and downloads the datasets from the several sources.

Alice also requests to grant her access to the LMS used by the school (a new teacher account is created by the LMS administrator). Before implementing her flipped classroom strategy, she contacts the school's DPO again to discuss any legal and ethical issues she needs to pay attention to. As advised by the DPO, she accesses the LMS and via the "*User agreements page*", she reviews the existing user agreements and confirms that **signed informed consent** has been given for all participating students (either parental consent on behalf of minors or directly by the students, as defined by [National Data Protection Authority](#)).

Alice realizes that she must update the current consent form based to the new General Data Protection Regulation Policy.

You need to help Alice to prepare a new consent form for the students participating in her flipped classroom model.

2.4.3 Step 2. Getting Familiar with the Assessment Rubric

Alice reviews the **Initial Consent Form**.

Please help Alice to evaluate this Initial Consent Form using the **Rubric for assessing the Consent Form** and to identify potential issues.

ACTIVITY/PRACTICE QUESTION (Reflect on) We encourage you to elaborate on your response about the evaluation of the Initial Consent Form created by Alice, in the following reflective task. You may reflect on:

1. *Does this consent form comply with GDPR consent requirements?*
2. *If not, what would you advise Alice to modify, so that this consent form is GDPR compliant and limits her school's exposure to regulatory penalties?*

2.4.3.1 Initial Consent Form

Introduction

Welcome to Athens Experimental High School (the “School” or “We”) Learning Management System (LMS). The School provides this LMS to you subject to the following Terms of Use and Privacy Policy (together, the “Terms”). When you use this LMS, you agree to abide by these Terms. If you do not agree to abide by these Terms, you may not use this LMS. Please read the Terms carefully.

The School reserves the right to make changes to this LMS and to modify the Terms at any time at its sole discretion. We encourage you to review the Terms frequently for modifications. By your use of this LMS, you agree to abide by any such modifications to the Terms, which are binding on you.

Privacy Policy

This Privacy Policy describes the School’s agreement with you regarding how we will handle certain information on the LMS. This Privacy Policy does not address information obtained from other sources such as submissions by mail, phone or other devices or from personal contact. By accessing the LMS and/or providing information to the School on the LMS, you consent to the collection, use and disclosure of certain information in accordance with this Privacy Policy.

Information Collected on Our LMS:

If you merely download material or browse through the LMS, our servers may automatically collect certain information from you which may include: (a) the name of the domain and host from which you access the Internet; (b) the browser software you use and your operating system; and (c) the Internet address of the website from which you linked to the LMS. The information we automatically collect may be used to improve the LMS to make it as useful as possible for our visitors; however, such information will not be tied to the personal information you choose to provide to us.

We do collect and keep personally identifiable information when you choose to voluntarily register to the LMS and submit such information. After your registration, we retain the information you submit for our records and to contact you from time to time. Please note that if we decide to change the manner in which we use or retain personal information, we may update this Privacy Policy, at our sole discretion.

Disclosure of Personal Information to Third Parties:

The School does not rent or sell personal information that you choose to provide to us nor does the School disclose credit card or other personal financial information to third parties other than as necessary to complete a credit card or other financial

transaction or as required by law. The School does engage certain third parties to perform functions and provide services, including, without limitation, hosting and maintenance, customer relationship, database storage and management, payment transaction and direct marketing campaigns. We will share your personal information with these third parties, but only to the extent necessary to perform the functions and provide the services, and only pursuant to binding contractual obligations requiring such third parties to maintain the privacy and security of your data.

Receiving Promotional Materials:

We may send you information or materials such as newsletters, ebooks, whitepapers by e-mail or postal mail when you submit your address via the LMS. By your registration in the LMS, you are consenting to our sending you such information or materials.

If you do not want to receive promotional information or material, please send an email with your name, mailing address and email address to athens.expschool.online@gmail.com. When we receive your request, we may take reasonable steps to remove your name from such lists.

Cookies

A cookie is a small text file that a website can place on your computer's hard drive for record-keeping or other administrative purposes. Our LMS may use cookies to help to personalise your experience on the LMS. Although most web browsers accept cookies automatically, usually you can modify your browser setting to decline cookies. If you decide to decline cookies, you may not be able to fully use the features of the LMS. Cookies may also be used at certain sites accessible through links on the LMS.

Links to Other Websites:

The School is not responsible for the practices or policies of the websites linked to or from the LMS, including without limitation their privacy practices or policies. If you elect to use a link that accesses another party's website, you will be subject to that website's practices and policies.

Terms of Use

For Informational Purposes Only

The School makes available the information on this Website for informational purposes only. You are solely responsible for the information you provide on this Website and for the information you use that you view on this Website. Information on this Website is not intended to be a replacement for direct consultation with the School; if you have questions or concerns, please contact the School directly.

Copyright and Trademark Information

The content included on this LMS, such as data, text, graphics, logos, images and software and its compilation is the property of the School and/or its content suppliers and is protected by copyright and trademark laws. In the event you upload any content including, without limitation, photographs or videos to this LMS, you (i) represent to the School and its affiliates that you have all rights necessary to upload the content; (ii) agree to indemnify the School and its affiliates for any third party infringement or other claims related thereto; and (iii) hereby license to the School and its affiliates a perpetual non-cancellable royalty-free license to use such uploaded content for any purposes in any media now existing or hereafter developed.

License for Your Use

For any period of time that you use this LMS and abide by these terms, the School grants to you a limited, revocable and nonexclusive license to access this LMS for your use but not to copy, download or modify it, or any portion of it, except with the express written consent of the School. This LMS or any portion of this LMS may not be reproduced, duplicated, copied, sold, visited or otherwise exploited without the express written consent of the School. You may not utilize framing to enclose any trademark, logo, content or other proprietary information contained on this LMS without the express written consent of the School. You may not use any meta tags or any other “hidden text” utilizing the School or its affiliates’ name or trademarks without the School’s express written consent.

You agree to use this LMS only for lawful purposes, and you acknowledge that your failure to do so may subject you to civil or criminal liability. You are responsible for ensuring that any materials you upload, post or submit to this LMS do not violate the copyright, trademark, trade secret or other personal or proprietary rights of any third party and you hereby agree to indemnify the School for any third party infringement or personal rights claims. You agree not to disrupt, modify, or interfere with this LMS or its associated software, hardware and servers in any way and you agree not to impede or interfere with others’ use of this LMS. You further agree not to alter or tamper with any information or materials on or associated with this LMS. Any unauthorized use or violation of these terms automatically terminates any permission or license granted by the School to access and use this LMS.

External Links

This LMS may provide links or references to third party websites or applications, including without limitation, third party websites or applications of advertisers or of providers of informational articles or other users. The School is not responsible for any information you choose to provide to those third party websites or applications; any information, products or services you acquire from those third party websites or applications, or any damages arising from your access to or use of those third party websites or applications.

Any links to third party websites and applications are provided as a convenience to the visitors of this LMS and any inclusion of any such links in this Website does not imply an endorsement or warranty of the third party websites or applications or their security, content, products, offerings or services. You are cautioned that any third party websites or applications are governed by their own terms of use and privacy policies, so when linking you should make sure to visit the appropriate pages of those third party websites or applications to determine what terms of use and privacy policies will apply to your use.

- **YES, I GIVE CONSENT FOR MY CHILD TO PARTICIPATE IN THE ONLINE COURSE AND AGREE TO THE CONSENT AS NOTED ABOVE.**
- **NO, I DO NOT GIVE CONSENT FOR MY CHILD TO PARTICIPATE IN THE ONLINE COURSE AND AGREE TO THE CONSENT AS NOTED ABOVE.**

Adapted from: <https://www.whitbyschool.org/privacy-policy>

2.4.3.2 Rubric for Assessing the Consent Form

Criteria	1 Unacceptable	3 Good/Solid	5 Exemplary
Language	The consent request is presented neither in a clear, nor in a concise way, using language that is not easy to understand	The consent request is presented in a quite clear and concise way, using language that is quite easy to understand	The consent request is presented in a very clear and concise way, using language that is very easy to understand
Explicit and Distinguishable	The consent request is not explicit or distinguishable from other pieces of information.	The consent request is quite distinguishable from other pieces of information but is not given via a positive act.	The consent request is clearly distinguishable from other pieces of information, given via an electronic tick-box that the individual has to explicitly check online
Freely given consent	The individual does not have a free choice.	The individual has a free choice and it is quite clear how to refuse consent without being at a disadvantage.	The individual has a free choice and it is very clear how to refuse consent without being at a disadvantage.
Possibility to withdraw the given consent	The consent form does not include the possibility to withdraw consent	The consent form includes the possibility to withdraw consent, but does not explain how to do it.	The consent form includes the possibility to withdraw consent and explains clearly how to do it.

(continued)

Criteria	1 Unacceptable	3 Good/Solid	5 Exemplary
Rights of the data subject	The individuals are not informed about their rights as a data subject (GDPR Art.12 to 23)	Rights of the data subject (GDPR Art.12 to 23) are somehow stated but the modalities to exercise these rights are not clear.	Individuals are clearly informed about their rights as a data subject (GDPR Art.12 to 23) and they can effectively exercise these rights
Identity of the organisation processing data	The consent form does not include the identity of the organisation processing data	The consent form includes quite clearly the identity of the organisation processing data	The consent form includes very clearly the identity of the organisation processing data
Purposes for which the data is being processed	The consent form does not explain the purposes for which the data is being processed	The consent form explains quite clearly the purposes for which the data is being processed	The consent form explains very clearly the purposes for which the data is being processed
Describes the type of data that will be processed	The consent form does not describe the type of data that will be processed	The consent form describes the type of data that will be processed	The consent form describes in detail the type of data that will be processed
International transfer of data	The consent form does not include information about whether the consent is related to an international transfer of your data	The consent form includes quite clearly information about whether the consent is related to an international transfer of your data	The consent form includes clearly information about whether the consent is related to an international transfer of your data

2.4.4 Step 3. Prepare Your Answer

Please assist Alice in preparing a consent form for the students participating in the online course for the flipped classroom initiative.

ACTIVITY/PRACTICE QUESTION (Reflect on) We encourage you to elaborate on your response about the preparation of the consent form for Alice’s students participating in the online course for the flipped classroom initiative, in the following reflective task. You may reflect on:

1. *How should the consent form be formulated so that Alice can obtain consent compliant with GDPR requirements?*
2. *What are the key features to create an effective opt-in consent form that works under GDPR?*

2.4.5 Step 4. Review a Sample Solution

Please review a sample of an **Exemplary solution** that follows the criteria specified in the **Rubric for assessing the Consent Form**.

ACTIVITY/PRACTICE QUESTION (Reflect on) We encourage you to elaborate on your response about the Exemplary solution that follows the criteria specified in the Rubric for assessing the Consent Form, in the following reflective task. You may reflect on:

1. *Do you identify any GDPR requirements that you did not take under consideration when creating your consent form?*

2.4.5.1 Exemplary Sample Solution

Consent Form to Register and Participate in the Online Course for the English Language Course of the ninth Grade of Athens Experimental High School.

In order to register and participate in the online course that will be offered for the English Language Course of the ninth Grade, you are invited to indicate your consent for the collection and processing of your personal data for the purposes of the online course, administered by Athens Experimental High School.

Athens Experimental High School (or “we”) uses a variety of resources to support student learning. Moodle™ software has been adopted as Athens Experimental High School’s Learning Management System (LMS). Moodle™ software is free and open source, and allows educators to create a private space online, filled with tools that easily create courses and various activities, all optimised for collaborative learning. In order to provide access to our students to the online course for the English Language Course of the ninth Grade on this platform/site, we need to collect and store personal information about them. You may also refer to <https://moodle.com/privacy-notice/>.

Please note:

1. The online course for the English Language Course of the ninth Grade will be carried out from 15/09/2021 to 15/06/2021.
2. Before you proceed to the registration to this online course, you will be asked to indicate your consent for the collection and processing of your personal data for the purposes of the course.
3. For the purposes of GDPR Regulation: ‘personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); ‘profiling’ means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person’s performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements; ‘controller’ means the natural or legal

person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law.

4. The Data Controller for data processed under this Notice is:
 Athens Experimental High School (VAT 021 27 76 45).
 20 Makrygianni Road.
 11,676 Athens.
 Greece.
 email: athens.expschool.online@gmail.com

Legal basis for processing the personal and sensitive data:

Personal Data:

In connection with this online course, the Athens Experimental High School's collection and processing of the following Personal Data is lawful based on.

- Article 6.1(a), GDPR, Consent.
- Article 6.1(b), GDPR, Contract.
- Article 6.1(c), GDPR, Legal Obligation.
- Article 6.1(f), GDPR, Legitimate Interest:

- Name, Surname, Email Address.
- User activity and contribution data.

Sensitive Data:

In connection with this research, the Athens Experimental High School's collection and processing of the following Sensitive Data is lawful based on consent (Article 9.2(a), GDPR):

- Gender.

Potential Benefits:

The participation in this online course enables data subjects (students) to effectively collaborate with their peers, and tutor(s) to collect data, efficiently provide resources, timely feedback and differentiated learning opportunities.

Potential Risk or Discomforts:

We do not perceive of any risk or discomfort in participating in the online course.

Storage of Data:

The installation of the Moodle™ software platform is hosted in a secure server at Athens Experimental High School's premises. The collected data is also stored in this secure server for the time required by the purposes described in this notice, for maximum 5 years.

Data transfer outside the European Union:

We may share some of the data collected with services located outside the European Union, in particular through the aforementioned Moodle™ software services.

Right to Withdraw:

Your participation in this online course is voluntary. You are under no obligation to participate in this online course and you may withdraw consent at any time,

without being at a disadvantage, by contacting the Athens Experimental High School Data Controller for this online course in athens.expschool.online@gmail.com.

Rights of Data Subject:

Whilst Athens Experimental High School is in possession of or processing your personal data, you, the data subject, have the following rights:

- Right of access – you have the right to request a copy of the information that we hold about you.
- Right of rectification – you have a right to correct data that we hold about you that is inaccurate or incomplete.
- Right to be forgotten – in certain circumstances you can ask for the data we hold about you to be erased from our records. The erasure of your information shall be subject to the Athens Experimental High School’s need to retain certain information pursuant to any other identified lawful basis.
- Right to restriction of processing – where certain conditions apply to have a right to restrict the processing.
- Right of portability – you have the right to have the data we hold about you transferred to another organisation.
- Right to object – you have the right to object to certain types of processing such as direct marketing.
- Right to object to automated processing, including profiling – you also have the right to be subject to the legal effects of automated processing or profiling.
- Right to judicial review: in the event that Athens Experimental High School refuses your request under rights of access, we will provide you with a reasonable explanation.

by contacting the Athens Experimental High School Data Controller for this online course in athens.expschool.online@gmail.com.

If the Athens Experimental High School’s use of your information is pursuant to your consent, you have the right to withdraw consent without affecting the lawfulness of the Athens Experimental High School’s use of the information prior to receipt of your request.

If you think your data protection rights have been breached you have the right to lodge a complaint with Athens Experimental High School Data Controller for this online course in athens.expschool.online@gmail.com and/or your [national Data Protection Authority \(DPA\)](#).

Data Subject Concerns and Reporting:

If you have any questions concerning the online course or experience any discomfort related to the online course, please contact the Athens Experimental High School Data Controller for this online course in athens.expschool.online@gmail.com.

Conflict of Interest

We do not perceive any conflicts of interest in the development of this online course.

Compensation:

There is no compensation for data subjects in this online course.

Confidentiality:

The only people processing your data will be the tutor(s) involved in the Athens Experimental High School's online course(s). The tutor(s) undertake to keep any information provided herein confidential, not to let it out of our possession and to report on the findings from the perspective of the entire participating group and not from the perspective of an individual. Please note that confidentiality cannot be guaranteed while data is in transit over the Internet.

Purposes for which the data is being collected and processed:

The data which is collected and processed via the online course in the Course Management System (Moodle) is being used by the Athens Experimental High School to facilitate teaching and learning. For this, online teaching resources are uploaded where the data subjects (students) enrol and study the lecture material at home. The material is in the form of videos, small activities with automatic feedback (online quizzes), and forum discussions. The data subjects (students) can undertake some additional homework online to further check their understanding and extend their learning. Though this online course and via the usage of CMS tools the tutor(s) monitor the data subjects (students) learning process, discover patterns, find indicators for success and indicators for poor marks or drop-out and proceed with recommendations and revisions of the course's online learning activities and educational resources, aiming to improve data subjects' (students') academic performance.

We ensure that the information we collect, process and use is appropriate for these correspondence purposes.

By indicating consent to participate in this online course you also indicate consent for the possible use of data for automated decision making, such as profiling, to identify data subjects' (students') progress against a range of indicators and activities identified to have an impact on data subjects' (students') success in the online course.

Consent to register and participate in the Online Course for the English Language Course of the ninth Grade of Athens Experimental High School.

Selecting "YES, I AGREE" below indicates that:

- You have read the above information;
- You voluntarily agree to participate in this online course;
- You understand the procedures described above;
- You give consent for the use of your Personal Data for the purposes outlined in this notice;
- You give consent for the use of your Sensitive Data for the purposes outlined in this notice;
- You are at least 15 years of age.

- **YES, I AGREE**
- **NO, I DO NOT AGREE**For students who are less than 15 years of age, consent from a parent or guardian is necessary
- **YES, I GIVE CONSENT FOR MY CHILD TO PARTICIPATE IN THE ONLINE COURSE AND AGREE TO THE CONSENT AS NOTED ABOVE.**
- **NO, I DO NOT GIVE CONSENT FOR MY CHILD TO PARTICIPATE IN THE ONLINE COURSE AND AGREE TO THE CONSENT AS NOTED ABOVE.**

2.4.6 Step 5. Self-Evaluate Your Answer

Now that you have seen the **Exemplary sample solution**, please rate your initial answer (evaluate the consent form you created), using the criteria in the **Rubric for assessing the Consent Form**.

Language

1. The consent request is presented neither in a clear, nor in a concise way, using language that is not easy to understand
2. The consent request is presented in a quite clear and concise way, using language that is quite easy to understand
3. The consent request is presented in a very clear and concise way, using language that is very easy to understand

Explicit and Distinguishable

1. The consent request is not explicit or distinguishable from other pieces of information.
2. The consent request is quite distinguishable from other pieces of information but is not given via a positive act.
3. The consent request is clearly distinguishable from other pieces of information, given via an electronic tick-box that the individual has to explicitly check online

Freely given consent

1. The individual does not have a free choice.
2. The individual has a free choice and it is quite clear how to refuse consent without being at a disadvantage.
3. The individual has a free choice and it is very clear how to refuse consent without being at a disadvantage.

Possibility to withdraw the given consent

1. The consent form does not include the possibility to withdraw consent
2. The consent form includes the possibility to withdraw consent, but does not explain how to do it.
3. The consent form includes the possibility to withdraw consent and explains clearly how to do it.

Rights of the data subject

1. The individuals are not informed about their rights as a data subject (GDPR Art.12 to 23)
2. Rights of the data subject (GDPR Art.12 to 23) are somehow stated but the modalities to exercise these rights are not clear.
3. Individuals are clearly informed about their rights as a data subject (GDPR Art.12 to 23) and they can effectively exercise these rights

Identity of the organisation processing data

1. The consent form does not include the identity of the organisation processing data
2. The consent form includes quite clearly the identity of the organisation processing data
3. The consent form includes very clearly the identity of the organisation processing data

Purposes for which the data is being processed

1. The consent form does not explain the purposes for which the data is being processed
2. The consent form explains quite clearly the purposes for which the data is being processed
3. The consent form explains very clearly the purposes for which the data is being processed

Describes the type of data that will be processed

1. The consent form does not describe the type of data that will be processed
2. The consent form describes the type of data that will be processed
3. The consent form describes in detail the type of data that will be processed

International transfer of data

1. The consent form does not include information about whether the consent is related to an international transfer of your data
2. The consent form includes quite clearly information about whether the consent is related to an international transfer of your data
3. The consent form includes clearly information about whether the consent is related to an international transfer of your data

References

- Barbaro, M., & Zeller, T. (2006, August 9). A face is exposed for AOL searcher No. 4417749. The New York Times. Retrieved from https://www.nytimes.com/2006/08/09/technology/09aol.html?pagewanted=all&_r=0
- Moore Barlow. (2018, June 26). GDPR and safeguarding in schools: What you need to know. Retrieved from <https://www.moorebarlow.com/gdpr-and-safeguarding-in-schools-what-you-need-to-know/>
- BrightBytes. (2017, June 29). *BrightBytes acquires trusted IPaaS provider authentica solutions*. Retrieved from <https://www.brightbytes.net/resources-archive/pressrelease>
- Broad, E., Smith, A., & Wells, P. (2017). *Helping organisations navigate ethical concerns in their data practices*. Open Data Institute. Retrieved from <https://www.scribd.com/document/358778144/ODI-Ethical-Data-Handling-2017-09-13#download>
- Castanedo, F. (2015). *Data preparation in the big data era: Best practices for data integration*. O'Reilly Media. Retrieved from https://www.tamr.com/wp-content/uploads/2015/09/Data_Preparation_in_the_Big_Data_Era_Tamr.pdf
- Cearley, D. (2017, June 23). Devise an effective cloud computing strategy by answering five key questions. Retrieved from <https://www.gartner.com>
- Chakrabarti, S., Cox, E., Frank, E., Güting, R. A., Han, J., Jiang, X., Kamber, M., Lightstone, S. S., Nadeau, T. P., Neapolitan, R. E., Pyle, D., Refaat, M., Schneider, M., Teorey, T. J., & Witten, I. H. (2009). *Data mining: Know it all*. M. Kaufmann.
- COMPARE THE CLOUD. (2018, February 21). 6 Pros and cons of cloud storage for business. Retrieved from <https://www.comparethecloud.net>
- Cybersecurity and Digital Business Risk Management. (2020, December 29). Retrieved from <https://www.gartner.com>
- D'Acquisto, G. D., Domingo-Ferrer, J., Kikiras, P., Torra, V., de Montjoye, Y. A., & Bourka, A. (2015). *Privacy by design in big data: An overview of privacy enhancing technologies in the era of big data analytics (European Union Agency for Network and Information Security (ENISA))*. ENISA. <https://doi.org/10.2824/641480>
- Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J.-H., Métayer, D. L., Tirtea, R., & Schiffner, S. (2014). *Privacy and data protection by design – from policy to engineering (European Union Agency for Network and Information Security (ENISA))*. ENISA. <https://doi.org/10.2824/38623>
- Digital Preservation Coalition. (2015). *Digital preservation handbook* (2nd ed.). Retrieved from <https://www.dpconline.org/handbook/digital-preservation/why-digital-preservation-matters>
- Drachler, H., & Greller, W. (2016). Privacy and Analytics – it's a DELICATE Issue. A Checklist for Trusted Learning Analytics. In *LAK 16 Conference Proceedings of the Sixth International Conference on Learning Analytics & Knowledge Conference*. ACM. <https://doi.org/10.1145/2883851.2883893>
- European Commission (2013). Ethics for researchers. In *Facilitating research excellence in FP7*. Publications Office of the European Union.
- European Commission. (2014). Report to the European Commission on New modes of learning and teaching in higher education. *High Level Group on the Modernisation of Higher Education. Luxembourg: Publications Office of the European Union*. <https://doi.org/10.2766/81897>
- European Commission. (2018a, January). National data protection authorities. Retrieved from https://edpb.europa.eu/about-edpb/board/members_en
- European Commission. (2018b, January 4). 2018 reform of EU data protection rules. Retrieved from https://ec.europa.eu/info/law/law-topic/data-protection/reform_en
- European Commission. (2018c, March 22). Data protection and online privacy. Retrieved from https://europa.eu/youreurope/citizens/consumers/internet-telecoms/data-protection-online-privacy/index_en.htm

- European Commission. (2018d, August 01). Can personal data about children be collected? Retrieved from https://ec.europa.eu/info/law/law-topic/data-protection/reform/rights-citizens/how-my-personal-data-protected/can-personal-data-about-children-be-collected_en
- European Commission. (2018e, August 01). For how long can data be kept and is it necessary to update it? Retrieved from https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/principles-gdpr/how-long-can-data-be-kept-and-it-necessary-update-it_en
- European Commission. (2018f, August 01). How much data can be collected? Retrieved from https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/principles-gdpr/how-much-data-can-be-collected_en
- European Commission. (2018g, August 01). Rights for citizens. Retrieved from https://ec.europa.eu/info/law/law-topic/data-protection/reform/rights-citizens_en
- European Commission. (2018h). The GDPR: new opportunities, new obligations: What every business needs to know about the EU's General Data Protection Regulation. https://ec.europa.eu/info/sites/default/files/data-protection-factsheet-sme-obligations_en.pdf
- European Commission. (2018i, August 01). What are Data Protection Authorities (DPAs)? Retrieved from https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-are-data-protection-authorities-dpas_en
- European Commission. (2018j, August 01). What are my rights? Retrieved from https://ec.europa.eu/info/law/law-topic/data-protection/reform/rights-citizens/my-rights/what-are-my-rights_en
- European Commission. (2018k, August 01). What are the responsibilities of a Data Protection Officer (DPO)? Retrieved from https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/data-protection-officers/what-are-responsibilities-data-protection-officer-dpo_en
- European Commission. (2018l, August 01). What does data protection 'by design' and 'by default' mean? Retrieved from <https://ec.europa.eu/info/law/law-topic/data-protection/>
- European Commission. (2018m, August 01). What is personal data? Retrieved from https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en
- European Commission. (2018n, August 01). What personal data is considered sensitive? Retrieved from https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_en
- European Commission. (2018o, August 01). When is a Data Protection Impact Assessment (DPIA) required? Retrieved from https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/when-data-protection-impact-assessment-dpia-required_en
- European Commission. (2018p, August 01). When is consent valid? Retrieved from https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/grounds-processing/when-consent-valid_en
- European Union. (2018). *It's your data – take control*. Publications Office of the European Union.
- GDPR.EU. (2019, February, 13). Does the GDPR apply to companies outside of the EU? Retrieved from <https://gdpr.eu/companies-outside-of-europe/>
- Griffiths, D., Drachler, H., Kickmeier-Rust, M., Steiner, C., Hoel, T., & Greller, W. (2016, January 6). Is privacy a show-stopper for learning analytics? A review of current issues and their solutions. *Learning Analytics Review*, ISSN: 2057–7494. LACE project.
- Hansell, S. (2006, August 8). AOL removes search data on group of web users. *The New York Times*. Retrieved from <https://www.nytimes.com/2006/08/08/business/media/08aol.html>
- Harish, A. (2019, March 21). When NASA lost a spacecraft due to a metric math mistake. [blog post]. Retrieved from <https://www.simscale.com/blog/2017/12/nasa-mars-climate-orbiter-metric/>
- Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation*, 3(1), 134–140. <https://doi.org/10.2218/ijdc.v3i1.48>

- Hill, K. (2012, February 16). How target figured out a teen girl was pregnant before her father did. *Forbes*. Retrieved from <https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/?sh=df77ff666686>
- Hill, A. (2014, September 15). A day in the life of a data mined kid [Web log post]. Retrieved from <https://www.marketplace.org>
- ICPSR. (2018). Data management & curation. *Institute for Social Research at the University of Michigan*. Retrieved from <https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/index.html>
- iKeepSafe. (2017). Data privacy in education: An iKeepSafe educator training course. Retrieved from <https://ikeepsafe.org/wp-content/uploads/2018/02/Data-Privacy-in-Education-Full-Curriculum-2017-2.pdf>
- Intersoft Consulting. (2018, October 5). General Data Protection Regulation (GDPR): Chapter 3 Rights of the data subject. Retrieved from <https://gdpr-info.eu/chapter-3/>
- Interuniversity Consortium for Political and Social Research (ICPSR). (2009). *Principles and good practice for preserving data* (International Household Survey Network, IHSN Working Paper No 003). Retrieved from <http://www.ihsn.org/principles-and-good-practice-for-preserving-data>
- IT Governance UK. (2016, October 3). Data protection impact assessments under the GDPR. Retrieved from <https://www.itgovernance.co.uk/privacy-impact-assessment-pia>
- Jisc. (2006). Digital preservation: Continued access to authentic digital assets. Retrieved from <https://www.webarchive.org.uk/wayback/archive/20140613220103/http://www.jisc.ac.uk/media/documents/publications/digitalpreservationbp.pdf>
- Johnston, L. R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R., & Stewart, C. (2018). How Important is Data Curation? Gaps and Opportunities for Academic Libraries. *Journal of Librarianship and Scholarly Communication*, 6(1), eP2198. <https://doi.org/10.7710/2162-3309.2198>
- Kononow, P. (2018, September 16). What is metadata (with examples). *Dataedo*. Retrieved from <https://dataedo.com/kb/data-glossary/what-is-metadata>
- Krueger, M. (2017, October 13). 3 Steps to identify and protect sensitive data for the GDPR [Web log post]. Retrieved from <https://www.pacificdataintegrators.com/insights/3-Steps-to-Identify-and-Protect-Sensitive-Data-for-GDPR>
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). John Wiley & Sons, Inc.. <https://doi.org/10.1002/9781119013563>
- Loon, R. (2018, February 6). Digital transformation requires a data-driven culture. Retrieved from <https://www.thedigitaltransformationpeople.com/channels/people-and-change/digital-transformation-requires-data-driven-culture/>
- Maletic, J. I., & Marcus, A. (2000). Data cleansing: beyond integrity analysis. *Fifth Conference on Information Quality (IQ 2000)*, 200–209.
- Marr, B. (2017, November 15). Outside insight: Why external data is the fuel of tomorrow's business success. Retrieved from <https://www.forbes.com/sites/bernardmarr/2017/11/15/outside-insight-why-external-data-is-the-fuel-of-tomorrows-business-success/?sh=21c35a0a5e1d>
- Narayanan, A., & Shmatikov, V. (2008). *Robust deanonymization of large datasets (how to break anonymity of the netflix prize dataset)*. The University of Texas at Austin.
- Open Data Institute. (2018a, March 24). No one owns data: we need to strengthen our rights. Retrieved from <https://theodi.org/article/no-one-owns-data-we-need-to-strengthen-our-rights/>
- Open Data Institute. (2018b, April 25). Updating the data ethics Canvas. Retrieved from <https://theodi.org/article/updates-the-data-ethics-canvas/>
- Panetta, K. (2019, October 10). Is the cloud secure? Retrieved from <https://www.gartner.com/>
- Pentland, A. S. (2013). The data-driven society. *Scientific American*, 309(4), 78–83.
- Phillips, M., Bailey, J., Goethals, A., & Owens, T. (2013). The NDSA levels of digital preservation: An explanation and uses. *National Digital Stewardship Alliance (NDSA)*. Retrieved from https://nds.a.org/documents/NDSA_Levels_Archiving_2013.pdf
- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection

- Regulation). Official Journal L. 119. Retrieved from <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32016R0679>
- Riley, J. (2017). *Understanding metadata: What is metadata, and what is it for?* Baltimore.
- Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. J. D. (2010). *Handbook of Educational Data Mining*. CRC Press.
- Romero, C., Romero, J. R., & Ventura, S. (2014). A survey on pre-processing educational data. In *Educational data mining studies in computational intelligence* (pp. 29–64). Springer. https://doi.org/10.1007/978-3-319-02738-8_2
- Slater, N. (2017). Consent and the GDPR: what approaches are universities taking?. *Effective Learning Analytics*. Retrieved from <https://analytics.jiscinvolve.org/wp/2017/06/30/consent-and-the-gdpr-what-approaches-are-universities-taking/>
- Slater, N., & Bailey, P. (2015). Code of practice for learning analytics. Jisc. Retrieved from <https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics>
- Shacklock, X. (2016). *From bricks to clicks: The potential of data and analytics in higher education*. Policy Connect – Higher Education Commission.
- Slade, S., & Tait, A. (2019). *Global guidelines: Ethics in Learning Analytics*. International Council for Open and Distance Education.
- Stonebraker, M. (2014, October 16). Three approaches to scalable data curation. *Strata Hadoop World in New York 2014*. Retrieved from <https://www.tamr.com/blog/three-approaches-scalable-data-curation-stonebraker-stratahadoop/>
- Sweeney, L. (2000). *Simple Demographics Often Identify People Uniquely (Data Privacy Working Paper 3)*. Carnegie Mellon University.
- UCL. (2018, November 15). GDPR – Anonymisation & Pseudonymisation. Retrieved from <https://www.ucl.ac.uk/data-protection/guidance-staff-students-and-researchers/practical-data-protection-guidance-notices/anonymisation-and>
- UK, Information Commissioner’s Office (ICO). (2012, November). Anonymisation: Managing data protection risk code of practice. Retrieved from <https://ico.org.uk/media/1061/anonymisation-code.pdf>
- USGS. (2014). Guidelines for the preservation of digital scientific data. Retrieved from https://nds.a.org/documents/USGS_Guidelines_for_the_Preservation_of_Digital_Scientific_Data_Final.pdf
- SaiGayatri Vadali. (2017, December 28). Day 7: Data cleaning – All you need to know about it. Retrieved from <https://becominghuman.ai/day-7-data-cleaning-all-that-you-need-to-know-about-it-23b05738abe7>
- Venezia, P. (2014, June 23). Murder in the Amazon cloud. Retrieved from <https://www.info-world.com/>
- Wes, M. (2018). Looking to comply with GDPR? Here’s a primer on anonymization and pseudonymization. Retrieved from <https://iapp.org/news/a/looking-to-comply-with-gdpr-heres-a-primer-on-anonymization-and-pseudonymization/>
- Woody, C. (2004). Risk methodology K-12. Retrieved from <https://studylib.net/doc/8097924/k-12-risk-methodology>

Useful Video Resources

- External Video: Data wrangling for faster, more accurate analysis [1:47].
- External Video: Meta... what? metadata! [5:25].
- External Video: Learn more about data interoperability [1:12].
- External Video: ICPSR 101: What is data curation? [1:29].
- External Video: Data curation @UCSB [2:29].
- External Video: Why digital preservation is important for everyone [2:51].

- External Video: Public cloud vs private cloud vs hybrid cloud [3:28].
- External Video: How toy story 2 almost got deleted: Stories from pixar animation: ENTV [2:25].
- External Video: Introduction to data ethics [3:23].
- External Video: What is the GDPR? [1:11].
- External Video: Why develop a data science code of ethics? [3:06].
- External Video: Privacy overview for K12 teachers and Administrators [5:26].
- External Video: Who uses student data? [2:30].
- External Video: Protecting student-data privacy: An expert's View [3:44].

Further Readings

- Acaps. (2016). Technical brief data cleaning. Retrieved from Acaps: https://www.acaps.org/sites/acaps/files/resources/files/acaps_technical_brief_data_cleaning_april_2016_0.pdf
- Beagrie, N., Charlesworth, A., & Miller, P. (2015). *The national archives guidance on cloud storage and digital preservation* (2nd ed.). Crown copyright licensed under the Open Government Licence v2.0.
- Chapman, A. D. (2005a). *Principles and methods of data cleaning – primary species and species occurrence data, version 1.0*. Report for the Global Biodiversity Information Facility, Copenhagen.
- Chapman, A. D. (2005b). *Principles of data quality*, version 1.0. Report for the global biodiversity information facility, Copenhagen.
- Common Education Data Standards (CEDS). (2017). Retrieved from <https://ceds.ed.gov/dataModel.aspx?CrowdFlower>.
- CrowdFlower. (2016). *Data Science Report* (Rep.). Retrieved from Crowd Flower website https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf
- Consortium for School Networking (CoSN). (2017). Cybersecurity. Retrieved from <https://www.cosn.org/cybersecurity>
- Data Cleaning: In-Depth Guide [Web log post]. (2019, January 1). Retrieved from <https://blog.aimultiple.com/data-cleaning/>
- Data Quality Campaign. (2013). Roadmap to safeguarding student data key focus areas for state education agencies. Retrieved from <https://2pido73em67o3eytaq1cp8au-wpengine.netdna-ssl.com/wp-content/uploads/2016/03/DQC-roadmap-safeguarding-data-June24.pdf>
- Data Quality Campaign. (2018a). Education data legislation review: 2018 state activity. Retrieved from <https://2pido73em67o3eytaq1cp8au-wpengine.netdna-ssl.com/wp-content/uploads/2018/09/2018-DQC-Legislative-Summary.pdf>
- Data Quality Campaign. (2018b). TIME TO ACT 2018: Using data to meet ESSA Goals. Retrieved from <https://2pido73em67o3eytaq1cp8au-wpengine.netdna-ssl.com/wp-content/uploads/2018/12/DQC-Time-to-Act-2018.pdf>
- Data Science Primer. (2017, June 9). Chapter 3: Data cleaning steps and techniques. Retrieved from <https://elitedatascience.com/data-cleaning>
- de Jonge, E., & van der Loo, M. (2013). *An introduction to data cleaning with R*. Statistics Netherlands.
- DQC. (2015). A stoplight for student data use. Retrieved from <https://2pido73em67o3eytaq1cp8au-wpengine.netdna-ssl.com/wp-content/uploads/2016/03/DQC-FERPA-Stoplight.pdf>
- Elgabry, O. (2019, February 28). The ultimate guide to data cleaning. When the data is spewing garbage. Retrieved from <https://towardsdatascience.com/the-ultimate-guide-to-data-cleaning-3969843991d4>
- Freitas, A., & Curry, E. (2015). Big data curation. In J. Cavanillas, E. Curry, & W. Wahlster (Eds.), *New horizons for a data-driven economy: A roadmap for usage and exploitation of big data in Europe* (pp. 87–118). Springer. https://doi.org/10.1007/978-3-319-21569-3_6

- Gartner. (2018). Plan your GDPR Journey. Retrieved from https://blogs.gartner.com/smarter-withgartner/files/2018/05/PR_435169_Beadle_Are_You_Ready_for_GDPR_Infographic_final.png
- GDPR Report. (2017, November 07). Data masking: Anonymisation or pseudonymisation? Retrieved from <https://gdpr.report/news/2017/11/07/data-masking-anonymisation-pseudonymisation/>
- Hollidge, R. (2018, February 21). Where are you storing your data and is your storage method GDPR compliant?. Retrieved from <https://www.instantonit.com>
- Hswe, P., & Musser, L. (n.d.). 2.1 What is metadata? Penn State University Libraries. Retrieved from <https://www.e-education.psu.edu/dmpt/node/660>
- Johnson, L. R., et al. (2018). How important are data curation activities to researchers? Gaps and opportunities for academic libraries. *Journal of Librarianship and Scholarly Communication*, 6(General Issue), eP2198. <https://doi.org/10.7710/2162-3309.2198>
- Ladd, M. (2018, April 5). What is metadata and how does it impact education? [Blog post].
- Larkin, A. (2018, June 26). Disadvantages of cloud computing. Retrieved from <https://cloud-academy.com>
- Mason, J. (2004). Context and metadata for learning, education, and training. In R. McGreal (Ed.), *Online education using learning objects* (pp. 168–182). Routledge Farmer.
- Miller, K., Miller, M., Moran, M., Dai, B. (2018). Data management life cycle: Final report. In *PRC 17–84F*. A&M Transportation Institute.
- National Forum on Education Statistics. (2009). *Forum guide to metadata: The meaning behind education data (NFES 2009–805)*. U.S. Department of Education. National Center for Education Statistics.
- OpenRefine. (2010, July 11). Introduction to OpenRefine. Retrieved from <http://openrefine.org/>
- Osborne, J. W. (2010). Data cleaning basics: Best practices in dealing with extreme scores. *Newborn and Infant Nursing Reviews*, 10(1), 37–43. <https://doi.org/10.1053/j.nainr.2009.12.009>
- Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research, and Evaluation*, 9(6).
- Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, 74(4), 525–556. <https://doi.org/10.3102/00346543074004525>
- Richter, T., & Pawlowski, J. (2007). The need for standardization of context metadata for e-learning environments. *e-ASEM Conference*. Seoul.
- Shacklett, M. (2016, August 12). Data curation takes the value of big data to a new level. Retrieved from <https://www.techrepublic.com/article/data-curation-takes-the-value-of-big-data-to-a-new-level/>
- Student Data Principles. (2014). Retrieved from <https://studentdataprinciples.org/>
- UK, Information Commissioner’s Office (ICO). (2017, October 20). Overview of the General Data Protection Regulation (GDPR). Retrieved from <https://ico.org.uk/media/for-organisations/data-protection-reform/overview-of-the-gdpr-1-13.pdf>
- UK, Information Commissioner’s Office (ICO). (2019a, May 17). Guide to the General Data Protection Regulation (GDPR). Retrieved from <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/>
- UK, Information Commissioner’s Office (ICO). (2019b, April 30). Individual rights. Retrieved <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/>
- Verbert, K., Manouselis, N., Drachsler, H., & Duval, E. (2012). Dataset-driven research to support learning and knowledge analytics. [Electronic version]. *Educational Technology & Society*, 15(3), 133–148.
- Webber, M. (2018). The GDPR’s impact on the cloud service provider as a processor. *Privacy & Data Protection Journal*, 16(4).
- Whitelegg, D. (2018, May 25). A developer’s guide to the GDPR. Understand how the GDPR impacts you. Retrieved from <https://developer.ibm.com/articles/s-gdpr1/>
- Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, 14(10). <https://doi.org/10.18637/jss.v059.i10>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 3

Learning Analytics



3.1 Introduction and Scope

3.1.1 Scope

The goals on this chapter are to:

- introduce the basics of **methods** and **tools** for **analyzing** and **interpreting** online learners' data to facilitate their personalized support,
- focus on **organizing**, **analyzing**, **presenting** and **interpreting learner-generated data** within their learning context, and
- elaborate on **ethical concerns** and **policies** for protecting learner-generated data from mistreatment and misuse.

3.1.2 Chapter Learning Objectives

This chapter learning objectives	Learn2Analyse Educational data Literacy Competence profile
Know what the common measurements of learner data and their contexts are, and understand the processes needed to collect both learner and context data in online and/or blended learning settings	1.1
Be able to identify and describe the limitations and quality measures on collecting learners' data in online and/or blended learning settings	1.2
Know methods for learners' data analysis and modelling as part of learning analytics methods	3.1
Know and understand learner-generated data presentation methods	3.2

	Learn2Analyse Educational data Literacy Competence profile
This chapter learning objectives	
Know and understand learners' data properties in learning analytics	4.1
Be able to identify and discriminate statistics commonly used for the interpretation of educational data in learning analytics	4.2
Be able to elaborate on the insights from learners' data analysis	4.3
Know and understand the methods that can be used to protect individuals' data privacy, confidentiality, integrity and security in learning analytics	6.2

3.1.3 Introduction

At the heart of this chapter is the so-called **Learning Analytics**. Learning analytics has been a hot topic for a while in educational communities, organizations and institutions. There are four essential elements involved in all learning analytics processes: **data**, **analysis**, **report** and **action** (Fig. 3.1).

1. Data, as the **primary analytics asset**, are the **raw material** that **gets transformed** into analytical insights; in the educational domain, they include information that is (usually) gathered as the learning processes are taking place, and is **about the learners, the learning environment, the learning interactions, and the learning outcomes**. A complete view of educational data has been provided in Chap. 1.
2. Analysis is the process of **transforming the collected data to obtain actionable information** from them, using, for this purpose, a set of **mathematical and statistical algorithms** and techniques; during data analysis, the data are **cleansed, transformed** and **modelled** with the goal of **discovering meaningful information** and supporting decision-making and action.
3. Report is used to **summarize what the analysis of the collected data can tell about learning** and to **present this information** in a meaningful manner; it is a set of processes for **organizing and presenting the results** of the analysis of learners' and learning data into charts and tables. Reporting learners' and learning data will provide **insights** about the learners' states during learning; interpreting those insights can guide data-driven decision making to action taken.

Fig. 3.1 The basic elements of learning analytics



4. Action is the ultimate goal of any learning analytics process; it is the **set of the informed decisions and the practical interventions** that the **educational stakeholders will undertake**. The results of **follow-up actions** will determine the success or failure of the analytical efforts. Learning analytics is useful **only if there is “action”** as a result of its implementation.

The increased need to **inform decisions and take actions based on data**, points out the significance of understanding and **adopting learning analytics in everyday educational practice**. And in order to treat educational data in a respectful and protected manner, the **policies for learning analytics** play a major role and need to be explicitly clarified.

3.2 Using Learner-Generated Data and Learning Context for Extracting Learning Analytics

3.2.1 Definition and Objectives of Learning Analytics

Learning analytics is defined by SOLAR as “*the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs*” (SOLAR, 2011). In other words, it is an **ecosystem of methods and techniques** (in general procedures) that successively **gather, process, report and act** on machine-readable data **on an ongoing basis** in order to **improve the learning environments and experience**.

As described in the [Learning Analytics](#) video (in the useful video resources), like any other context-aware process, learning analytics procedures **track and record data** about learners and their contexts, **organize and monitor** them, and **interpret and map** the real current state of those data, **to use them** for providing “**actionable intelligence**”, i.e., insights to act upon.

Based on the shared common understanding of learning analytics, it is important to clarify and discuss what learning analytics can do, what they can be used for, why one needs to use learning analytics, or in other words, what are the **objectives** of learning analytics? Some simple examples from everyday experience can showcase those objectives.

- In *traditional classroom settings*, it’s often hard to identify each student’s individual strengths and weaknesses, learning disabilities and prior subject knowledge, and subsequently **tailor and personalize instruction** accordingly. It’s also hard to **recommend** personalized learning resources to the individuals.
- In *online learning settings*, it’s common that the students **drop-out** early. It’s also hard to detect students’ **emotions** or enhance students’ **social learning skills**.

- In *blended learning settings*, the students might not know how to **self-regulate their learning**, and they often **procrastinate**. It's also hard to **monitor each student's progress** and provide **feedback** accordingly.

These deficiencies are identified immediately with learning analytics. More specifically, learning analytics aim to (Chatti et al., 2012; Papamitsiou & Economides, 2014) and those objectives are illustrated in Fig. 3.2:

- **Monitor** learners' progress
- **Model** learners/learners' behaviour
- **Detect** affects/emotions of learners
- **Predict** learning performance/dropout/retention
- **Generate** feedback
- **Provide** recommendations
- **Guide** adaptation
- **Increase** self-reflection/ self-awareness
- **Facilitate** self-regulation

Overall, **learning analytics are important** because *every "trace"* within an electronic learning environment may be valuable information that can be **tracked, analyzed and combined with external learner data**; every simple or more complex action within such environments can be **isolated, identified and classified** through computational methods **into meaningful patterns**; every type of interaction can be **coded into behavioural schemes** and **decoded into interpretable guidance** for decision making.

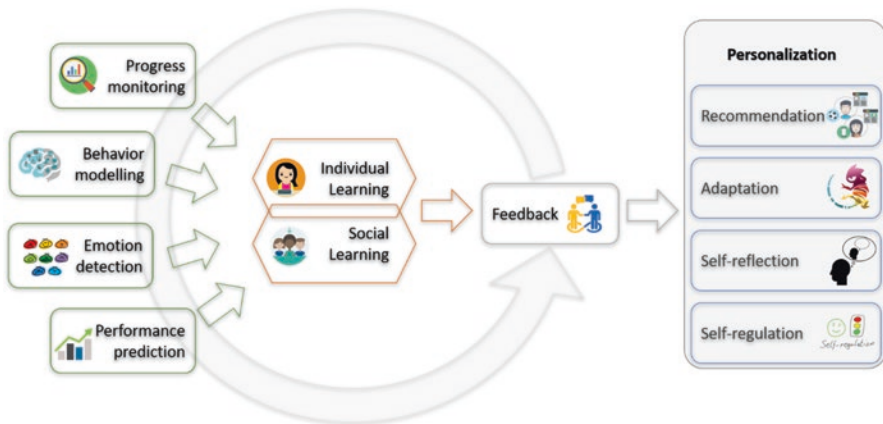


Fig. 3.2 The objectives of learning analytics

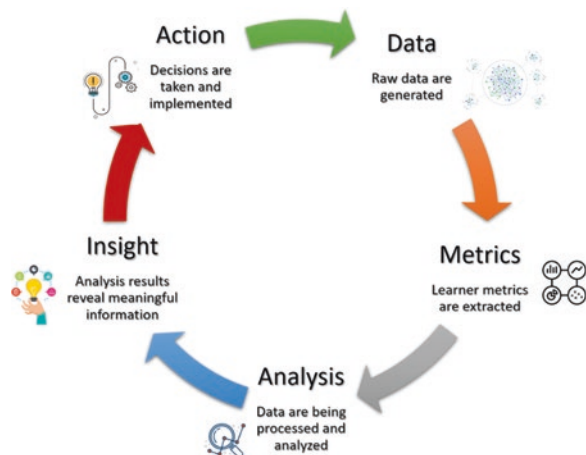
3.2.2 Measurements as Indicators of Learners' Current Learning States

Learning analytics seeks to produce “**actionable intelligence**”; the key is the action that is taken. Campbell and Oblinger (2007) have pointed out **five steps** in learning analytics: **Capture, Report, Predict, Act, Refine**. From (a) capturing and gathering the raw data, to (b) introducing metrics for sharing a common understanding of the data in educationally meaningful ways, to (c) analyzing the metrics for predicting the future states of the learners, to (d) gaining insights into the learning processes, and to (e) acting upon the data-based evidence for delivering personalized learning to each individual, the **cyclical process of learning analytics is fed with the continuously generated learner data**, illustrated in Fig. 3.3.

Learning analytics are about learners and their learning. As such, Clow (2012) proposed a cycle for learning analytics that starts with **learners**. The next step is the generation and capture of **data** about or by the learners. The third step is the **processing of this data into metrics** or analytics, which provide some insight into the learning process. The cycle is not complete until these metrics are used to drive one or more **interventions** (actions) that have some effect on learners.

This learning analytics cycle can provide a **data-perspective to strong learning theories**. For instance, the cycle can be viewed as a data-driven aspect of Kolb's Experiential Learning Cycle (1984): taking the system as a whole, there is a direct correspondence: actions by or about learners (*concrete experience*) generate data (*observation*) from which metrics are derived (*abstract conceptualization*), which are used to guide an intervention (*active experimentation*). The role of the learner is fundamental in this process. And, since learning analytics are extracted from the learners' and learning data, two steps need to be clarified: a) **what is the learner's data that will be used in learning analytics**, and b) **what types of learning analytics can be formed from the learner's data**.

Fig. 3.3 The cycle of learning analytics



As already explained, learning analytics is a cyclical process. Learners generate data that can be processed into metrics and analyzed for patterns such as success, weakness, overall personal or comparable performance, and learning habits. Educators can administer “interventions” based on the data analyzed, and the process then repeats itself.

Before beginning to analyze data, one should **understand what data are collected, and why it is needed to collect them**: data collection should have **specific objectives and outcomes**. The collected data on their own cannot give meaningful insights, unless they are associated with specific measurements, **depending on what one wants to measure**: learning outcomes, goal attainment, performance, behavioural changes, engagement, motivation, cognition, abilities, emotions, etc. **Metrics are what one measures, the measurements**.

There are many types of data that support student learning – and they are so much more than test scores. The type of information the educational data often include, and the sources the data can be collected from, usually are linked with a straightforward relation. For instance, **student characteristic data and/or contextual information** are usually collected from enrolment records, student profiles, or attendance rolls; **student perception data** can be found in surveys and interviews; **student activity data** are available in logs from the LMS and interaction records; **student achievement data** lay within various kinds of assessment data such as rubrics, scores or observation notes; **student wellbeing data** capture students’ social and emotional development, or school climate, and can be found in sources such as biosignals or social networks. Educational data and the respective data sources are explained in Chap. 1.

But **individual data points don’t give the full picture** needed to support the incredibly important education goals of parents, students, educators, and policy-makers. The [What Is Student Data?](#) Video (see useful video resources) explains in simple terms what student data is about and when they can be used effectively. As explained in this video and in Chap. 1, there are learner and context data that can be captured **within the learning environment** (e.g., log-files, quiz scores, login data, content access, file downloads, discussion participation, etc.), and there are also other types of data that are **external to the learning environment** (e.g., survey-demographic data, biosensor data, online discussion forums, social network data, etc.). In addition, **aggregating/integrating different data sources to increase validity and relevance, and to reduce biases (improve reliability)** is also important. Once one understands what data need to be collected, one will be able to **locate and select the most appropriate data sources to extract them from**. Those data will **feed the learning analytics cycle**.

It has been explained in previous sections what student data are about and how they can be combined together to show the whole picture of student learning, which is deeply related to the context itself. Learning analytics is a context-aware process. **Both learner and context data are necessary** in this process. Different types of **data can come together** – under different objectives – **to form a full picture of student learning**. When used effectively, data empowers everyone. The first step is to **understand why one is collecting data and associate the data with metrics**

according to the learning concept one **aims to measure and shed light on**. Each of **these measurements** referred to as **learning analytics metrics, can be associated with one or more learning analytics objectives** (see Sect. 3.2.1), summarized in Fig. 3.4.

To understand that, let’s take the following simple and generic example. How many views make an educational YouTube video a success? How about 300 K? That’s how many views a video you posted got. It featured some well-known and successful professionals, who prompted young people to enrol in a Data Science course. It was twice as popular as any video you had posted to date. Success! Then came the data report: only eight viewers had signed up to take the course, and zero actually completed it. Zero completions. From 300 K views. Suddenly, it was clear that views did not equal success. In terms of completion rates, the video was a complete failure. **What happened?**

Well, not all important things in life can be measured and not everything that can be measured is important. **If one is measuring something, but not necessarily all the right things**, the end result could still not be right, or one is relying on the wrong data to make the case. The critical question is *which measurements are the “right” ones*. There is a difference between numbers and numbers that matter. This is **what separates data from metrics**. One can’t control the educational data one is collecting, but can control what one measures. When we talk about learning analytics metrics and measurements, we’re typically referring to gathering data on three areas: *efficiency, effectiveness, and outcome* (Robbins, 2017), illustrated in Fig. 3.5.

- **Efficiency** is generally thought of as *learning-centric* activity metrics—number of learners, time on task, frequencies of resources downloads, quiz scores, attempts, hint usage, etc.
- **Effectiveness** metrics are *evaluation-focused* and include aspects like learner engagement, quality of deliverables, knowledge acquisition, collaboration, progress, performance, etc.
- **Outcome** looks at *bottom-line results*. To the extent that efficiency and effectiveness metrics matter, they provide validation and explanation for the outcome.

Learning efficiency refers to more granular metrics, closer to raw data; their objective is to describe learners’ actions at the task or activity level (**micro-level**), and they cannot sufficiently reveal a lot about learning (as a more general objective) on

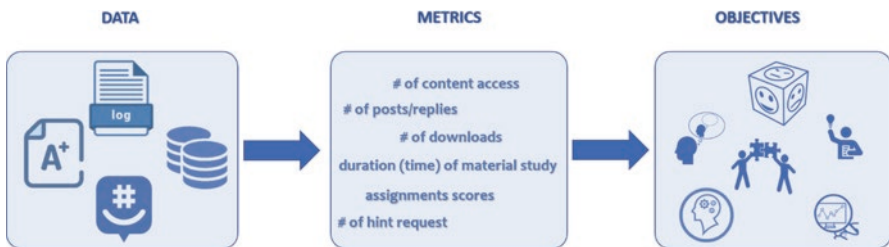


Fig. 3.4 Associating data to learning analytics metrics and objectives

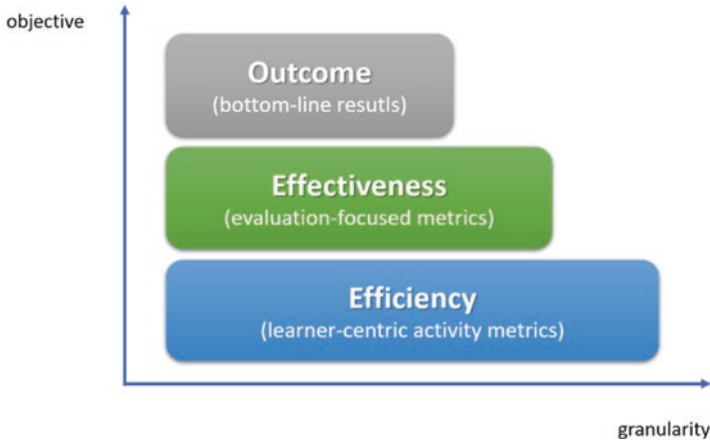


Fig. 3.5 Categories of learning analytics metrics

their own. Combining these metrics can contribute to understanding more complex learning constructs, such as engagement and collaboration. The metrics used to refer to this **meso-level** (activity or course) of more abstract and complex concepts are synopsis under the *learning effectiveness* metrics, and their objective is to quantify less fine-grained constructs. Finally, *learning outcome* can be described with metrics from previous categories that are combined to give insight and explain the results of the learning processes (**macro-level**).

Depending on the goals (i.e., the learning analytics objective), the learning analytics metrics will be **obtained from the same or different learner and context data**. The types/levels of the metrics will be decided according to their sophistication, the complexity of the analysis method employed, and the value they add for human decision-making (Lang et al., 2017; Scapin, 2015; Soltanpoor & Sellis, 2016):

- **Descriptive analytics:** use data aggregation and data mining to provide insight into the past and answer: “What has happened?” (e.g., reports and descriptions).
- **Diagnostic analytics:** dissect the data with methods like data discovery, data mining and correlations to answer the question “Why did it happen?” (e.g., interactive visualizations).
- **Predictive analytics:** utilize a variety of data to make the prediction and apply sophisticated analysis techniques (such as machine learning) to answer the question “What is likely to happen?” (e.g., trends and predictions).
- **Prescriptive analytics:** utilize an understanding of what has happened, why it has happened and a variety of “what-might-happen” analysis to help the user determine the best action to take and answer the question “What do I need to do?” (e.g., alerts, notifications, recommendations).

Figure 3.6 illustrates the types of learning analytics based on their complexity and value for decision-making.



Fig. 3.6 Types of learning analytics based on their complexity and value for decision-making

Questions and Teaching Materials

1. What are the three core questions to ask before using learning analytics?

- What information to use? What has happened? What is likely to happen?
- What information to use? How is it gathered? What is likely to happen?
- What information to use? How is it gathered? How is it combined?
- What information to use? How is it gathered? What has happened?

Correct answer: c.

2. What can Learning Analytics do, and what can they be used for?

- Monitor progress, predict performance, create content, facilitate self-regulation
- Predict dropout, increase self-awareness, guide adaptation, detect emotions
- Generate feedback, model learners, support game-based learning, predict retention
- Evaluate learning, provide recommendations, assess collaboration, increase effort

Correct answer: b.

3. **How can Learning Analytics (LA) provide a data-driven perspective to strong learning theories:**
- (a) **LA helps teachers develop more appropriate interventions and learning opportunities for target learners (e.g., experiential learning)**
 - (b) **LA helps learners become aware of their progress on different tasks by combining the learning data that are generated during the process (e.g., self-regulated learning)**
 - (c) **LA make use of data generated by learners' online activity to identify behaviours and patterns within the learning environment that signify effective process (e.g., social learning).**
 - (d) **All the above**

Correct answer: d.

4. **Which of the following are learning data?**
- (a) **Survey-demographic data, biosensor data**
 - (b) **Gender, socioeconomic status, special education needs**
 - (c) **Test scores, educational file downloads, educational content access**
 - (d) **Enrolment records, emotional development, social network data**

Correct answer: c.

5. **What is the difference between data and metrics?**
- (a) **Data are measurements (numbers/calculations) to help make decisions about how to move forward, whilst metrics are indicators of progress and achievement**
 - (b) **Data is the set of raw numbers or calculations gathered, whilst metrics are proxies for what ultimately matters (i.e., what we measure)**
 - (c) **Data is a mapping of observations into numbers, whilst metrics are numerical approximations of objectives**
 - (d) **Data is the raw measurements, whilst metrics are trends in the data**

Correct answer: b.

6. **Consider the following metrics: time on task, frequencies of resources downloads, quiz scores, attempts, hint usage. What category of metrics are they?**

- (a) **Learning efficiency (micro-level)**
- (b) **Learning effectiveness (meso-level)**
- (c) **Learning outcome (macro-level)**
- (d) **Those are not metrics – they are raw data**

Correct answer: a.

7. **What type of learning analytics would you use to help you determine that all of the student’s actions—low interaction time, low forum participation, and low scores—point to low engagement in the activity?**
- (a) **Descriptive analytics**
 - (b) **Diagnostic analytics**
 - (c) **Predictive analytics**
 - (d) **Prescriptive analytics**

Correct answer: b.

8. **ACTIVITY/PRACTICE QUESTION (Reflect on)**

We encourage you to elaborate on your response on data collection in the following reflective task.

You may reflect on:

1. **Can you associate the educational data with the learning analytics objectives? Please provide specific examples of how educational data can be used to address specific learning analytics objectives.**
2. **Can you explain how the same or different educational data can be used as different types of learning analytics? Please provide specific examples of data used as learning analytics metrics for descriptive, diagnostic, predictive and prescriptive analytics**

3.2.3 *Limitations and Data Quality Issues of Learners’ Data Measurements in Open and Blended Courses*

As already explained in Chap. 1, data often suffer from inaccuracies, biases or even manipulations; the educational data, apart from being **relevant** to be used for decision making (fit-for-purpose), should also be **reliable** and **valid**. According to [Wikipedia](#) (Data Quality, 2022), **data is generally considered high quality** if it is “fit for [its] intended uses in operations, decision making and planning and data is



Fig. 3.7 The dimensions of data quality

deemed of high quality if correctly represents the real-world construct to which it refers.”

Like in all kinds of organizations, data quality is critical for educational institutes, as well. In online and blended learning settings, many factors are additive to the existing difficulty in **handling educational data quality**. For example, such factors often are heterogeneous educational data sources, high volumes of learner and learning data, and a myriad of unstructured data types extracted. The [Data Quality Matters – Tech Vision 2018 Trend](#) video (see useful video resources) explains the critical issues of data quality from a more general perspective. As discussed in this video, there are many aspects to data quality, including **completeness, consistency, accuracy, timeliness, validity, and uniqueness**, synopsized as follows (Mihăiloaie, 2015; Pipino et al., 2002) and illustrated in Fig. 3.7:

- **Completeness:** there are no gaps in the data from what was *expected to be collected* and what was *actually collected*, i.e., there are no missing data – the collected dataset is complete.
- **Consistency:** the data types must align and be compatible with the expected versions of the data being collected, i.e., there are *no contradictions in the data types* and the data are *usable*.
- **Accuracy:** collected data are correct, relevant and *accurately represent what they should*.
- **Timeliness:** the data should be *received at the expected time* for the information to be utilized efficiently.

- **Validity:** a measurement is *well-founded* and likely corresponds accurately to the real world.
- **Uniqueness:** there should be *no data duplicates* reported.

Among the 6 dimensions, completeness and validity usually are easy to assess, followed by timeliness and uniqueness. Accuracy and consistency are the most difficult to assess. The critical question is how those data limitations relate to learning analytics and why does quality matter. Here, we will focus on how these principles/limitations apply in learning analytics.

Specifically, in the learning analytics cycle, learner and contextual data are collected and transformed into metrics (analytics), according to the learning objective that needs to be addressed; the different types of metrics shall next guide human decision-making and interventions. Yet, *the higher the need for data-driven decision-making is, the more the integrity and quality of data become critical* (National Forum for the Enhancement of Teaching and Learning in Higher Education, 2017). The following example demonstrates in simple terms the impact of data limitations and quality for learning analytics.

Let's examine the case of an educator who wants to understand learners' engagement with an activity. To measure engagement on the activity level, it is common practice to use learners' participation data (e.g., frequency of logins, session duration, posts on the activity forum, etc.). If the learners' ID is missing from the data that are available via the LMS (the data are *incomplete*), then the educator shall not be able to identify each learner's participation. Similarly, if each learner's data would be stored in different formats (e.g., dates: MM/DD/YY vs. DD/MM/YY) this would result in confusion about the validity of the data and their interpretation (the data are *not valid*). In the same example, this *inconsistency* in the data format would also result to *inaccurate* data— *when did the learner really log in to the activity?* – i.e., it would be unclear *what the correct values of the stored data are*. Furthermore, if the learners' data during the activity would not become timely available, the educator would not gain insight to what the learners are doing during that activity (violation of *timeliness*), making it impossible to intervene in a timely manner. Similarly, if the same learners' data are stored multiple times (e.g., each time a learner logs in the activity, the login is duplicated) and all the information is considered for analysis, the results would be misleading (violation of *uniqueness*).

It is important to clarify that raw data quality strongly affects the analytics quality; learning analytics metrics are **transformations of** the raw learner and learning **data** collected, according to the objectives set. These metrics will next be **treated as “data”** themselves, and they will be subjected to further processing. Just like with any kind of data, **quality also matters for learning analytics metrics**: what the specific metrics **can reveal** is strongly dependent on their quality. In most cases, limited quality will have the direct result of **lack of trust** in the metrics, and consequently, **poor decisions** and **gradual abandonment** of the data-driven educational decision-support system. **Poor quality data is troublesome** (The data quality benchmark report, 2015). Educators **cannot and will not trust insights** that are acquired by processing corrupted, duplicate, inconsistent, missing, broken, or incomplete

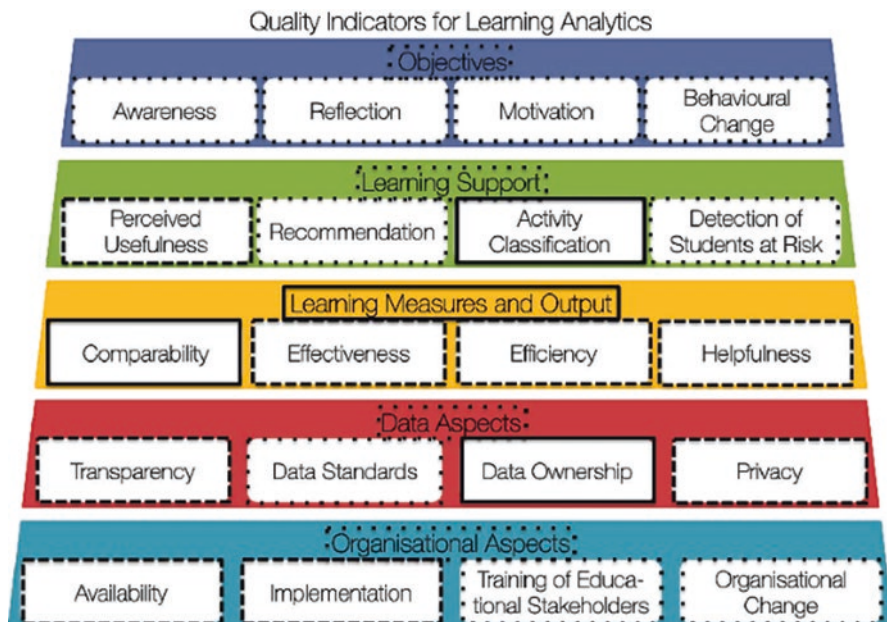


Fig. 3.8 Quality indicators for learning analytics. (Adopted from: Scheffel et al., 2015)

data. Learning analytics metrics quality is expected to **increase the value of the learner and learning data** and the opportunities to use them properly.

The following approaches were developed to discuss the exact concerns of **quality issues in learning analytics metrics**. In particular, the LACE project developed a proposal for a framework of quality indicators for learning analytics that contributes towards a standardized and holistic approach for the evaluation of learning analytics tools (Scheffel et al., 2015). It potentially can act as a means for providing evidence on the impact of learning analytics on educational practices. The suggested framework is *generic* and considers **multiple learning analytics aspects**, ranging from their objectives to organizational issues. For the measures and data aspects, the framework highlights *comparability, effectiveness, efficiency, and helpfulness, as well as transparency, data standards, data ownership, and privacy, respectively* (Fig. 3.8).

From a *more “data-oriented” approach to “quality”* aspects for *learning analytics metrics*, the above indicators can be combined and merged with those identified before (illustrated in Fig. 3.7), as follows:

- **Learning analytics metrics quality indicators:** Standards (comparability, consistency), Completeness, Accuracy (effectiveness, efficiency), Validity, Timeliness, Uniqueness.
- **Learning analytics metrics ethics considerations:** Privacy, Ownership, Transparency, Consent.

Fig. 3.9 Ethical considerations of learning analytics on top of quality indicators



The “*quality indicators*” refer to how appropriate the learning analytics metrics are, how fit-for-purpose they are as data that will be used in the decision-making process in turn; the “condition” of the data themselves – the degree to which a set of characteristics of data fulfils requirements.

The “*ethics considerations*” refer to systemising, defending, and recommending concepts of right and wrong conduct in relation to data; they are considerations that tackle the potential for data misuse, and issues about the right, legitimate, and proper ways to use data. Ethics considerations are placed on top of quality indicators, since the latter are relevant to the data, whilst the former are relevant to the usage of the data (Fig. 3.9).

Like any kind of data, learning analytics metrics should be protected from misuse, mistreatment, or violations. The **quality of learning analytics** as (data) metrics themselves matters in terms of impacting the quality of the outcome as a data-driven decision. **Mostly it is important to control** who has access to those metrics, what can and cannot be done with the metrics, and for how long access is granted after the collection and analysis of the raw learning and context data occurs. Therefore, along with the learning analytics metrics quality indicators, the ethical limitations should be considered, as well.

Questions and Teaching Materials

1. Data completeness refers to:

- (a) **Data that are well-founded and likely correspond accurately to the real world**
- (b) **Correct and relevant data that accurately represent what it should.**

- (c) **There are no gaps in the data from what was expected to be collected and what was actually collected**
- (d) **The data types must align and be compatible with the expected versions of the data being collected**

Correct answer: c

2. Assume an LMS's database is a huge file, which has an important index located 20% of the way through and saves content data at the 75% mark. Consider a scenario where an e-tutor comes and creates new content (e.g., adds new exercises) at the same time a backup is being performed, which is being made as a simple "file copy" which copies from the beginning to the end of the large file(s) – and at the time of the content edit, it is 50% complete. The new content is added to the content space (at the 75% mark) and a corresponding index entry is added (at the 20% mark). What is the data quality problem that raises in this scenario?
- (a) **Data consistency**
 - (b) **Data completeness**
 - (c) **Data accuracy**
 - (d) **Data timeliness**

Correct answer: a.

3. The goal of the quality indicators framework is:
- (a) **To evaluate how appropriate the learning analytics metrics are, how fit-for-purpose they are as data that will be used in the decision-making process in turn**
 - (b) **To contribute towards a standardized and holistic approach for the evaluation of learning analytics tools.**
 - (c) **To systemise, defend, and recommend concepts of right and wrong to tackle the potential for data misuse, and issues about the right, legitimate, and proper ways to use data**
 - (d) **To control who has access to the metrics, what can and cannot be done with the metrics, and for how long access is granted after the collection and analysis of the raw learning and context data occurs.**

Correct answer: b.

4. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to elaborate on your response about learning analytics metrics limitations and quality, in the following reflective task:

1. **Can you provide examples of how the limitations of analytics quality apply when addressing learning objectives using specific learning analytics metrics? You can use the example in this section as guidance.**
2. **Do you understand the difference between limitations as quality measures for learning analytics and the ethical limitations for learning analytics? Please, provide specific examples of each category of data quality**

3.2.4 *Ethical Treatment of Learner-Generated Data and Measurements*

Learning analytics provides tremendous opportunities to assist learners – but they also pose ethical implications that shouldn't be ignored. The practical challenge of learning analytics metrics is the question of privacy of the learner and how to protect the learner from potential harm due to data misuse. Questions abound:

- Who has **access** to the learner's data? Who **owns** individuals' data?
- To what degree do you need to **inform** users that their data are being collected?
- Do you need learners' **permission** to use their data?
- Where should the data be **stored**? How **secure** does it need to be?
- Is **identification** of individuals possible from metadata?
- What about **misinterpretation** of data, or other data errors?

Towards addressing these issues, the [Learning Analytics: The need for a code of ethics](#) video (see useful video resources) elaborates on the need to establish a code of ethics for learning analytics. This code of practice aims to set out the responsibilities of educational institutions to ensure that learning analytics is carried out responsibly, appropriately and effectively, addressing the key legal, ethical and logistical issues which are likely to arise.

Slade and Prinsloo (2013) identified three broad classes of ethical issues: (a) the location and interpretation of data; (b) informed consent, privacy, and the de-identification of data; and (c) the management, classification, and storage of data. As we have explicitly explained, in the learning analytics cycle, data are collected about individuals and their learning activities, and metrics are constructed; the data will be analysed and interventions (might) take place. This entails opportunities for positive impacts on learning, as well as risks for misunderstandings, misuse of data and adverse impacts on students.

When learners perform learning tasks within a learning environment to increase their knowledge and develop skills and competences, they expect to receive support to overcome gaps in knowledge/competences. They also expect **to be in a "safe"**

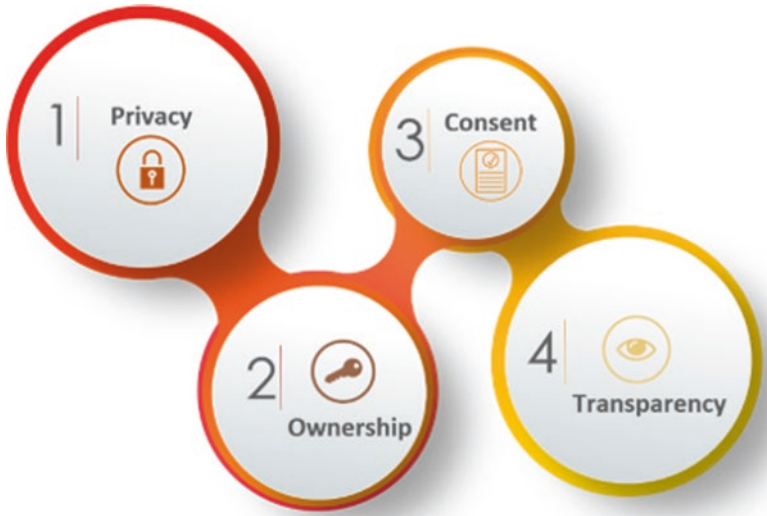


Fig. 3.10 Ethical considerations in learning analytics

environment where their mistakes will be treated with respect, without serious consequences or unfair and unjustified discrimination against them, as individuals. Two critical issues are hidden in the implied “*safety*” of the learning environments: (a) the learners should feel “**secure**” and maintain the “**privacy**” of their data (integrity of the self), and (b) the learners’ data should be treated in an “**ethical**” manner. Drachler and Greller (2016) provided a clear differentiation between ethics and privacy: “*Ethics is the philosophy of morality that involves systematizing, defending, and recommending concepts of right and wrong conduct [...] privacy is a living concept made out of continuous personal negotiations with the surrounding ethical environment*”. The main ethics considerations are illustrated in Fig. 3.10 and are outlined as follows:

- **Privacy:** the regulation of how personal digital information is being observed by the self or distributed to other observers – protection from unauthorized intrusion. Anonymize and de-identify individuals.
- **Ownership:** the act of having legal rights and complete control over a single piece or set of data – information about the rightful owner of data assets and the acquisition, use and distribution policy implemented by the data owner.
- **Consent:** documentation that clearly describes the processes involved in data collection and analysis. Explain how the data will be used, and why – and how it won’t be used – and get consent from each individual before any data are collected.
- **Transparency:** the regulation about the purposes for which data will be collected and used, under which conditions, who will have access to data, the measures through which individuals’ identity will be protected, and how sensitive data will be handled.

Ethics provides us with guides on what is the right thing to do in all aspects of life, while the law generally provides more specific rules so that societies and their institutions can be maintained (Tsachuridou, 2015).

Over the past 5 years or so, a number of guidelines, codes of practice and policies have been developed in response to this. Slade and Prinsloo (2013) established one of the earliest frameworks with a focus on ethics in learning analytics. Others have followed, including JISC's code of practice in 2015, the **Learning Analytics Community Exchange (LACE)** framework in 2016 (Drachler & Greller, 2016) and a learning analytics policy development framework for the EU by the **SHEILA project** (Tsai & Gasevic, 2017). More recently and in the light of the rapid development of Learning Analytics on a global basis, **International Council for Open and Distant Education (ICDE)** has taken the initiative to produce a set of guidelines for ethically-informed practice that would be valuable to all regions of the world (March 2019).

To address the issues raised earlier in this section and **demystify the ethics and privacy limitations** around learning analytics, the LACE project published the **DELICATE instrument** to be used by any educational institution. The instrument includes **policies and guidelines** regarding privacy, legal protection rights or other ethical implications that address learning analytics. The DELICATE checklist helps to investigate the obstacles that could impede the rollout of learning analytics and the implementation of trusted learning analytics for higher education. The eight points are shown in Fig. 3.11 and include:

1. **D-etermination:** Decide on the purpose of learning analytics for your institution.
2. **E-xplain:** Define the scope of data collection and usage.
3. **L-egitimate:** Explain how you operate within the legal frameworks, refer to essential legislation.
4. **I-nvolve:** Talk to stakeholders and give assurances about the data distribution and use.
5. **C-onsent:** Seek consent through clear consent questions.
6. **A-nonymise:** De-identify individuals as much as possible.
7. **T-echnical aspects:** Monitor who has access to data, especially in areas with high staff turn-over.
8. **E-xternal partners:** Make sure externals provide highest data security standards.

The EU SHEILA project focused on developing a learning analytics policy development framework for the EU under the 6 dimensions of the Rapid Outcome Mapping Approach (ROMA) (Ferguson et al., 2014; Macfadyen et al., 2014), and consisting of 49 **action points**, 69 **challenges**, and 63 **policy questions**. The ROMA dimensions, as considered by the SHEILA framework, include: (1) The political context of an institution, i.e., identifying the 'purposes' for adopting learning analytics in a specific context; (2) The involvement of stakeholders, i.e., the implementation of learning analytics in a social environment involves collective efforts; (3) A vision of behavioural change and potential impacts; (4) Strategic planning, including resources, ethics & privacy, and stakeholder engagement and buy-in; (5) Institutional



Fig. 3.11 The delicate checklist

capacity to affect change, i.e., assessing the availability of existing resources; (6) A framework to monitor and evaluate the efficacy and continue learning.

In addition, the ICDE report on Ethics in Learning Analytics identified several core issues that are important on a global basis for the use and development of Learning Analytics in ethics-informed ways. Those issues are shown in Fig. 3.12 and include:

- **Transparency**: how learners' data are collected, analysed and used to shape learners' paths.
- **Data ownership and control**: the presumption is often that data collected are owned by the institution. However, "*data are not considered as something a student owns but rather is. Students do not own their data but are constituted by their data*" (Prinsloo & Slade, 2017). Therefore, institutions do not own the student data that they hold but have temporary *stewardship*.
- **Accessibility of data**: can relate to both the determination of who has access to raw and analysed data, and to the ability of students to access and correct their own data. Within a learning analytics context, we might expect that data are



Fig. 3.12 Ethics in learning analytics based on ICDE report

accessed on a ‘need-to-know’ basis to facilitate the provision of academic and other support services.

- **Validity and reliability of data:** Datasets should be kept valid, reliable, accurate, representative of the issue being measured, current, complete and sufficient.
- **Institutional responsibility and obligation to act:** how access to knowing and understanding more about how students learn brings with it a moral obligation to act.
- **Communications:** care should be taken when communicating directly with students on the basis of their analytics.
- **Cultural values:** measures established as being correlated with successful or unsuccessful outcomes are likely to differ in different geographies and cultures.
- **Inclusion:** Learning Analytics should be primarily used to support students, in student-centred ways that minimize the risk to legitimise exclusion.
- **Consent:** In line with GDPR, consent is not required for the use of non-sensitive data for analytics, is required for use of sensitive data, and would be required to take interventions directly with students on the basis of the analytics.
- **Student agency and responsibility:** it is recommended that institutions seek to engage students in applications of learning analytics so as students can be actively involved in helping the institution to design and shape interventions that will support them.

Questions and Teaching Materials

1. Carrie is an instructional designer and a book writer. She creates content for online courses and she also prepares a printed version of her book. From the student data available on the online platform that she uses for the courses, she can identify students who are in need of additional learning support, and she decides to promote and sell her book to those students and make profit from it. What are the ethical/legal issues raised here with regard to student data?
 - (a) The location and secure storage of data
 - (b) Misinterpretation of data, or other data errors
 - (c) Informed consent, privacy, and ownership of student data
 - (d) The regulation about the purposes for which data will be collected and used

Correct answer: d

2. What is the difference between “data ownership” and “data privacy”?
- Data privacy requires us to, at least conceptually, agree that you as the data subject own your data and the data you generate. Data ownership in itself does not necessitate that privacy be respected by default.
 - Data privacy is the regulation of how personal data will be collected and used, under which conditions, and who will have access to data. Data ownership is the act of having legal rights and complete control over a single piece or set of data.
 - Data privacy is the right of a citizen to have control over how personal information is collected and used. Data ownership is the regulation of how personal data will be collected and used, under which conditions, and who will have access to data.
 - Data privacy is the act of having legal rights and complete control over a single piece or set of data. Data ownership defines and provides information about the rightful owner of data assets and the acquisition, use and distribution policy implemented by the data owner.

Correct answer: a.

3. Match the appropriate definition (from the right column), to the respective “point” in the left column

1. External partners	a. Define the scope of data collection and usage.
2. Determination	b. Seek consent through clear consent questions.
3. Anonymise	c. Make sure externals provide highest data security standards.
4. Technical aspects	d. Explain how you operate within the legal frameworks, refer to essential legislation.
5. Explain	e. Talk to stakeholders and give assurances about the data distribution and use.
6. Legitimate	f. De-identify individuals as much as possible.
7. Consent	g. Decide on the purpose of learning analytics for your institution.
8. Involve	h. Monitor who has access to data, especially in areas with high staff turn-over.

Correct answer: 1.c / 2.g / 3.f / 4.h / 5.a / 6.d / 7.b / 8.e.

4. **Read the paper Tsai et al. (2018) and focus on sect. 4 Results. What are the identified challenges for stakeholders in the three case studies?**
- (a) **It was difficult to define ownership and responsibilities among professional groups within the university.**
 - (b) **The provision of opt-out options conflicts with the goal to tackle institutional challenges that involve all institutional members.**
 - (c) **Anonymised data could potentially be reidentified when matched with other pieces of data.**
 - (d) **All the above**

Correct answer: d.

5. **Which of the following statements is correct?**
- (a) **The SHEILA framework is used to inform the development of policies for learning analytics, but strategies are not covered**
 - (b) **The DELICATE checklist addresses issues of power-relationship, data ownership, anonymity, data security, privacy, data identity, transparency, and trust.**
 - (c) **The ICDE report on Ethics in Learning Analytics identifies which core principles relating to ethics are core to all, unless there is legitimate differentiation due to separate legal or more broadly cultural environments.**
 - (d) **All the above**

Correct answer: b.

6. **ACTIVITY/PRACTICE QUESTION (Reflect on)**

We encourage you to elaborate on your response about learning analytics ethical considerations and policies, in the following reflective task:

1. **Read the SHEILA-research-report and choose 2 action points, 2 challenges and 2 policy questions that you find most interesting. Please, elaborate on your choices.**
2. **Study the DELICATE framework and the ICDE framework and discuss the overlap between them.**

3.3 Analyzing Data and Presenting Learning Analytics

3.3.1 *Methods for Analyzing the Learner-Generated Data and the Measurements Over Them*

As already explained, the learning analytics cycle describes the whole process from collecting the learner and context data to taking data-driven actions and interventions. The raw learner and context data do not tell a lot on their own, but when converted to metrics, they have the potential to reveal what we don't know about our learners.

Good metrics have three key attributes: their data are consistent, clean, and valid to use (see Sect. 3.2.3). Data cleaning and management is a demanding task (see Chap. 2). Given that good and clean data are available, next the data analysis method needs to be selected. Here we explain **what methods** can be used for **analysing** the educational data and learning analytics. This step is the main “game” of Data Science; it requires the procedures under the umbrella of Data Science. Data Science is a blend of various tools, algorithms, and machine learning principles with the goal to discover hidden patterns from the raw data (Sharma, 2019). The main generic categories of methods of this step are shown in Fig. 3.13 and include (but are not limited to):

- **Statistical methods**
- **Data mining**
- **Machine learning**
- **Qualitative methods**
- **Social Network Analysis**
- **Visualization** – This step is related to the output procurement and will be extensively presented in the next section.

However, not all data analysis methods can yield the results one is seeking. To achieve that, **a number of criteria need to be specified**, e.g., the learning analytics objective you want to address (modelling learners, prediction of performance, adaptation, recommendation, etc., see Sect. 3.2.1), the metrics you have to compute

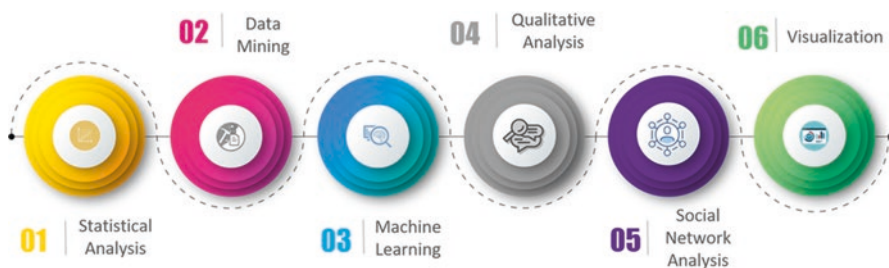


Fig. 3.13 The basic data analysis methods in learning analytics

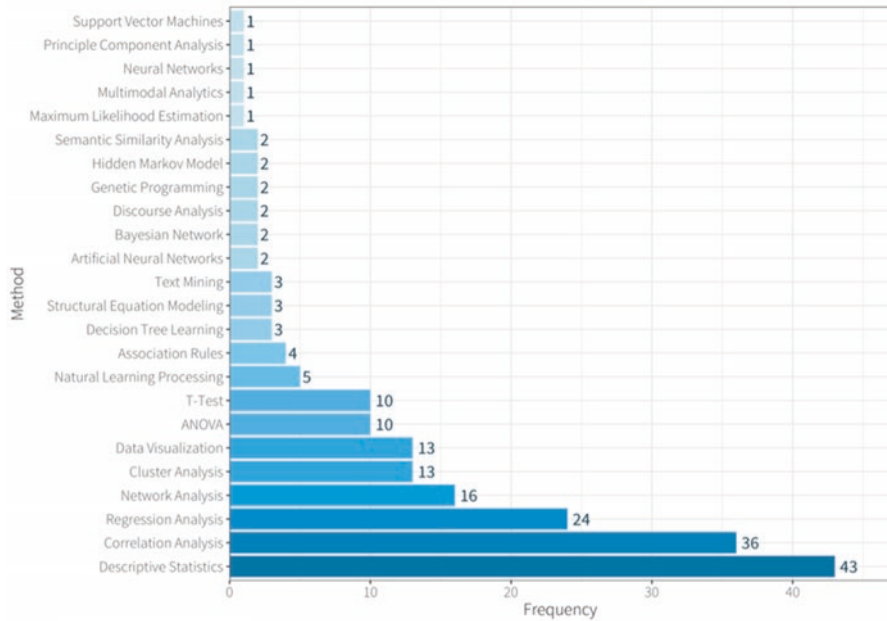


Fig. 3.14 Frequency of data analysis methods in learning analytics. (Data Source: <http://bora.uib.no/handle/1956/17740>)

(effective, efficient, outcome, see Sect. 3.2.2), and the type of analytics you want to use (descriptive, diagnostic, predictive, etc., see Sect. 3.2.2). The analysis methods will be utilized to form a better understanding of the educational settings and learners: learning analytics focus on the **application of known methods and models** to address issues affecting student learning and the environments in which it occurs.

Before we explain how the appropriate analysis method can be chosen according to the needs, we briefly introduce (in simple terms) the approaches commonly used in learning analytics. Specifically, the learning analytics metrics come from data related to learners' interactions with course content, other learners, and instructors. Different techniques are applied to **detect interesting patterns hidden** in the educational data sets.

Among the analysis techniques, some have received increased attention in the last couple of years, namely *statistics*, *data mining*, *machine learning*, *qualitative analysis*, *social network analysis*, and *visualizations* (Chatti et al., 2012; Khalil & Ebner, 2016; Papamitsiou & Economides, 2014). In a recent report on the current state-of-the-art in learning analytics, a corpus of 100 studies was considered (Misiejuk & Wasson, 2017). Figure 3.14 shows the frequency of the data analysis methods used in the corpus.

By far, *statistics* is the most commonly used method, including *descriptive statistics* (43%), *correlation analysis* (36%), *ANOVA* (10%) and *T-Test* (10%). Data mining methods like *regression analysis* (24%) and *cluster analysis* (13%) are also

common techniques, followed by *network analysis* (16%) and *data visualisations* (13%). The remainder of the methods were reported 1–5 times. Some of these less used approaches are *machine learning methods* such as neural networks and support vector machines. More recently, *multimodal analysis* uses more sophisticated data such as video, gaze, gestures, and combines various methods such as computer vision, machine learning, etc.

Although the different analysis methods are **inherently technical**, they can **provide pedagogical insights** if properly used. For example, *descriptive statistics* (such as the mean, median and standard deviation) can be used to showcase the students' interaction with a learning system (the usage), as it is coded with *efficiency metrics* (see Sect. 3.2.2) like the time online, total number of visits, distribution of visits over time, frequency of students' postings/replies, percentage of material read, etc. Statistical methods can also be used to signify the importance of the analysis results (e.g., analysis of variance – ANOVA, and t-tests), or to explain more complex constructs of learning (*effectiveness metrics*), such as engagement (e.g., Principal Component Analysis – PCA). *Data mining* methods like classification and clustering can be used to model and explain learner performance (*outcome metric*), and *machine learning* techniques can be successfully applied to detect learners' affective states (*effectiveness metrics*) during the learning activities.

Next, we focus on how the most commonly used **statistical methods** can **tell the story** in the data. In particular, **statistics** are used for measuring, controlling, communicating and understanding the data (Davidian & Louis, 2012). It is a mathematical science including methods of collecting, organizing and analyzing data in such a way that meaningful conclusions can be drawn from them. In general, statistics begin with data collection using a **sampling method** (you have learned about that in Chap. 1), and next, for understanding the collected data, its investigations and analyses fall into two broad categories called **descriptive** and **inferential statistics**. Furthermore, **descriptive statistics** deals with the processing of data **without attempting to draw any inferences** from it (Kenton, 2018). Finally, **inferential statistics** is a scientific discipline that uses mathematical tools to **make forecasts** and make generalizations about the larger population of subjects by analyzing the given data (Kuhar, 2010).

The **Statistics – Introduction to Statistics** video (see useful video resources) presents a brief introduction to statistics. Before advancing to more sophisticated techniques, we elaborate more on the fundamentals of statistical analysis and how they can tell the story in learning data analytics.

As already explained, descriptive statistics are used to summarize data in a way that makes sense. Descriptive statistics are, as their name suggests, descriptive: they illustrate what the data shows but do not generalize beyond the data considered. Here is a list of commonly used descriptive statistics (Dillard, 2017):

- **Frequencies** – a count of the number of times a particular score or value is found in the data set. For example, how many students (within all participants) have scored 5 out of 10 on a test.
- **Percentages** – used to express a set of scores or values as a percentage of the whole.

- **Mean** – numerical average of the scores or values for a particular variable, e.g., the average score that the students achieved on a test. Taken alone, the mean is a dangerous tool. In some data sets, the mean is also closely related to the mode and the median (two other measurements near the average). However, in a data set with a high number of outliers or a skewed distribution, the mean simply doesn't provide the accuracy you need for a nuanced decision.
- **Median** – the numerical midpoint of the scores or values that is at the center of the distribution of the scores.
- **Mode** – the most common score or value for a particular variable, e.g., the most common score that was achieved among all students.
- **Minimum and maximum values (range)** – the highest and lowest values or scores for any variable.
- **Standard deviation (σ)** – quantifies the amount of variation or dispersion of a set of data values, or otherwise, how close the data points are to the mean – the measure of a spread of data around the mean. A low standard deviation indicates that the data points tend to be close to the mean of the set, while a high standard deviation indicates that the data points are spread out over a wider range of values.

Mean, median and mode are measures of central tendency, while range and standard deviation are measures of dispersion.

Descriptive statistics may be sufficient if the results do not need to be generalized to a larger population, e.g., outside the specific assignment; when comparing the percentage of students that have solved an assignment correctly versus wrongly, descriptive statistics may be sufficient. Most analytics fall into the basic data evaluation category, and there is tremendous value here, and opportunities for some huge wins.

However, using only this kind of statistics entails the risk of '*picking the low hanging fruit*' of learning analytics – descriptive information or **simple statistics** that *values what can be easily measured rather than measuring what values*. If it matters to understand, not only what happened, but also why it happened, utilizing the data to make inferences or predictions about learners is needed, and using inferential statistics is required.

Inferential statistics can be used to generalize the findings from sample data to a broader population, and examine the differences and relationships between two or more samples of the population (Kuhar, 2010). These are more complex analyses and are looking for significant differences between variables and the sample groups of the population. Inferential statistics allow testing hypotheses and generalizing results to the population as a whole. Following is a list of basic inferential statistical tests (Rathi, 2018):

- **Correlation** – seeks to describe the nature of a relationship between two variables, such as strong, negative positive, weak, or statistically significant. If a correlation is found, it indicates a relationship or pattern, but keep in mind that *it does not indicate or imply causation*.
- **Analysis of Variance (ANOVA)** – tries to determine whether or not the difference in the means of two sampled groups is statistically significant or due to random chance. For example, the test scores of two groups of students are

examined and proven to be significantly different. The ANOVA will tell you if the difference is significant, but it does not speculate regarding “why”.

- **Regression** – used to determine whether one variable is a predictor of another variable. For example, a regression analysis may indicate to you whether or not participating in a test preparation program results in higher ACT scores for high school students. It is important to note that regression analysis are like correlations in that *causation cannot be inferred* from it.

Questions and Teaching Materials

1. Which of the following statements best explains the generic role of Data Science in Learning Analytics?
 - (a) Data Science is used to convert the raw student data into learning analytics metrics
 - (b) Data Science uses mathematical tools to make forecasts about the larger student population by analyzing their data
 - (c) Data Science is a blend of various tools, algorithms, and machine learning principles with the goal to discover hidden patterns from the raw student data and to understand learning in online and blended learning environments
 - (d) Data Science uses complex analyses and is looking for significant differences between variables for the sample groups of the student population

Correct answer: c.

2. What data analysis method would you use to signify the differences in on-task effort exertion (e.g., in time-spent to complete the task) between different student groups?
 - (a) Median and standard deviation
 - (b) t-tests and/or ANOVA
 - (c) Principal Component Analysis
 - (d) Machine Learning

Correct answer: b

3. Which are the generic categories of statistical methods?
 - (a) Simple statistics, Complex statistics, Inferential statistics
 - (b) Sampling methods, Simple statistics, Complex statistics
 - (c) Sampling methods, Descriptive statistics, Inferential statistics
 - (d) Descriptive statistics, Complex statistics, Inferential statistics

Correct answer: c.

4. Match the appropriate definition (from the right column), to the respective “descriptive statistic” in the left column

1. Mode	a. numerical average of the scores or values for a particular variable, e.g., the average score that the students achieved on a test.
2. Median	b. a count of the number of times a particular score or value is found in the data set.
3. Mean	c. Quantifies the amount of variation or dispersion of a set of data values, or otherwise, how close the data points are to the mean – the measure of a spread of data around the mean.
4. Standard deviation	d. the most common score or value for a particular variable, e.g., the most common score that was achieved among all students.
5. Frequencies	e. Used to express a set of scores or values as a percentage of the whole.
6. Percentages	f. the numerical midpoint of the scores or values that is at the center of the distribution of the scores.

Correct answer: 1.d / 2.f / 3.a / 4.c / 5.b / 6.e.

5. Calculate the frequency of the students (within all participants) who have scored above 5 (>5) out of 10 in all assignments, from the table below.

- (a) **3**
- (b) **7**
- (c) **8**
- (d) **12**

	Assign.1	Assign.2	Assign.3	Mid-term Test	Assign.4	Assign.5	Final Test
Stud1	5	7	4	4	6	7	5
Stud2	7	9	7	9	8	8	8
Stud3	5	5	7	5	6	5	5
Stud4	4	5	3	4	4	3	2
Stud5	7	8	6	7	4	5	4
Stud6	5	6	5	3	4	5	3
Stud7	7	7	7	6	7	5	5
Stud8	7	7	3	7	6	7	8
Stud9	3	7	6	5	3	3	2
Stud10	6	8	7	4	6	7	6
Stud11	6	9	5	4	4	7	5
Stud12	6	6	6	5	6	5	5

(continued)

	Assign.1	Assign.2	Assign.3	Mid-term Test	Assign.4	Assign.5	Final Test
Stud13	7	8	7	6	6	4	4
Stud14	7	7	4	6	5	2	4
Stud15	5	6	8	7	7	4	5
Stud16	6	7	8	5	5	4	4
Stud17	4	7	4	6	4	5	3
Stud18	7	6	9	5	3	3	3
Stud19	5	5	7	8	8	7	6
Stud20	7	9	8	9	10	8	9
Stud21	6	7	4	3	4	2	3
Stud22	6	5	4	2	4	5	4
Stud23	4	4	4	4	5	5	6
Stud24	7	6	7	6	5	3	6
Stud25	6	5	5	7	5	4	5

Correct answer: a.

6. Mariana is an instructional designer, and she needs to redesign the educational material for a course, on which the majority of students failed. She has available student data from the previous time the course was available. What statistical method should she use to predict students' scores from their participation variables (e.g., time on assignments, number of assignments completed, frequencies of logins, etc.)?
- Mean and standard deviation of the participation variables: they illustrate what the data shows
 - ANOVA of the participation variables: determine whether or not the difference in the means of the sampled groups is statistically significant or due to random chance
 - Regression: determine whether the participation variables can explain the scores
 - None of the above: more advanced data analysis methods are required

Correct answer: c.

7. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to elaborate on your response about the analysis methods employed in learning analytics, in the following reflection task:

- Provide 2 examples of learning analytics metrics and explain why you would use the mean and standard deviation to describe their values. Please, elaborate on your choices.

2. **Provide examples of learning analytics metrics that could be used to explain a learning outcome, and elaborate on the statistical method you would use to explore the relationship.**

3.3.2 *Presentation Methods for Reporting on Learner Data Analytics*

Now the educational data that were collected have been analyzed. *How did students perform in an assignment? How did they perform compared to the previous assignment? How many of them downloaded the material that was made available online? How much time did the students spent on studying the online material compared to the score they achieved on the assignments?*

These are common questions that can be answered when the educational data that have been collected, are analyzed using the respective metrics. The collected learner and context data and learning can be presented in many different ways to help make it **easier to understand** and **more interesting to read**. After collecting and organizing data, the next step is to **display them in an easy to read manner** – highlighting similarities, disparities, trends, and other relationships, or the lack of, in the dataset.

Data can be used to make data-driven and informed educational decisions, but all the data in the world **won't help if one cannot understand what the insightful analysis can present**. The first step to presenting data is to understand that **how data is presented matters** (Kiss, 2018). Take these two visuals. They display the results of the scores that 250 students achieved on the five assignments and the mid-term exams during one semester, on a scale 0–100. The first one (infographic style – Fig. 3.15) is “prettier.” However, the visual is difficult to understand unless one actually reads the information on it. Pretty, but not helpful...

On the other hand, the second one (Fig. 3.16) uses simple bars to display the same information. Helpful, and still pretty...

In this section we elaborate on the different ways used to represent educational data and learning analytics metrics in a meaningful manner.

As already explained, displaying the analysis results and what is within the educational dataset in a clear way, is helpful in telling the story and making sense of the data that have been collected. Data reports present the data, analyses, conclusions and recommendations in an easy to decipher and digest format (Lebied, 2016).

The methods commonly used to display data include **tables, charts, bar graphs, pie graphs, and line plots**. Other commonly used ways to present data are **histograms, box-plots, scatterplots, and stem-and-leaf plots**. Sometimes, a combination of the graphical representations is used as a **dashboard**: presenting data results together should tell a story or reveal insights together, that isn't possible if left apart.

Why do we use tables, diagrams or charts to display the learner/learning information?

Fig. 3.15 Infographic style visualization of learning data

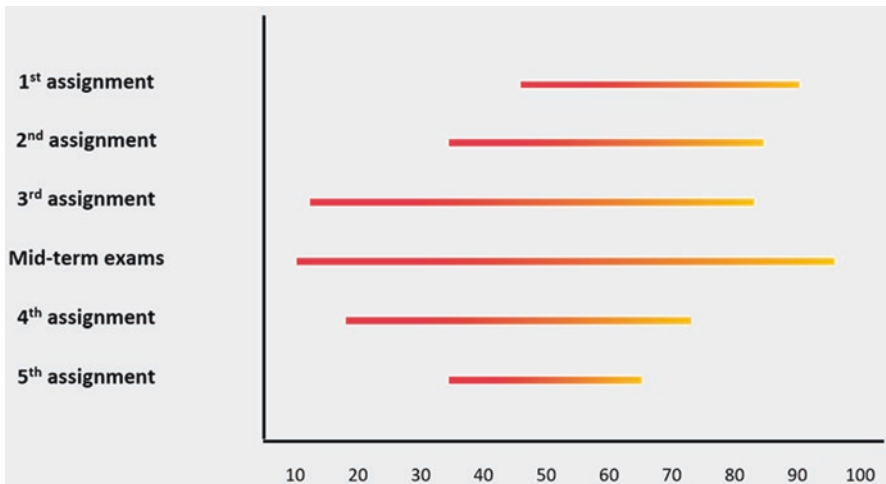
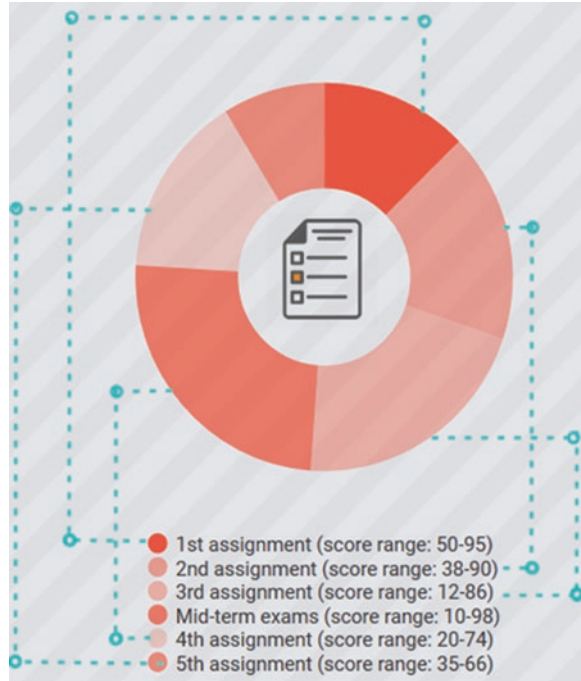


Fig. 3.16 Simple visualization of learning data

- Displaying data visually (with pictures) can make it **easier to understand**.
- It makes the information **stand out on a page**.
- It is easier to **display using pictures, rather than lots of words**. For example, it is easier to show someone the layout of a town using a map, rather than describing it in words.

Data can be presented in various forms **depending on the type of data collected**. For example, a **frequency distribution table** shows **how often** each value (or set of values) of the variable occurs in a dataset. A **frequency table** is used to summarize categorical or numerical data. Frequencies are also presented as **relative frequencies**, that is, the **percentage** of the total number in the sample. Except from the tables, there are other, **graphical ways** to present data. Analytics presented visually make it easier for decision makers to grasp difficult concepts or identify new patterns.

The **Value of Data Visualization** video (see useful video resources) provides a quick introduction to the value of data visualization. Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, **data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data**. Data visualization is a powerful tool, especially in a world desperate for hard facts. When it comes to making sense of learning analytics and understanding learning patterns in the educational data, one can start from simple graphs that can demonstrate this information. For example, quiz submission data, discussion interaction data (e.g., participation in the forum), data from the access to the learning management system, assignment completion data have been gathered and analyzed. What's next is to answer questions like the following:

- How well an individual student did in comparison to the entire class?
- What was the overall performance on a quiz?
- Is there a relationship between quiz performance and content access?

To address these questions, **graphic representations that are easy to interpret** are needed (Blits, 2017). Figure 3.17 illustrates the most common data visualization types.

A **bar graph** is a way of summarizing a set of **categorical data**. It displays the data using a number of rectangles, of the same width, each of which represents a particular category. Bar graphs can be displayed horizontally or vertically, and they are usually drawn with a gap between the bars (rectangles). For example, to answer

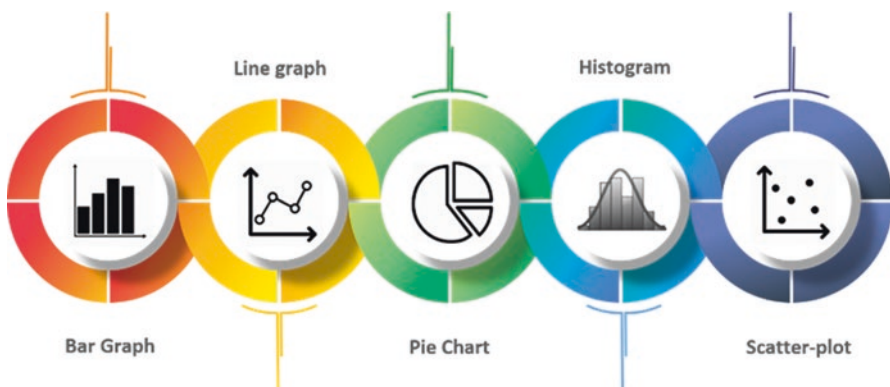


Fig. 3.17 Data visualization types

to how well an individual student did in comparison to the entire class, a bar graph can be used, where each student in the classroom is represented by a bar.

A **line graph** is particularly useful when we want to show the **trend of a variable over time**. Time is displayed on the horizontal axis (x-axis) and the variable is displayed on the vertical axis (y-axis). In the above example, a line graph can be used to showcase the overall performance on a quiz.

A **pie chart** is used to display a set of categorical data. It is a **circle**, which is divided into **segments**. Each segment represents a particular category. The area of each segment is **proportional** to the number of cases in that category. For example, a pie chart can be used to display the successful completion of an assignment.

A **histogram** is a way of summarizing data that are **measured on an interval scale** (either discrete or continuous). It is often used in **Exploratory Data Analysis (EDA)** to illustrate the features of the distribution of the data in a convenient form. In the above example, a histogram can be used to show the distribution of scores of students on the final exams.

A **scatter-plot** displays values for typically two variables for a set of data. The data are a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis. The scatter-plot is usually used to determine if a correlation exists between the data, and how strong it is. For example, a scatter-plot can show if there is a relationship between quiz performance and content access, or if there is a relationship between assignment completion and quiz performance.

It needs to be clarified that, in statistics, exploratory data analysis (EDA) is a preliminary data analysis approach to summarize the main characteristics of a given dataset, often with visual methods. EDA refers to a critical process of performing initial investigations on data to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations. It is a good practice to understand the data first and try to gather as many insights from it.

In most cases, a single graph does not contain all the information that is hidden in the data, cannot provide all the insights that might be needed to understand students' learning behaviour or outcomes, and is not sufficient for informed decision-making. The solution is to use combined graphs of the learning analytics metrics that all together can tell the story in the data. These combined graphs are called **dashboards**. *"A dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance"* (Few, 2004). Here are five examples of learning analytics dashboard implementations, in relation to the educational objective they aim to address.

LAPA – Learning Analytics for Prediction & Action The goal of LAPA dashboard is to inform learners' online learning behaviour to learners themselves and the instructor and guide their learning in a smart, personalized way. The first version of LAPA (Fig. 3.18) consists of 7 graphs. The graph chosen for the online activity summary is the scatterplot, where individual learners can choose the X-axis and



Fig. 3.18 The LAPA dashboard. (Source: Park & Jo, 2015)

Y-axis to locate their position in class. The other 6 graphs are provided with a trend line of their activity every week along with the average activity information of their peers. All graphs in LAPA are updated every week until end of semester (Park & Jo, 2015).

LADA – Learning Analytics Dashboard for Advisers LADA is a learning analytics dashboard that supports academic advisers in compiling a semester plan for students based on their academic history. LADA also includes a prediction of the academic risk of the student (Gutiérrez et al., 2018). LADA visualizes two categories of information: a) The chance of success and prediction quality components b) The various information card components designed to support the adviser (Fig. 3.19).

LISSA – Learning Dashboard for Insights and Support during Study Advice LISSA provides an overview of every key moment in chronological order up until the period in which the advising sessions are held: the grades of the positioning test (a type of entry-exam without consequence), mid-term tests, January



Fig. 3.19 The LADA dashboard. (Source: Gutiérrez et al., 2018)

exams, and June exams. A general trend of performance is visualised at the top: the student path consists of histograms showing the position of the student among their peers per key moment (Charleer et al., 2018). LISSA is shown in Fig. 3.20.

SmartKlass (Moodle) SmartKlass™ is a Learning Analytics dashboard for Institutions, Teachers and Students. By analyzing student’s behavioural data SmartKlass™ creates a rich picture of the evolution of the students in an online course: it can help teachers to identify the students lagging behind, help teachers to identify the students that content is not challenging enough for them, help teachers to compare participation and results to other courses, so the teachers can take action (Fig. 3.21). Students can also learn about their performance, individually and compared with the group.

Acrobatiq The Learning Dashboard (Fig. 3.22) generates summary graphs, tables and reports and dynamically displays student learning estimates, engagement data and activity data in real time. It enables faculty, students, and other stakeholders to visualize and act on student learning performance. It can be used for revealing what students did/not learn, quantifying how well students have learned each skill, identifying consequential patterns in students’ learning behaviours, and measuring effectiveness of instructional and design choices.

Signals Course Signals was developed to allow instructors the opportunity to employ the power of learner analytics to provide real-time feedback to a student. Course Signals relies not only on grades to predict students’ performance, but also



Fig. 3.20 The LISSA dashboard. (Source: Charleer et al., 2018)

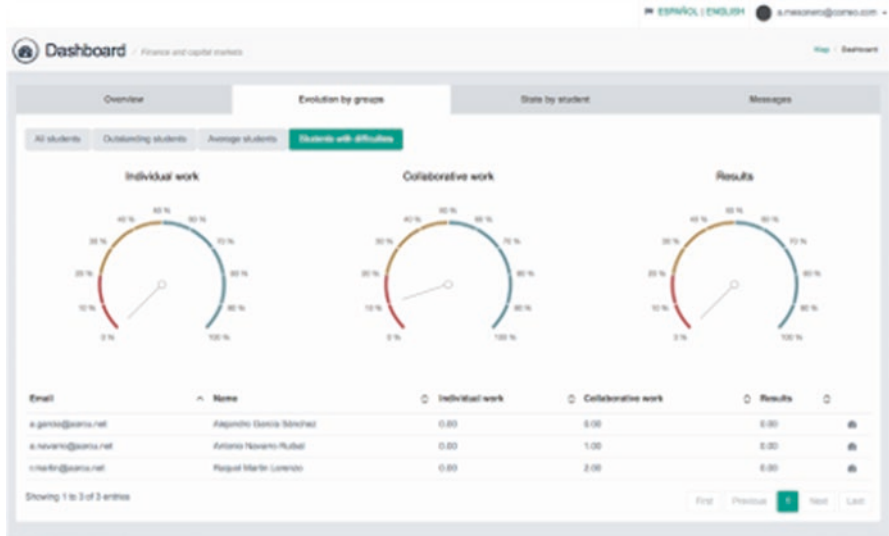


Fig. 3.21 The SmartKlass. (Source: https://moodle.org/plugins/local_smart_klass)

demographic characteristics, past academic history, and students’ effort as measured by interaction with Blackboard Vista, Purdue’s learning management system (Arnold & Pistilli, 2012). The [Course Signals Explanation](#) video (see useful video resources) is a brief introduction to Signals.

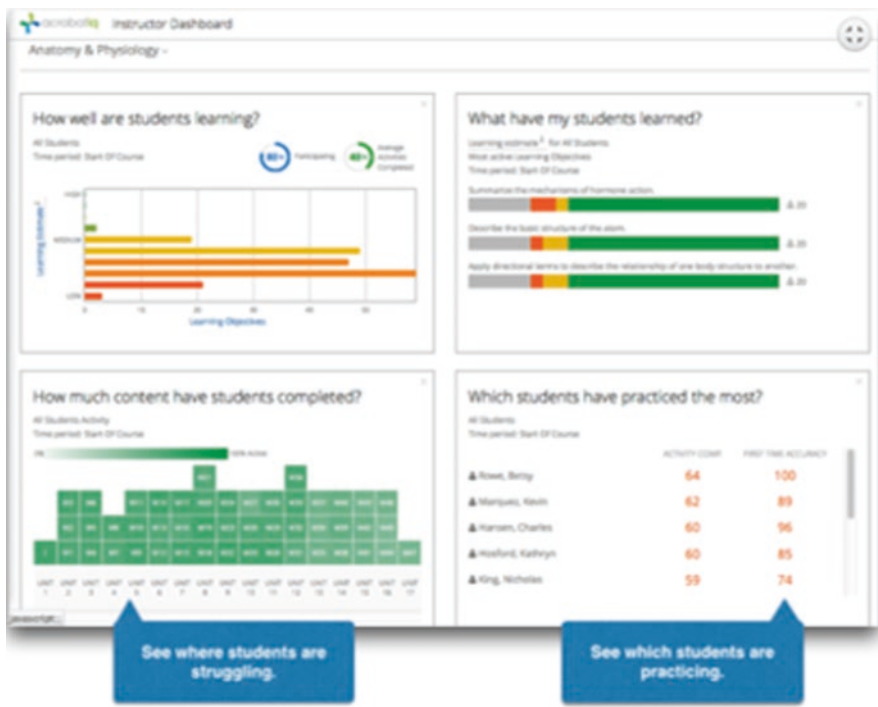


Fig. 3.22 The Acrobatiq. (Source: <https://www.acrobatiq.us/products/the-learning-dashboard.html>)

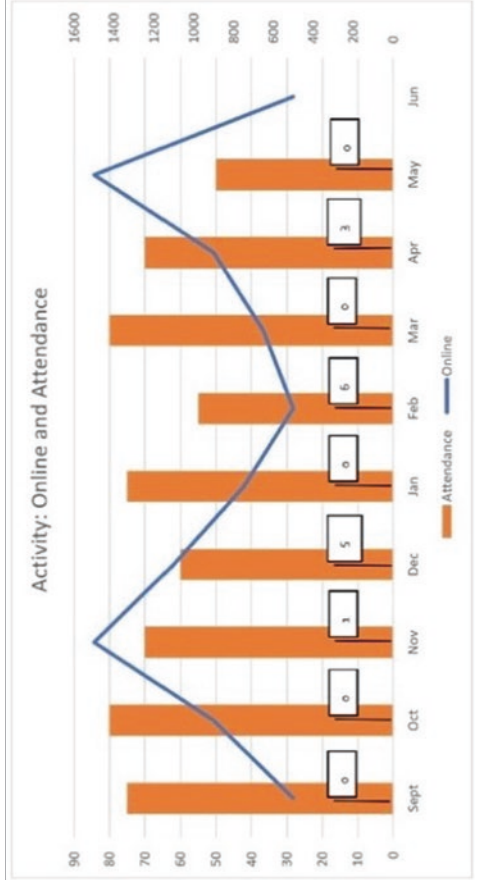
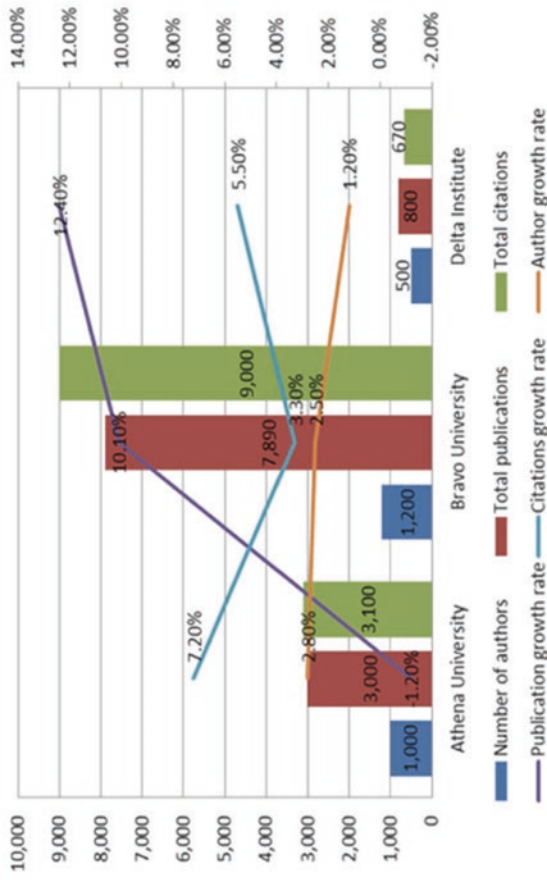
KlassData The learning process in virtual environments is more complex to analyze, but the generated data unlocks the power of learning analytics and opens the door to personalized paths in education. The [KlassData: Learning Analytics for Education](#) video (see useful video resources) application explains how KlassData works.

Questions and Teaching Materials

1. Match the visualizations (from the left column) to the respective evaluation of data presentation clarity (i.e., “Easy to understand” / “Difficult to understand” in the right column)?

Correct answer: 1.b / 2.b / 3.b / 4.a.

(a) Easy to understand



2. What is the purpose of dashboards?

- (a) **To present data results together so that they tell a story or reveal insights together, that isn't possible if left apart**
- (b) **To summarize categorical or numerical data**
- (c) **To grasp difficult concepts or identify new patterns**
- (d) **To produce and deliver richly interactive visualizations**

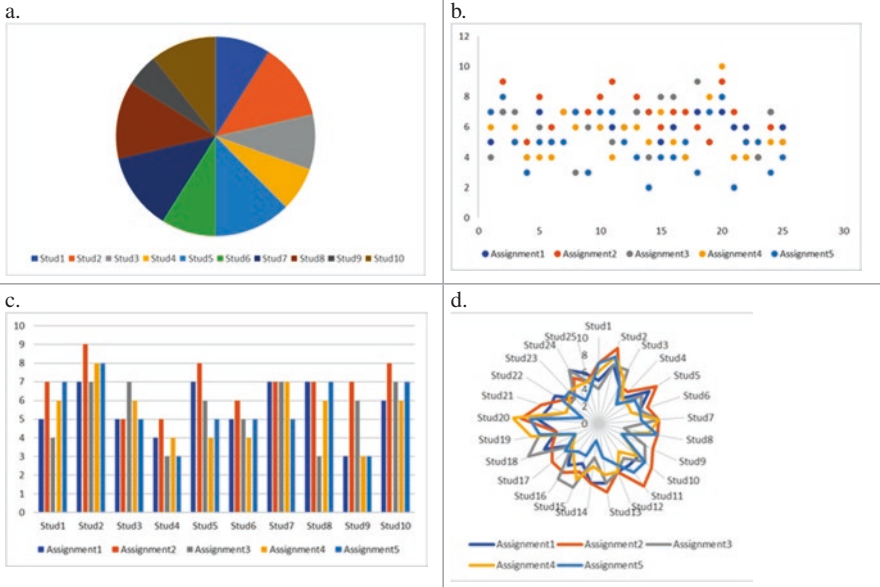
Correct answer: a.

3. What type of graph is more appropriate to present all students' scores on monthly assignments and the average class performance, and what for the distribution of the scores?

- (a) **A scatter plot to visualize the students' scores on monthly assignments and the average class performance, and a histogram for the distribution of the scores**
- (b) **A bar graph with a line to visualize the students' scores on monthly assignments and the average class performance, and a histogram for the distribution of the scores**
- (c) **A bar graph with a line to visualize the students' scores on monthly assignments and the average class performance, and a pie for the distribution of the scores**
- (d) **A scatter plot to visualize the students' scores on monthly assignments and the average class performance, and a pie for the distribution of the scores**

Correct answer: b.

4. Select the visualization that better illustrated the performance of all students on all assignments



Correct answer: c.

5. Match the objective (from the right column), to the respective visualization dashboard in the left column.

1. SmartKlass	a. supports academic advisers in compiling a semester plan for students based on their academic history.
2. LISSA	b. Reveals what students learn, quantifies how well students have learned each skill, identifies patterns in students' learning behaviours, and measures effectiveness of instructional and design choices.
3. LAPA	c. Help teachers to identify the students lagging behind, help teachers to identify the students that content is not challenging enough for them, help teachers to compare participation and results to other courses, so the teachers can take action.
4. Acrobatiq	d. Informs learners' online learning behaviour to learners themselves and the instructor and to guide their learning in a smart and personalized way.
5. LADA	e. Provides an overview of every key moment in chronological order up until the period in which the advising sessions are held.

Correct answer: 1.c / 2.e / 3.d / 4.b / 5.a.

6. What is the main focus of the visualization dashboard systems that have been developed?
- (a) To capture moment-by-moment learning and students' achievements
 - (b) To increase students' awareness of their own progress, guide self-learning, and support self-regulation of learning
 - (c) To predict students' progress during the semester and make content recommendations
 - (d) To monitor individual students' learning and reveal gaps, misunderstandings, or difficulties and help teachers tailor their instruction to the students' needs

Correct answer: d.

7. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to elaborate on your response about the data representation techniques in learning analytics, in the following reflective task:

1. Provide 2 examples of learning analytics metrics and explain what type of representation method you would employ to demonstrate their role. Please, elaborate on your choices.
2. Assume that you want to get insight about learners' engagement in an online activity. What learning analytics metrics you would consider and what visualizations would you provide on a dashboard to monitor how these metrics change? Please, elaborate on your decisions/suggestions.

3.4 Interpreting Learning Analytics and Inferring Learning Changes

3.4.1 *Making Sense of Learners' Data Analytics and Analysis Results*

The **intersection of learning science with data and analytics** enables more sophisticated ways of **making meaning** to support student learning. All these available learner and context data “carry” so much knowledge about the learners and the learning processes, that remains hidden and waits to be revealed. But data from tracking systems are not inherently intelligent. *Hit counts* and *access patterns* do not really explain anything. **The intelligence is in the interpretation of the data;** what all those statistics about the learner's data and measurements can inform us about. For example, *login frequencies*, *time-spent on tasks* or *numbers of forum posts* do **not measure the impact** on students' learning. However, the data analysis



Fig. 3.23 The path from learners to knowledge

techniques can reveal **potential relationships between metrics** that otherwise, in a human-analysis perspective, would be **undiscoverable** or even ignored. In the above example, learning analytics metrics such as *time-spent* or *frequencies of attempts* can be used to identify *specific units of study or assignments* in a course that are difficult (or trivial) for most of the students, and reveal *the correlation between task-difficulty and student behaviour*. Ideally, data analysis techniques enable the visualization of interesting data that in turn sparks the investigation of this data. Figure 3.23 illustrates the path from learners and their data to the interpretation of learning analytics.

The statistical analysis uses a *combination of potentially actionable metrics* to predict an outcome that needs attention and improvement. For example, to predict the successful completion of an assignment, metrics can include *measurable events*, such as time-spent on-task, on-task mental effort, number of attempts to solve a task, frequency of question posing, frequency of help-seeking, etc. *Less obvious data* can also be used, such as non-cognitive variables, like stress levels, emotional intensity, attention, etc. Analyses provide a score for each student, so students can be grouped objectively into categories needing high-, medium- or no-intervention to successfully complete the assignment. **The analysis cannot say that the learning analytics metrics caused the outcome, but it can show what combination of indicators is related to the outcome.** Your data reports and visualizations will help you to *identify historical trends and correlations*, which you can use to understand *what happened* and (probably) *why*.

Behavioural data can also be used to track *students' approaches to study*. For example, *frequency and sequence of interactions* can be tracked, as students engage with learning tasks. While this may **not directly measure student learning**, it can provide insights on the student's on-task activity and help to **identify strategies** that could improve *how they plan* and *regulate their study*.

Data science is promising to have a substantial influence on the understanding of learning in online and blended learning environments. This, of course, implies a shift on the typical role of educators, from being instructors and facilitators to performing some of the tasks data analysts usually hold (Fig. 3.24). They need to be able to discover the patterns in the data and convey the meaning in educational terms, that is to interpret the analysis results into meaningful learning schemas.



Fig. 3.24 The different roles of the educator in relation to data

The more an educator will use the learning analytics metrics, tools and visualization dashboards, the more she will understand what the story that the data can tell is, and what the most important patterns in the data are in explaining students' engagement, progress and outcomes. The analysis might reveal correlations between metrics that the educator had never thought of before, and behavioural patterns that are repeated from student to student and from class to class.

As the educator moves from efficiency metrics to effectiveness metrics to outcomes (see Sect. 3.2.2), she should keep in mind that all **metrics are proxies for what ultimately matters**. The different types of analytics facilitate the selection of the most appropriate metrics and guide their interpretation. Next, we elaborate on how the analysis outcomes associate with the learning analytics objectives and the analytics types.

As already discussed, the common objectives of learning analytics include monitoring learners' progress, modelling learners/learners' behaviour, detecting learner's emotions, predicting learning performance/dropout/retention, generating feedback, providing recommendations, guiding adaptation, increasing self-reflection/self-awareness, and facilitating self-regulation. To address these objectives, four types of learning analytics can be used, namely descriptive, diagnostic, predictive and prescriptive analytics. The infographic by [CommLabIndia](#) and the [article by eLearningIndustry](#) give a comprehensive overview of different levels of learning analytics and of how bases of and approaches to using analytics can lead to deeper insights.

Each analytics type can be supported and facilitated by specific data analysis methods that are appropriate for that type of data transformations. For example,



Fig. 3.25 The learning analytics types with respect to the objectives and actions

descriptive statistics and simple visualizations (using bar graphs, histograms, etc.) are the suitable analysis technique to provide descriptive analytics. Similarly, correlation analysis better facilitates diagnostic analytics, whereas regression analysis is commonly used for prediction purposes, and as such it is an indicative analysis technique for predictive analytics. When it comes to prescriptive analytics, more sophisticated analysis techniques can be employed (e.g., heuristics, machine learning), which, however, require strong background in data science and are beyond the scope of this chapter. Depending on the objectives and the types of analytics used, the interpretation of the analysis results can vary from gaining insights, to making decisions, to taking actions (Fig. 3.25).

For example, let's assume that, in anticipation, an educator wants to early predict students' success in the final exams in order to provide them proactive feedback, recommendations, support their self-regulated learning strategies, and prevent failure or drop-out. Let's also assume that the educator has available all the data from the students' activity during the semester (online participation, assignments' completion, quizzes' scores, etc.). The learning management system the educator is using can provide all the *descriptive statistics* about students' misconceptions, engagement, achievement, progress, etc., and deliver this information using multiple visualizations of the different learning analytics metrics, demonstrating some critical interrelationships between them and facilitating some *diagnostic* operations. The dashboard can also provide the result from a regression analysis in graphical formats that considers the most critical metrics and forecasts the evolution of the *prediction* variable (e.g., success in final exams) and displays the tendencies in the metrics. If the educator combines all this graphical information, that is the result of the analytics processing, she will be able to *associate the numerical facts with each student's progress and learning needs*.

Questions and Teaching Materials**1. How can learning analytics contribute to human learning?**

- (a) Learning analytics can measure the impact on learning
- (b) Learning analytics can directly measure human learning
- (c) Learning analytics can show what combination of indicators is related to the outcome
- (d) Learning analytics metrics can show what has caused the learning outcome

Correct answer: c.

2. Steven is an etutor. For his online course, he wants to identify areas that require improvement – e.g., learner engagement or the effectiveness of course delivery – and he also wants to identify gaps and performance issues early, before they become problems. What type of analysis methods and analytics he should use?

- (a) Descriptive statistics (e.g., mean, standard deviation, min, max) – descriptive analytics (e.g., course enrolments, course compliance rates, what learning resources are accessed and how often)
- (b) Correlation analysis (e.g., ANOVA, t-test) – descriptive analytics (e.g., course enrolments, course compliance rates, what learning resources are accessed and how often)
- (c) Regression analysis – predictive analytics (e.g., high/low performance, high/low engagement)
- (d) Machine learning (e.g., classification) – predictive analytics (e.g., high/low performance, high/low engagement)

Correct answer: a.

3. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to elaborate on your response about the learning analytics interpretations, in the following reflective task:

1. Provide 2 examples of learning analytics objectives and explain what learning analytics type you would employ to achieve those objectives. Please, elaborate on your choices

3.4.2 Explaining the Data Analysis Results in an Educationally Meaningful Manner to Understand Learners and the Environment they Learn In

What analytics cannot do by themselves is improve instruction. While they can point to areas in need of improvement and they can identify engaging practices, the numbers cannot make suggestion for improvements. This requires a human intervention.

Intervention should be personalized to the learner – based on their engagement and/or performance data and any personal information you may have. For example, if the educator notices that a student stopped participating in online forums just before their performance began to drop, it would be proper to encourage the student to resume their involvement in the forums. At the same time, it could be helpful to get feedback from the student to find out why they stopped participating. There may have been an event in the course or some other obstacle that the educator should address in order to facilitate the student's involvement in the online forums.

Effective intervention may involve adapting teaching styles. If students tend to do better with certain kinds of media, interactivity, or assessments, the course design should be adapted to enable better learning. However, some learning professionals are hesitant to initiate a learning analytics practice for two reasons: the perception that they must address everything at once, and the concern that leadership will use the insights in a penalizing way. The [Learning Analytics to inform teaching practice](#) video (see useful video resources) explains how learning analytics can be used to inform teaching practise.

If a metric is not informing a decision, there's no need to keep gathering it. If it is, optimize the specific data and learn how to turn it into insights that inform decisions that matter. Over time, add more metrics, always keeping in mind the decisions they inform. The data one collects should be a combination of engagement and performance data – but it is important to make sure that one is not collecting information that will not use. The [Jisc Learning Analytics: Making data useful](#) video (see useful video resources) demonstrates an example of how data can be effectively used and how one can give meaning to data.

Questions and Teaching Materials

1. What is the first step teachers should consider for using learning analytics?
 - (a) **How to design the feedback and intervention using learning analytics?**
 - (b) **What data should they collect to transform into learning analytics metrics?**
 - (c) **What learning analytics metrics should they use to solve the problem at hand?**
 - (d) **What kind of problem or aspect of learning you they want to detect and act on the learning environment?**

Correct answer: d.

3.5 Concluding Self-Assessed Assignment

3.5.1 Introduction

In order to proceed, you are requested to complete a concluding self-assessed assignment. This self-assessed assignment is a real-life scenario activity (based on the use case of the instructional designer David), using a rubric across three proficiency levels and an exemplary solution rating. When you have completed this assignment, you will assess it yourself, following the rubric which will list the criteria required and give guidelines for the assessment.

This self-assessed assignment procedure consists of 5 steps:

- **Step 1.** Real life scenario
- **Step 2.** Getting familiar with the assessment rubric
- **Step 3.** Prepare your answer
- **Step 4.** Review a sample solution
- **Step 5.** Self-evaluate your answer

3.5.2 Step 1. Real Life Scenario

David is an instructional designer. He always aims to create engaging learning activities and compelling course content. Recently he has been organizing the educational material and learning and assessment activities for a new course, and he wants to design a dashboard to monitor progress, engagement, and performance, both for individual students and for the whole class, that will advance the learning experience. He has available several types of student data tracked by the LMS during students' activities (e.g., login data, content/ educational material access, time-stamp for each activity, file downloads, assignments completed, correctness of assignments, grades on assignments, posting on online forums, quiz scores, discussion participation, etc.), as well as demographic and enrolment data (e.g., age, gender, socioeconomic status, special education needs, course enrolment, etc.). It is important for David to deliver a dashboard that will increase students' self-awareness about their progress, motivate them to self-reflect and identify their needs, and finally enhance their retention and performance.

However, David is new in learning analytics and educational data literacy. Help David design a dashboard that will integrate students' needs and will address the above learning objectives.

3.5.3 Step 2. Getting Familiar with the Assessment Rubric

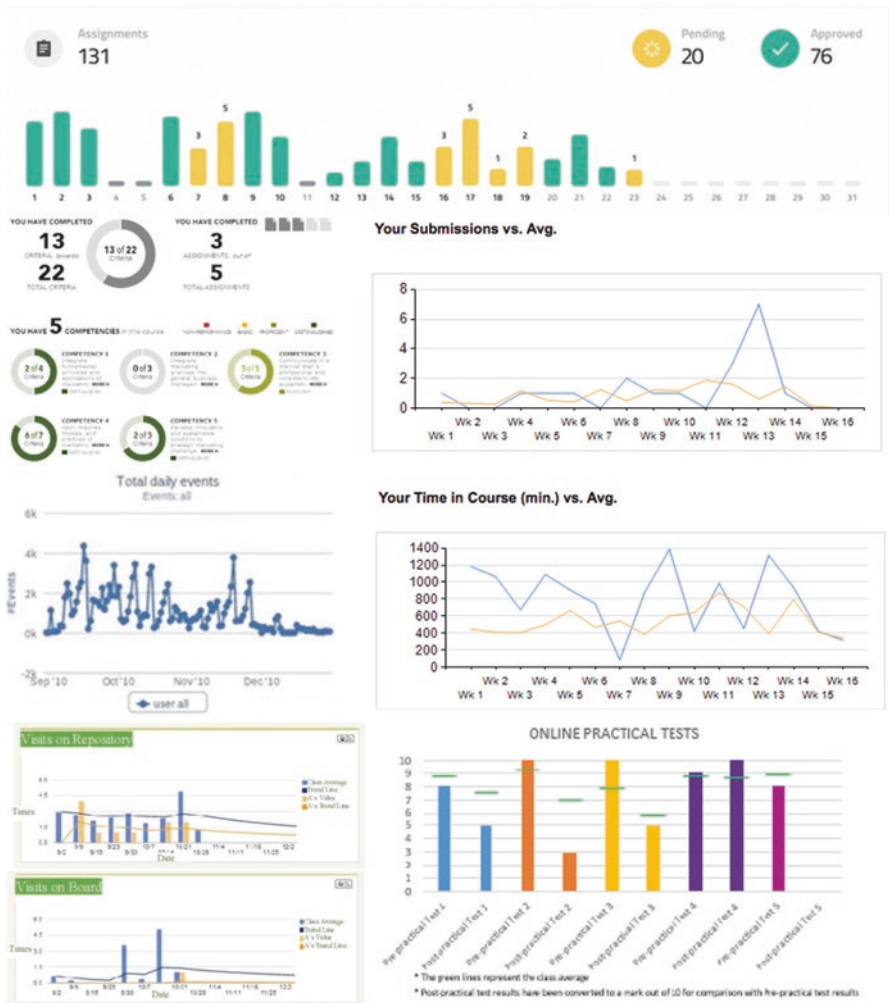
David has searched on the Internet for Learning Analytics Dashboards samples, to get some design inspiration, and designs and Initial ExampleDB.

Please help David to evaluate this Initial ExampleDB using the Rubrics for assessing the dashboard and to identify potential issues.

ACTIVITY/PRACTICE QUESTION (Discussion) We encourage you to elaborate on your response about the evaluation of the Initial ExampleDB created by David, in the following discussion task, by posting your thoughts on the discussion board. You may discuss:

1. Does this example dashboard comply with the dashboard design criteria in the Rubric?
2. If not, what would you advise David to modify, so that this dashboard serves and addresses the learning objectives he has set?

3.5.3.1 Initial Example DB



* The green lines represent the class average
 * Post-practical test results have been converted to a mark out of 10 for comparison with pre-practical test results

3.5.3.2 Rubric for Assessing the Example DB

Criteria	Unacceptable (1)	Good/solid (3)	Exemplary (5)
Clarity: Graphs and charts answer the specific question/ address the specific objective.	Graphs and charts do not have clearly defined topics and fail to address specific questions.	Graphs and charts have somewhat clearly defined topics but fail to address specific questions.	Graphs and charts have concise and clearly defined topics that address specific questions.
Information quality: Graphs and charts complement each other – there is no information redundancy.	Graphs and charts are not relevant to each other and there is information redundancy.	Graphs and charts are relevant to each other but there is information redundancy.	Graphs and charts complement each other well, without redundant information.
Appropriateness: Graphs and charts types are appropriate for the data types and scale.	None/a few of the graphic types used are suited for the type and scale of the data they represent.	Most graphic types used are well suited for the type and scale of the data they represent.	All graphic types used are well-suited for the type and scale of the data they represent.
Interpretability: Graphs and charts convey meaningful information to the viewer and facilitate decision making.	Graphs and charts are overwhelmed by text, color, and symbolism, that are irrelevant to the question the visualization seeks to answer.	Graphs and charts contain some color, symbolism, or text that is irrelevant to the question the visualization seeks to answer.	Graphs and charts contain no color, symbolism, or text that is irrelevant to the question the visualization seeks to answer.
Organization: Graphs and charts are well organized and easy to follow.	Graphs and charts are a bit of a mess. The dashboard is not easy to follow.	Graphs and charts are visually appealing and somewhat well organized. The dashboard is somewhat easy to follow.	Graphs and charts are visually appealing and well organized. The dashboard is easy to follow.
Usability: Legends describe and explain every graphic variable type employed.	Either there is no legend, or it does not describe any of the graphic variable types present in the visualization.	Legend describes a few/most of the graphic variable types present in the visualization.	Legend describes every graphic variable type present in the visualization.
Aesthetics: Visualization makes appropriate use of color.	More than 12 colors are used. Similar colors are adjacent.	Fewer than 12 colors are used, but similar colors are not adjacent.	Fewer than 8 colors used in visualization, colors are discrete.

3.5.4 Step 3. Prepare Your Answer

Please assist David to design a prototype of the dashboard that will integrate students' needs and will address the above learning objectives. For this purpose, you will have to design a detailed prototype of the dashboard (using pen and paper and/or any tool of your preference). Please, consider that David (and you!) has available all types of student data he might need, and help him select the most appropriate ones for each learning objective, mapping the learning analytics metrics to the respective and most suitable type of graph and/or chart.

ACTIVITY/PRACTICE QUESTION (Reflect on) We encourage you to elaborate on your response about the prototype of the dashboard that David wants to design to increase students' self-awareness about their progress, motivate them to self-reflect and identify their needs, and finally enhance their retention and performance, in the following reflective task:

1. What are the key indicators to include (visualize) in the dashboard and can help students monitor their progress in the course?
2. What are the key indicators to include (visualize) in the dashboard and can help students monitor their performance on assignments and quizzes?
3. What are the key indicators to include (visualize) in the dashboard and can help students monitor their engagement in the course and course materials and tools?

3.5.5 Step 4. Review a Sample Solution

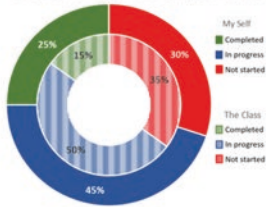
Please review a sample of an Exemplary Sample Solution that follows the criteria specified in the Rubrics for assessing the dashboard.

ACTIVITY/PRACTICE QUESTION (Reflect on) We encourage you to elaborate on your response about the Exemplary Sample Solution that follows the criteria specified in the Rubrics for assessing the dashboard, in the following reflective task: Do you identify any design requirements that you did not take under consideration when creating your dashboard prototype?

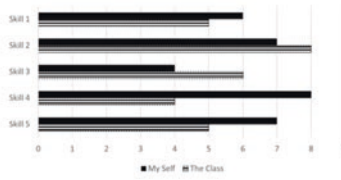
3.5.5.1 Exemplary Sample Solution

My Progress

Overview of Assignments Completion Status



Progress in Skills

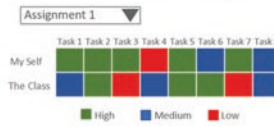


My Performance

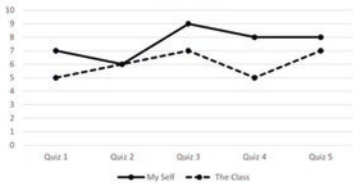
Completed Assignments Scores



Task Performance in Assignment



Quizzes Scores

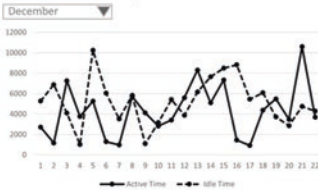


Quiz Summary

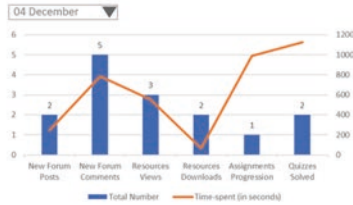


My Engagement

Daily Time in The Course



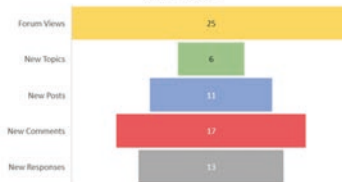
Summary of my Daily Events



Resources Usage

Resource	View	Download
Video 1	✓	✗
Video 2	✓	✓
Video 3	✗	✗
File 1	✓	✓
File 2	✗	✓
File 3	✓	✗
File 4	✓	✗

Forum Usage



3.5.6 Step 5. Self-Evaluate Your Answer

Now that you have seen the Exemplary Sample Solution, please rate your initial answer (evaluate the dashboard you created), using the Rubric table below.

Criteria	Unacceptable (1)	Good/solid (3)	Exemplary (5)
Clarity: Graphs and charts answer the specific question/ address the specific objective.	Graphs and charts do not have clearly defined topics and fail to address specific questions.	Graphs and charts have somewhat clearly defined topics but fail to address specific questions.	Graphs and charts have concise and clearly defined topics that address specific questions.
Information quality: Graphs and charts complement each other – there is no information redundancy.	Graphs and charts are not relevant to each other and there is information redundancy.	Graphs and charts are relevant to each other but there is information redundancy.	Graphs and charts complement each other well, without redundant information.
Appropriateness: Graphs and charts types are appropriate for the data types and scale.	None/a few of the graphic types used are suited for the type and scale of the data they represent.	Most graphic types used are well suited for the type and scale of the data they represent.	All graphic types used are well-suited for the type and scale of the data they represent.
Interpretability: Graphs and charts convey meaningful information to the viewer and facilitate decision making.	Graphs and charts are overwhelmed by text, color, and symbolism, that are irrelevant to the question the visualization seeks to answer.	Graphs and charts contain some color, symbolism, or text that is irrelevant to the question the visualization seeks to answer.	Graphs and charts contain no color, symbolism, or text that is irrelevant to the question the visualization seeks to answer.
Organization: Graphs and charts are well organized and easy to follow.	Graphs and charts are a bit of a mess. The dashboard is not easy to follow.	Graphs and charts are visually appealing and somewhat well organized. The dashboard is somewhat easy to follow.	Graphs and charts are visually appealing and well organized. The dashboard is easy to follow.
Usability: Legends describe and explain every graphic variable type employed.	Either there is no legend, or it does not describe any of the graphic variable types present in the visualization.	Legend describes a few/most of the graphic variable types present in the visualization.	Legend describes every graphic variable type present in the visualization.
Aesthetics: Visualization makes appropriate use of color.	More than 12 colors are used. Similar colors are adjacent.	Fewer than 12 colors are used, but similar colors are not adjacent.	Fewer than 8 colors used in visualization, colors are discrete.

References

- Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In S. B. Shum, D. Gasevic, & R. Ferguson (Eds.), *2nd international conference on learning analytics and knowledge (LAK '12)* (pp. 267–270). ACM.
- Blits, S. (2017). 10 useful ways to visualize your data with examples. <https://www.sisense.com/blog/10-useful-ways-visualize-data-examples/>
- Campbell, J. P., & Oblinger, D. G. (2007). Academic analytics. *EDUCAUSE Quarterly*, *October*, 2007.
- Charleer, S., Moere, A. V., Klerkx, J., Verbert, K., & De Laet, T. (2018). Learning analytics dashboards to support adviser-student dialogue. *IEEE Transactions on Learning Technologies*, *11*(3), 389–399. <https://doi.org/10.1109/TLT.2017.2720670>
- Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thiis, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning (IJTEL)*, *4*(5/6), 318–331.
- Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. In S. B. Shum, D. Gasevic, & R. Ferguson (Eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK '12)* (pp. 134–138). <https://doi.org/10.1145/2330601.2330636>
- Data quality – Wikipedia. (2022). https://en.wikipedia.org/wiki/Data_quality
- Davidian, M., & Louis, T. (2012). Why statistics? *Science*, *335*(6077), 12. <https://doi.org/10.1126/science.1218685>
- Dillard, J. (2017). 5 Most important methods for statistical data analysis. <https://www.bigskyasociates.com/blog/bid/356764/5-Most-Important-Methods-For-Statistical-Data-Analysis>
- Drachsler, H., & Greller, W. (2016). Privacy and Analytics – it’s a DELICATE issue. A Checklist to establish trusted Learning Analytics. In *6th Learning analytics and knowledge conference 2016, April 25–29, 2016*.
- Experian Data Quality. (2015). The data quality benchmark report. <http://cdn.qas.com/us-marketing/whitepapers/data-quality-benchmark-report-2015.pdf>
- Ferguson, R., Macfadyen, L. P., Clow, D., Tynan, B., Alexander, S., & Dawson, S. (2014). Setting Learning Analytics in Context: Overcoming the Barriers to Large-Scale Adoption. *Journal of Learning Analytics*, *1*(3), 120–144.
- Few, S. (2004). *Dashboard confusion*. Perceptual edge.
- General plugins (Local): SmartKlass™ Learning Analytics Moodle. https://moodle.org/plugins/local_smart_klass
- Gutiérrez, F., Seipp, K., Ochoa, X., Chiluiza, K., De Laet, T., & Verbert, K. (2018). LADA: A learning analytics dashboard for academic advising. *Computers in Human Behaviour*. <https://doi.org/10.1016/j.chb.2018.12.004>
- Kenton, W. (2018). Descriptive statistics. https://www.investopedia.com/terms/d/descriptive_statistics.asp
- Khalil, M., & Ebner, M. (2016). What is learning analytics about? A survey of different methods used in 2013-2015. CoRR, abs/1606.02878.
- Kiss, M. (2018). 10 tips for presenting data. <https://resources.observepoint.com/blog/10-tips-for-presenting-data>
- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice Hall.
- Kuhar, C.W. (2010). Experimental design: Basic concepts, encyclopedia of animal behaviour. <https://www.sciencedirect.com/topics/medicine-and-dentistry/inferential-statistics>
- Lang, C., Siemens, G., Wise, A., & Gasevic, D. (Eds.). (2017). *Handbook of learning analytics* (1st ed.). SOLAR.
- Lebied, M. (2016). How to build interesting data reports people love to read. <https://www.datapine.com/blog/data-reports-people-love-to-read/>
- Macfadyen, L. P., Dawson, S., Pardo, A., & Gašević, D. (2014). Embracing big data in complex educational systems: The learning analytics imperative and the policy challenge. *Research & Practice in Assessment*, *9*(2014), 17–28.

- Mihăiloaie, C. (2015). Data quality dimensions – From accuracy to uniqueness. <https://www.performancemagazine.org/data-quality-dimensions-from-accuracy-to-uniqueness/>
- Misiejuk, K., & Wasson, B. (2017). State of the field report on learning analytics. SLATE report 2017-2. Bergen.
- National Forum for the Enhancement of Teaching and Learning in Higher Education. (2017, August 15). “Guide to Data Quality,” in teachingandlearning.ie. Last Accessed May 5, 2019, <https://www.teachingandlearning.ie/resource/guide-to-data-quality-2/>
- Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data Mining in Practice: A systematic literature review of empirical evidence. *Educational Technology & Society*, 17(4), 49–64.
- Park, Y., & Jo, I. (2015). Development of the learning analytics dashboard to support Students’ learning performance. *J. UCS*, 21, 110–133.
- Pipino, L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communication ACM*, 45(4), 211–218. <https://doi.org/10.1145/505248.506010>
- Prinsloo, P., & Slade, S. (2017). Ethics and learning analytics: Charting the (un)charted. In L. Charles, S. George, W. Alyssa, & G. Dragan (Eds.), *Handbook of learning analytics* (pp. 49–57). SOLAR.
- Rathi, A. (2018). Inferential statistics for data science. <https://towardsdatascience.com/inferential-statistics-for-data-science-b0075670fc8a>
- Robbins, S. (2017). Learning analytics and the value of understanding L&D metrics.
- Scapin, R. (2015). Learning analytics in education: Using Student’s big data to improve teaching. <https://www.slideshare.net/rscapin/learning-analytics-47463622>
- Scheffel, M., Drachler, H., & Specht, M., 2015. Developing an evaluation framework of quality indicators for learning analytics. In *5th international conference on learning analytics and knowledge – LAK ‘15* (pp. 16–20). ACM Press.
- Sharma, H. (2019). What is data science? A beginner’s guide to data science. <https://www.edureka.co/blog/what-is-data-science/>
- Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *The American Behavioral Scientist*, 57(10), 1509–1528.
- SOLAR. (2011). Call for papers of the 1st international conference on Learning Analytics & Knowledge (LAK 2011). Retrieved from <https://tekri.athabascau.ca/analytics/>
- Soltanpoor, R., & Sellis, T. (2016). Prescriptive analytics for big data. In M. Cheema, W. Zhang, & L. Chang (Eds.), *Databases theory and applications. ADC 2016. Lecture notes in computer science* (Vol. 9877). Springer.
- The Importance Of Learning Analytics In Learning And Development. <https://elearningindustry.com/learning-analytics-benefits-ld>
- Tsachuridou, E. (2015). Why ethics and law are not the same thing | INTHEBLACK. <https://www.intheblack.com/articles/2015/04/01/why-ethics-and-law-are-not-the-same-thing>
- Tsai, Y. S., & Gasevic, D. (2017). Learning analytics in higher education—Challenges and policies: A review of eight learning analytics policies. In *Proceedings of the seventh international learning analytics & knowledge conference* (pp. 233–242). ACM.
- Tsai, Y. S., Moreno-Marcos, P. M., Jivet, I., Scheffel, M., Tammets, K., Kollom, K., & Gašević, D. (2018). The SHEILA framework: Informing institutional strategies and policy processes of learning analytics. *Journal of Learning Analytics*, 5(3), 5–20.

Useful Video Resources

External video. Learning analytics [4:02].

External video. What is student data? [2:59].

External video. Data quality matters – Tech vision 2018 trend [7:31].

External video. Learning analytics: The need for a code of ethics[9:59].

- External video. Statistics – Introduction to statistics [3:45].
- External video. The value of data visualization [1:42].
- External video. Course signals explanation [2:15].
- External video. KlassData: Learning analytics for education [2:23].
- External video. Learning analytics to inform teaching practice [7:44].
- External video. Jisc learning analytics: Making data useful [9:22].

Further Readings

- 3 tips to turn learning analytics into actionable outcomes. <https://elearningindustry.com/turn-learning-analytics-into-actionable-outcomes-3-tips>
- 5 Myths about the ‘Data Quality’ that could derail your analytics project. (2018). <https://www.cloudmoyo.com/5-myths-about-the-data-quality-that-could-derail-your-analytics-project/>
- Adaptive courses with learning analytics | Acrobatiq. <http://acrobatiq.com/products/the-learning-dasboard/>
- Bekker, A. (2017). 4 types of data analytics to improve decision-making. <https://www.scnsoft.com/blog/4-types-of-data-analytics>
- Blackboard Analytics for Learn | Blackboard. <https://www.blackboard.com/education-analytics/analytics-for-learn.html>
- Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. US Department of education, office of. *Educational Technology*, 1, 1–57.
- Bodily, R., Ikahihifo, T. K., Mackley, B., et al. (2018). The design, development, and implementation of student-facing learning analytics dashboards. *Journal of Computing in Higher Education*, 30(3), 572–598. <https://doi.org/10.1007/s12528-018-9186-0>
- Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal*, 14, 2. <https://doi.org/10.5334/dsj-2015-002>
- Corrin, L., Kennedy, G., de Barba, P. G., Lockyer, L., Gašević, D., Williams, D., Dawson, S., Mulder, R., Copeland, S., & Bakharia, A. (2016). *Completing the loop: Returning meaningful learning analytic data to teachers*. Office for Learning and Teaching.
- Danelutti, S. (2016). Making sense of sensemaking. <https://innerventur.es/2016/05/31/making-sense-of-sensemaking/>
- Data Presentation. <http://academic.sun.ac.za/emergencymedicine/trrm/module5/BS1-3.htm>
- Descriptive and Inferential Statistics. (2018). <https://statistics.laerd.com/statistical-guides/descriptive-inferential-statistics.php>
- Digital Learning and Analytics – How to Make Organizational Training Profitable. <https://www.valamis.com/blog/digital-learning-and-analytics-how-to-make-organization-training-profitable>
- Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, 28, 68–84. <https://doi.org/10.1016/j.iheduc.2015.10.002>
- Gosh, P. (2018). The impact of data quality in the machine learning era – DATAVERSITY. <https://www.dataversity.net/impact-data-quality-machine-learning-era/#>
- Johnson, A. (2017). Ethics and data - a powerful partnership. <https://www.baesystems.com/en/cybersecurity/blog/data-and-ethics-a-powerful-partnership>
- Kay, D., Korn, N., & Oppenheim, C. (2012). Legal, risk and ethical aspects of analytics in higher education. *JISC CETIS Analytics Series*. 1(6). <http://publications.cetis.ac.uk/2012/500>
- Klašnja-Miličević, A., Ivanović, M., & Budimac, Z. (2017). Data science in education: Big data and learning analytics. *Computer Applications in Engineering Education*, 25, 1066–1078. <https://doi.org/10.1002/cae.21844>
- Learning Analytics Examples. <https://isit.arts.ubc.ca/learning-analytics-examples/>
- Learning analytics specification – MoodleDocs. (2017). https://docs.moodle.org/dev/Learning_Analytics_Specification

- Maydon, T. (2017). The 4 types of data analytics. <https://insights.principa.co.za/4-types-of-data-analytics-descriptive-diagnostic-predictive-prescriptive>
- Merceron, A., Blikstein, P., & Siemens, G. (2015). Learning analytics: From big data to meaningful data. *Journal of Learning Analytics*, 2(3), 4–8.
- Open University. (2014). Policy on ethical use of student data for learning analytics. <http://www.open.ac.uk/students/charter/sites/www.open.ac.uk.students.charter/files/files/ecms/web-content/ethical-use-of-student-data-policy.pdf>
- Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450. <https://doi.org/10.1111/bjet.12152>
- Prieto, L. P., Sharma, K., Dillenbourg, P., & Rodríguez-Triana, M. J. (2016). Teaching analytics: Towards automatic extraction of orchestration graphs using wearable sensors. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK '16)* (pp. 148–157). ACM. <https://doi.org/10.1145/2883851.2883927>
- Quantitative data: definition, types, analysis and examples. <https://www.questionpro.com/blog/quantitative-data/>
- Reimann, P. (2016). Connecting learning analytics with learning research: The role of design-based research. *Learning: Research and Practice*, 2(2), 130–142. <https://doi.org/10.1080/023735082.2016.1210198>
- Rienties, B. (2018). The power of learning analytics to impact learning and teaching: A critical perspective. <https://iet.open.ac.uk/presentations/265>
- Rienties, B., Cross, S., Marsh, V., & Ullmann, T. (2017). Making sense of learner and learning big data: Reviewing five years of data wrangling at the Open University UK. *Open Learning: The Journal of Open, Distance and e-Learning*, 32(3), 279–293. <https://doi.org/10.1080/02680513.2017.1348291>
- Ruggiero, D. (2016). What metrics don't tell us about the way students learn. <https://theconversation.com/what-metrics-dont-tell-us-about-the-way-students-learn-59271>
- Sampson, D. (2016a). Teaching analytics: Analyze your lesson plans to improve them. <https://elearningindustry.com/teaching-analytics-analyze-lesson>
- Sampson, D. (2016b). Learning analytics: Analyze your lesson to discover more about your students – eLearning industry. <https://elearningindustry.com/learning-analytics-analyze-lesson>
- Scheffel, M., Drachler, H., Stoyanov, S., & Specht, M. (2014). Quality indicators for learning analytics. *Educational Technology & Society*, 17(4), 117–132.
- Slade, S., & Tait, A. (2019). *Global guidelines: Ethics in learning analytics*. International Council for Open and Distant Education (ICDE).
- Volaric, T., & Ljubic, H. (2017). Learner and course dashboards for intelligent learning management systems. In *25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)* (pp. 1–6).
- What is Learning Analytics & How can it be used?. <https://www.northeastern.edu/graduate/blog/learning-analytics/>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 4

Teaching Analytics



4.1 Introduction and Scope

4.1.1 Scope

The goal on this chapter is to:

- introduce the basics of methods and tools for analysing and interpreting educational data for facilitating educational decision making, including course and curricula design.

4.1.2 Chapter Learning Objectives

This chapter learning objectives	Learn2Analyse Educational data Literacy Competence profile
Know how to identify data sources within the educational design process	1.1
Be able to explain key concepts of data quality for data collected in the educational design process	1.2
Be able to design automated and semi-automated interventions based on educational data	4.4
Know and understand how to revise course tasks and contents based on educational data	5.1
Be able to construct adequate criteria and indicators for evaluating the impact of a data-driven intervention in educational design of online and blended courses	5.2

(continued)

This chapter learning objectives	Learn2Analyse Educational data Literacy Competence profile
Be able demonstrate awareness of data privacy and distinguish between different levels of data protection in educational design of online and blended courses	6.2
Be able to explain the differences between the concepts of authorship, ownership, data access, renegotiation, and data-sharing in education design	6.3

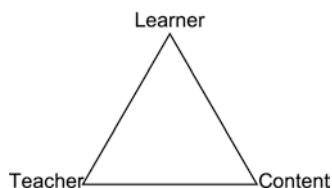
4.1.3 Introduction

This chapter will introduce the basics of methods and tools for analysing and interpreting educational data for facilitating educational decision making, including course and curricula design. Teaching analytics use static and dynamic information about the design of learning environments for near real-time modelling, prediction, and optimisation of learning artefacts, learning designs, learning processes, curriculum designs, and educational decision making.

- **The first topic focuses on data sources for supporting teaching analytics. You will reflect on the instructional design process and locate data sources for optimising learning environments as well as understand limitations and requirements for data quality.**
- **The second topic includes critical reflections on data ethics and privacy principles. You will build awareness toward data privacy, distinguish different levels of data protection and identify issues of authorship, ownership, data access and data-sharing.**
- **The third topic addresses the application and communication of educational data and analytics findings to various stakeholders. You will design and revise automated and semi-automated interventions as well as apply methodologies for improving the design of learning environments, teaching processes as well as curricula.**

In order to warm-up, explore the “didactic triangle” in Fig. 4.1 and reflect what data may stem from each of the key concepts and related interactions.

Fig. 4.1 Didactic triangle



4.2 Data Sources for Supporting Teaching Analytics

4.2.1 *Learning and Teaching*

According to Seel and Ifenthaler (2009), learning involves a stable and persisting change of what a person knows, requiring mental representations. The processes that result in learning (e.g., learning activities) can be and often are distinguished from the products of learning (e.g., learning outcomes), as discussed by Spector et al. (2014). Several theories of learning have been postulated over the 20th and 21st centuries: Behaviourism, Cognitivism, Constructivism, Connectivism. Figure 4.2 illustrates the theories of learning, how learning is conceptualised and what factors may influence learning.

Teaching is considered as deliberate actions undertaken with the intention of facilitating learning. Hence, when it comes to teaching, the relevant input and output characteristics for designing a learning environment need to be identified. The elementary parts of teaching include matching of content elements, psychological operations and didactic considerations (Scheerens et al., 2007). Doyle (1985) defines seven key criteria for effectiveness of teaching as follows:

1. **Teaching goals are clearly formulated;**
2. **The course material to be followed is carefully split into learning tasks and is placed in sequence;**
3. **The teacher explains clearly what the pupils must learn;**
4. **The teacher regularly asks questions to gauge pupils' progress and understanding;**
5. **Pupils have ample time to practice what has been taught, with much use of "prompts" and feedback;**
6. **Skills are taught until mastery is automatic;**
7. **The teacher regularly tests the pupils and calls on them to be accountable for their work.**

Table 4.1 provides an overview of phases in the structuring of teaching (Scheerens et al., 2007):

4.2.2 *Design of Learning Environments*

Learning environments are physical or virtual settings in which learning takes place. Learning theory provides the fundament for the design of learning environments. However, there is no simple recipe for designing learning environments (Ifenthaler, 2012). Generally, the design of learning environments includes the three simple questions: What is taught? How is it taught? How is it assessed? Yet, the design of learning environments is not simply asking the above stated three questions. Rather, it includes a systematic analysis, planning, development, implementation, and evaluation phases (see Fig. 4.3).

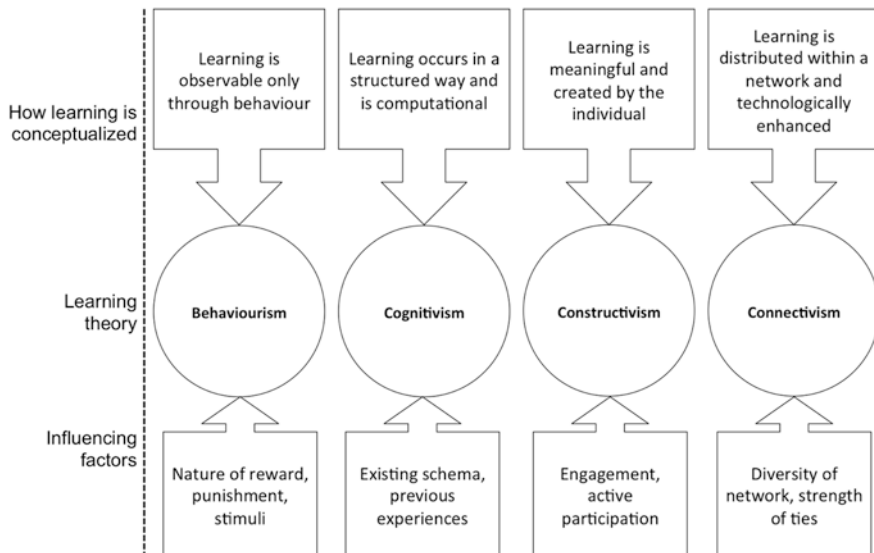


Fig. 4.2 Overview on learning theories. (Ifenthaler & Schumacher, 2016a, b)

Table 4.1 Structuring of teaching

Content dimension	Psychological dimension
decomposition of content in sequences that represent the structure of the subject matter area	taxonomy of cognitive, affective, and psychomotor operations that reflect increasing complexity
COMBINE BOTH DIMENSIONS IN	SEQUENCES OF INSTRUCTIONAL OBJECTIVES
creating tasks and task sequences with pedagogical potential	taking into consideration cognitive complexity and emotional meaning of tasks
COMBINE BOTH IN	LESSON PLANS AND SCRIPTS
actual teaching in which multiple representations and explanations of content elements are given	taking into consideration possible misconceptions, typical difficulties, and frequently made mistakes
COMBINE BOTH IN	TEACHING
constructing content elements for the development of items for formative and summative assessment instruments	adding representations of expected psychological operations, with different degree of complexity to each content element of item
COMBINE BOTH IN	ITEM BANKS AND TESTS IN WHICH DIFFICULTY LEVEL AND ABILITY ARE IDENTIFIABLE DIMENSIONS

Scheerens et al. (2007)

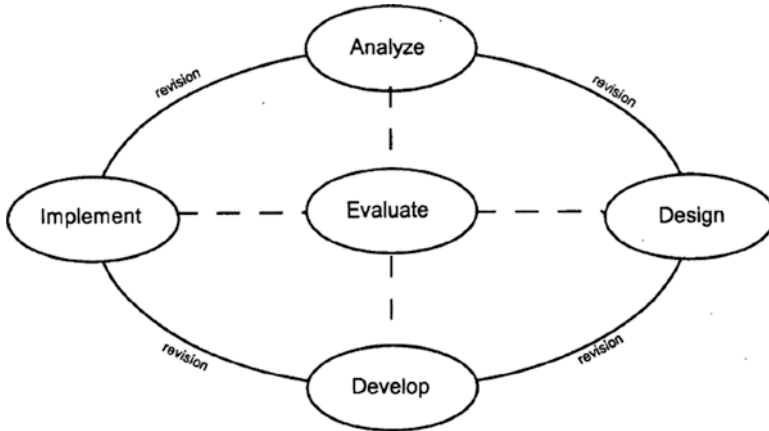


Fig. 4.3 The ADDIE model. (Gustafson & Branch, 2002)

The analysis phase includes needs analysis, subject matter content analysis, and job or task analysis. The design phase includes the planning for the arrangement of the content of the instruction. The development phase results in the tasks and materials that are ready for instruction. The implementation phase includes the scheduling of instruction, training of instructors, preparing time tables, and preparing evaluation parts. The evaluation phase includes various forms of formative and summative assessments.

4.2.3 Learning Design

Whereas instructional design is rooted in behaviourist learning theories and seems to on the one hand focus on learning products, such as learning objects and machine-readable representations and on the other hand on delivery systems and the advancement of the automation of designs, learning design is rooted in constructivist learning theories and seems to focus on making the design process explicit and shareable. Table 4.2 includes a list of definitions of learning design exemplifying the roots of this research field.

4.2.4 TPACK Model

At the heart of good teaching with technology are three core components: content, pedagogy, and technology, plus the relationships among and between them (Mishra & Koehler, 2006). The TPACK model (i.e., Technological Pedagogical Content Knowledge) describes the core components of teaching where content (what you

Table 4.2 Overview on definitions of learning design

Author(s)	Definition
Agostinho (2006, p. 3)	A learning design is a representation of teaching and learning practice documented in some notational form so that it can serve as a model or template adaptable by a teacher to suit his/her context.
Conole (2008, p. 191)	The range of activities associated with creating a learning activity and crucially provides a means of describing learning activities.
Conole (2013, p. 121)	A methodology for enabling teachers/designers to make more informed decisions in how they go about designing learning activities and interventions, which is pedagogically informed and makes effective use of appropriate resources and technologies. This includes the design of resources and individual learning activities right up to curriculum-level design. A key principle is to help make the design process more explicit and shareable. Learning design as an area of research and development includes both gathering empirical evidence to understand the design process, as well as the development of a range of learning design resource, tools and activities.
Dalziel (2008, p.8)	A framework to describe a sequence of educational activities in an online environment.
Dobozý (2013, p. 68)	A way of making explicit epistemological and technological integration attempts by the designer of a particular learning sequence or series of learning sequences.
Hale (2016, p. 1)	Learning design is the process of designing learning experiences (planning, structuring, sequencing) through facilitated activities that are pedagogically informed, explicit, and make better use of technologies in teaching.
Koper (2006, p. 13) (2008, p. 191)	The description of the teaching-learning process that takes place in a unit of learning. The key principle in learning design is that it represents the learning activities and the support activities that are performed by different persons (learners, teachers) in the context of a unit of learning. These <i>activities</i> can refer to different <i>learning objects</i> that are used during the performance of the activities (e.g. books, articles, software programmes, pictures), and it can refer to <i>services</i> (e.g. forums, chats, wiki's) that are used to collaborate and to communicate in the teaching-learning process.
Mor & Craft (2012, p. 86)	Learning design is the creative and deliberate act of devising new practices, plans and activities, resources and tools aimed at achieving particular educational aims in a given context.
Papadakis (2012, p. 258)	The creation of sequences of learning activities, which involve groups or learners interacting within a structured set of collaborative environments.

Ifenthaler et al. (2018)

teach) and pedagogy (how you teach) must be the basis for any technology that is used in a learning environment in order to support and enhance learning (see Fig. 4.4).

Pedagogical Content Knowledge (PCK) is the knowledge that teachers have about their content and the knowledge that they have about how teach that specific content. Technological Pedagogical Knowledge (TPK) is the set of skills which teachers develop to identify the best technology to support a particular pedagogical approach. Technological Content Knowledge (TCK) is the set of skills which teachers acquire to help identify the best technologies to support their students as they learn content.

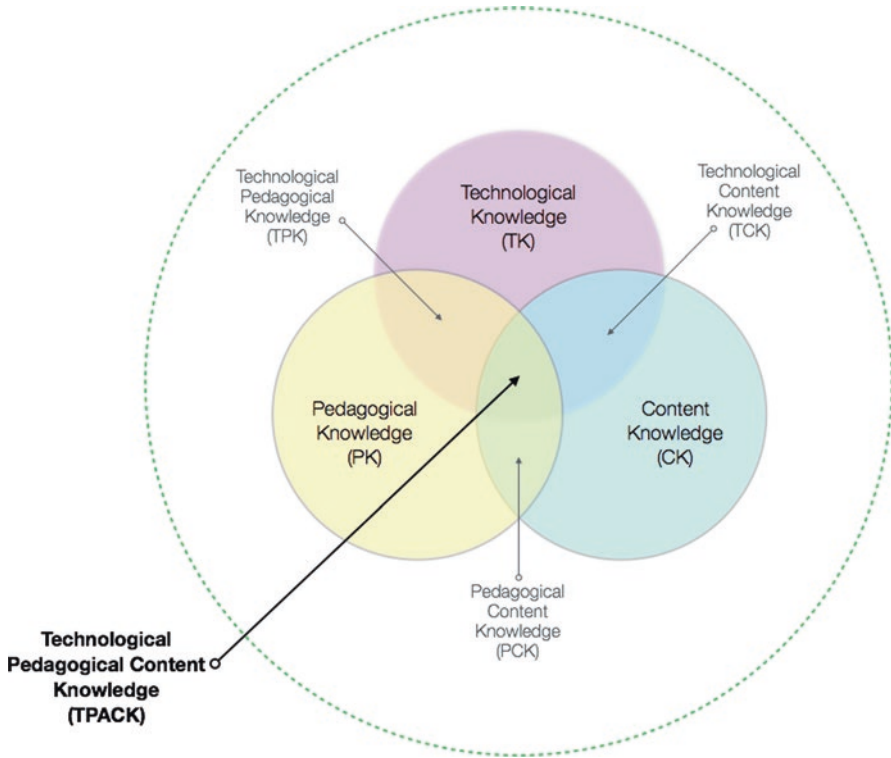


Fig. 4.4 The TPACK model. (Mishra & Koehler, 2006)

Questions and Teaching Materials

1. For each theory of learning, influencing factors for learning can be distinguished. Which of the following factors can be related to Behaviourism?
 - (a) Active participation and networking.
 - (b) Building ties for social networks.
 - (c) Providing rewards in relation to achievements.
 - (d) Active engagement and stimuli for social collaboration.

Correct Answer: c

2. Learning Design and Instructional Design have different origins and conceptual foundations. Still, the purpose of these disciplines can be summarised as follows:

- (a) **They include seven procedural steps for reviewing learning quality.**
- (b) **They include a systematic perspective on the planning, implementation and evaluation of learning environments.**
- (c) **They include assessment criteria for competences.**
- (d) **They include two features of learning strategies.**

Correct Answer: b

3. **The didactic triangle consists of ...**

- (a) **Learner**
- (b) **Teacher**
- (c) **Content**
- (d) **Technology**
- (e) **Environment**

Correct Answer: a, b, c

4. **Effective teaching includes ...**

- (a) **Time pressure**
- (b) **Formative assessment**
- (c) **Pure exploration**
- (d) **Clearly formulated goals**
- (e) **Sequenced learning tasks**

Correct Answer: b, d, e,

5. **A key principle of learning design includes ...**

- (a) **Limitation of learning time**
- (b) **Representation of learning activities**
- (c) **Real-time monitoring of performance**
- (d) **Governance of exam regulations**
- (e) **Exclusion of supportive technology**

Correct Answer: b

6. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to reflect on your teaching experience supported through data. You may reflect on:

- **Do you refer to different sources of data when designing your learning environments?**
- **Do you analyse data to inform your teaching practice in (near) real-time, i.e., while teaching a class (online or face-to-face)?**
- **Do you use specific tools to collect and analyse data to inform your teaching?**
- **Do you strictly follow one theory of learning (e.g., Behaviourism, Cognitivism) when designing your learning environments?**
- **Do you evaluate each phase of the instructional design (i.e., analysis, design, development, implementation) process before moving to the next phase?**

4.3 Data Sources Within the Instructional Design Process

4.3.1 *Broadening the Perspective for Data-Driven Education*

The idea of grounding instructional design decisions on educational data has been around for some time. Traditionally, evidence-based instruction has used summative evaluation data to (re-)design instructional programs and systems. Immediate interventions based on formative evaluations have been conducted significantly less frequently. Research on learning and instruction brought attention to additional data sources, as summarized in the 3P-model of teaching and learning (Biggs et al., 2001): “presage” data focuses on student factors and the teaching context, “process” data on learning focused activities, and “product” data on learning outcomes. Historically, most of this data has been collected with social science research methods. Surveys and questionnaires have been used most often, at times supplemented by different forms of observations.

Online teaching and learning has created a wide range of opportunities for data-driven education. A lot more data sources are now at hand, as well as new technologies for data handling and analysis. While it seems impossible to create a complete list of potential data sources, educational data and the respective data sources can be systematized with a number of attributes.

Educational data can be primary data (direct data), that is: data that is especially collected for the purpose of improving teaching and learning. Secondary (indirect) data, on the other hand, has been initially collected for other purposes, but can also be used for teaching analytics. Data can be collected candid and transparent. This means that the purpose of data collection is clear, as for example in a direct survey, interview or an eye-tracking study. Educational data can also be collected automatically and with little or no transparency, as it is the case with user trails within the system or logging data. Educational data can be oriented toward the learning

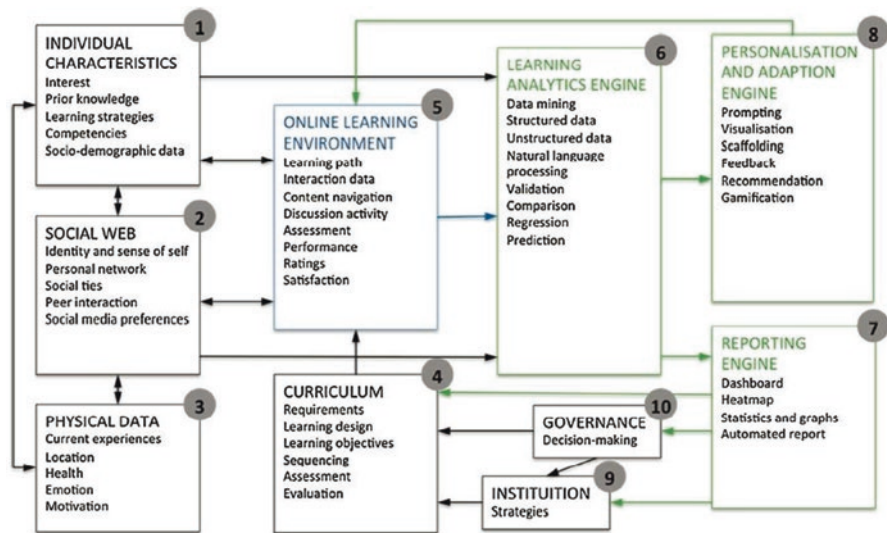


Fig. 4.5 Holistic learning analytics framework. (Ifenthaler & Widanapathirana, 2014)

outcome or the learning process. Educational data can be static, that is: stable over a defined period of time (e.g., personality traits). Educational data can be dynamic, that is: volatile over the course run (e.g., motivational and emotional states). Educational data can be sourced on the individual or on a collective level. Educational data can be idiosyncratic or generalizable. Educational data can refer to learner variables (person focus; i.e. personal learning goals), it can refer to contextual variables (environment focus; i.e. curricular learning objectives), or to learning behaviour (person-environment-interaction focus; i.e. course performance). Finally, educational data can be open and accessible to anyone (i.e., curriculum data, syllabi), or it can be protected (i.e., discussion posts within a course environment) – a distinction which is not always as straightforward as it may sound (Greller & Drachsler, 2012).

4.3.2 Data Sources Within a Holistic Analytics Framework

Ifenthaler and Widanapathirana (2014) developed and empirically validated a holistic learning analytics framework that connects a number of different data sources (#1 to #5). A major aim of this model is to create a link between learner characteristics (e.g., prior learning), learning behaviour (e.g., access of materials), and curricular requirements (e.g., learning objectives, sequencing of learning) (see Fig. 4.5).

4.3.3 Sources of Learner Data

Within the holistic learning analytics framework (see Fig. 4.5), three main areas of learner data and respective data sources have been differentiated. Characteristics of (1) individual learners include socio-demographic information, personal preferences and interests, responses to standardized inventories (e.g., learning strategies, achievement motivation, personality), demonstrated skills and competencies (e.g., computer literacy), acquired prior knowledge and proven academic performance, as well as institutional transcript data (e.g., pass rates, enrolment, dropout, special needs). Associated interactions with the (2) social web include preferences of social media tools (e.g., Twitter, Facebook, LinkedIn) and social network activities (e.g., linked resources, friendships, peer groups, web identity). Physical data (3) from outside the educational system is collected through various systems, for example through a library system (i.e., university library, public library). Other physical data may include sensor and location data from mobile devices (e.g., study location and time), or affective states collected through reactive tests (e.g., motivation, emotion, health, stress, commitments). Especially non-cognitive (i.e., emotional and motivational data) can provide deep insights into individual learning processes (D’Mello, 2017).

4.3.4 Sources of Online Learning Data

Furthermore, there are two areas of data and respective data sources related to online learning behaviour (see Fig. 4.6). Rich information is available from learners’ activities in the online learning environment (5) (i.e., learning management system, personal learning environment, learning blog). These mostly numeric data refer to logging on and off, viewing or posting discussions, navigation patterns, learning paths, content retrieval (i.e., learner-produced data trails), results on assessment tasks, responses to ratings and surveys. More importantly, rich semantic and context-specific information is available from discussion forums as well as from complex learning tasks (e.g., written essays, wikis, blogs). Additionally, interactions of facilitators with students and the online learning environment are tracked. Closely linked to the information available from the online learning environment is the curriculum information (5), which includes metadata of the online learning environment. These data reflect the learning design (e.g., sequencing of materials, tasks, and assessments), and learning objectives as well as expected learning outcomes (e.g., specific competencies). Ratings of materials, activities, and assessments as well as formative and summative evaluation data are directly linked to specific curricula, facilitators, or student cohorts (Ifenthaler & Widanapathirana, 2014).

In summary, teaching analytics use static and dynamic data sources for informing learning and teaching processes as well as outcomes. The Figure below summarises the profiles approach which includes static and dynamic data from students

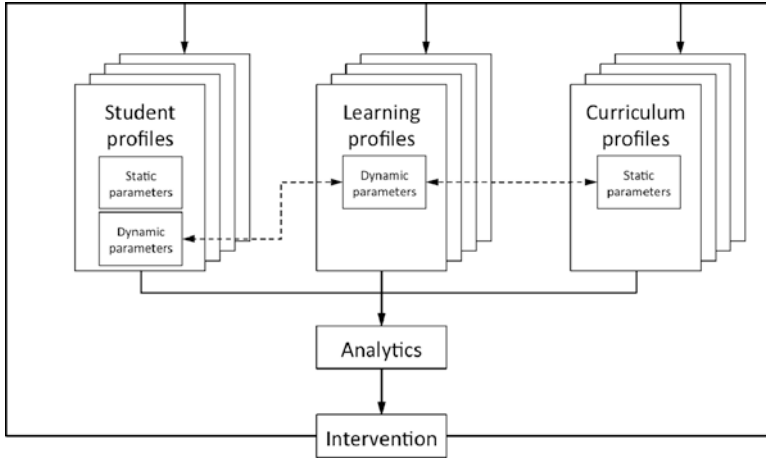


Fig. 4.6 Profiles approach using static and dynamic data. (Ifenthaler & Widanapathirana, 2014)

(e.g., demographic information, academic performance), dynamic data of learning behaviour (e.g., navigation pathways), and static data defined in the curriculum (e.g., learning outcomes, learning artefacts).

Questions and Teaching Materials

1. Which learning data can be related to the learning profile:

- (a) Forum activity, interaction with learning materials, assessment attempts
- (b) Forum posts and historical grades
- (c) Forum visits and learning objectives
- (d) Forum activity, emotional states, place of living

Correct Answer: a.

2. Why do teaching analytics require a reference to curricular statements, such as learning outcomes?

- (a) They help to understand the needs of a learner.
- (b) They function as benchmark for adaptive feedback a teacher can relate to.
- (c) They help the administrator to monitor the expertise of a teacher.
- (d) Active engagement and stimuli for social collaboration.

Correct Answer: d.

3. What outcomes can be produced from a reporting engine?

- (a) **Dashboard**
- (b) **Heatmap**
- (c) **Personalised help**
- (d) **Collaborative scaffolds**
- (e) **Automated report**

Correct Answer: a, b, e.

4. The profiles approach includes the following parameters

- (a) **Alpha-numeric parameters**
- (b) **Static parameters**
- (c) **Dynamic parameters**
- (d) **Component parameters**
- (e) **Change parameters**

Correct Answer: b, c.

5. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to reflect on your teaching experience supported through data. You may reflect on:

- **Are you able to access relevant data to inform your teaching anytime required?**
- **Are your students aware of data you are using for informing your teaching?**
- **A major aim of the holistic analytics model is to create a link between learner characteristics, learning behaviour, and curricular requirements. Please name three or more data sources for which it might be worthwhile to establish such a connection. Where do you see logical relationships that might be helpful for analytics?**
- **How would you try to collect emotional and motivational data? What could be feasible data sources?**

4.4 Key Concepts of Data Quality and Limitations of Data Meaningfulness

4.4.1 *Data Quality in Educational Contexts*

As the amounts of educational data grow larger, the issue of data quality is becoming more and more important. ‘Big Data’ in education is characterized by the same attributes as in other domains: Volume, Velocity, Variety, and Value (Katal et al., 2013). Volume refers to the tremendous volume of the data, usually measured in TB or above. Velocity means that data are being formed at an unprecedented speed and must be dealt with in a timely manner. Variety indicates that big data has all kinds of data types, and this diversity divides the data into structured data and unstructured data. Finally, Value represents low-value density. Value density is inversely proportional to total data size, the greater the big data scale, the less relatively valuable the data (Cai & Zhu, 2015).

Already on a smaller scale, data quality is of crucial importance for teaching and learning analytics, as ‘poor data’ can impede valid inferences and hamper subsequent educational interventions. However, there is no common definition of educational data quality to date. If the broad ISO 9000:2015 definition of quality is applied, data quality can be defined as the degree to which a set of characteristics of data fulfils pre-defined requirements. These requirements are usually described in quality dimensions, each with specific elements and indicators for measurement (Cai & Zhu, 2015).

Despite the complexity of the topic, the majority of the numerous frameworks on data quality share a common core of quality dimensions that can be transferred to education datasets (Akoka et al., 2007; Goasdoué et al., 2007; Laranjeiro et al., 2015): completeness, accuracy, consistency, freshness and relevancy.

4.4.2 *Core Dimensions of Data Quality*

Data Accuracy is defined as the correctness and precision used for representing real world data in an information system. Data needs to be precise, valid and errorfree. Three main accuracy definitions have been established in current research literature: (i) Semantic correctness which describes how well data represent states of the real-world, i.e., identifying the semantic distance between system-based data and real-world data. For instance, the recorded address “99, Main Street” is actually the address of Mary? (ii) Syntactic correctness related to the degree to which data is free of syntactic errors, for example misspellings and format discordances, i.e., identifying the syntactic distance between system-based data and expected data representation. For example, the address “99, Main Street” is valid and well written? (iii) Precision refers to the level of detail of data representation, i.e., identifying the gap between the level of detail of system-based data and its expected level of detail

(Peralta, 2006). For instance, the amount “€ 98” is a more precise representation of the cost of a product than “€ 100”.

Data Completeness is defined as the degree to which all relevant data have been recorded in an information system. It is expected that all relevant facts of the real world are represented in the information system (Gertz et al., 2004). Two aspects of completeness are differentiated: (i) Coverage meaning whether all required entities for an entity class are included; (ii) Density describing whether all data values are present (not null) for required attributes (Peralta, 2006).

Data Consistency refers to the degree to which data satisfies a set of integrity requirements. Common requirements of data constancy include check for null or missing values, key uniqueness or functional dependencies (Peralta, 2006).

Data Freshness introduces the idea of how old is the data: Is it fresh enough with respect to the user expectations? Has a given data source the more recent data? Is the extracted data stale? When was data produced? There are two main freshness definitions in the literature: (i) Currency describes how stale is data with respect to the sources. It captures the gap between the extraction of data from the sources and its delivery to the users. For example, given an account balance, it may be important to know when it was obtained from the bank data source. (ii) Timeliness describes how old is data (since its creation/update at the sources). It captures the gap between data creation/update and data delivery. For example, given a top-ten book list, it may be important to know when the list was created, no matter when it was extracted from sources (Akoka et al., 2007).

Data relevancy corresponds to the usefulness of the data. Among the huge volumes of data, it is often difficult to identify that which is useful. In addition, the available data is not always adapted to user requirements. This might lead to the impression of poor relevancy. Relevancy plays a crucial part in the acceptance of a data source. This dimension, usually evaluated by rate of data usage, is determined by the user and thus not directly measurable by quality tools.

4.4.3 Dimensions of Educational Data Quality

Valid examples for the core dimensions of educational data quality from the educational context could include the following (see Table 4.3):

4.4.4 Data Quality Problems

Laranjeiro et al. (2015) classify data quality problems with respect to the source of information: single or multiple. Single-source problems are related with the (wrong or absent) definition of integrity constraints. Multi-source problems relate with the integration of data from multiple sources, which, for instance, might hold different representations of the same values, or contradictions. Each of these two classes of

Table 4.3 Dimensions of educational data quality

Data quality dimension	Description	Example for educational data
Accuracy	Are the data free of errors?	Student number in a campus management system matches the student number in the learning management system
Completeness	Are necessary data missing?	Academic performance record includes all data points necessary to determine study progress (i.e. semester, courses passed, grades, ...)
Consistency	Are the data presented in the same format?	All requested event dates are delivered in a DD/MM/YY format
Freshness	Are the data up-to-date?	Learning analytics system reflects real-time behavior and performance data
Relevancy	Is the data useful for the task at hand?	Do I need the academic performance record to analyze student interactions?

problems are further divided into schema-level, which are related with defects in the definition of the data model and schema, and instance-level which are problems that are not visible at the schema level and cannot be prevented by restrictions at the schema level (or by redesign).

In exchange for the user-determined ‘relevancy’-Dimension, the authors added ‘Accessibility: The degree to which data can be accessed in a specific context of use’ to their synopsis of data quality problems (see Table 4.4).

Questions and Teaching Materials

1. An example for data accuracy is

- (a) **Academic performance record includes several data points of study progress**
- (b) **Event dates are stored in various formats**
- (c) **Student number in a campus management system matches the student number in the learning management system**
- (d) **Real-time user behaviour is stored for at least 10 days**

Correct Answer: c.

2. Volume is referring to

- (a) **The number of learners and teachers**
- (b) **The capacity of a human brain**
- (c) **The voice level related to data storage devices**
- (d) **The tremendous amount of the data, usually measured in TB or above**

Correct Answer: d.

Table 4.4 Data quality problems mapped into dimensions

Problem types		Data quality problems	Accessi- bility	Accuracy	Completeness	Consistency	Freshness	
Source	Level							
Single	Instance	Missing data		X	X			
		Incorrect data		X				
		Misspellings		X				
		Ambiguous data	X	X				
		Extraneous data	X				X	
		Outdated temporal data		X				X
		Misfielded values	X	X			X	
		Incorrect references		X				
		Duplicates	X					
	Schema	Domain violation		X				
		Violation of functional dependency		X				
		Wrong data type	X				X	
		Referential integrity violation	X	X			X	
		Uniqueness violation		X				
Multiple	Instance	Structural conflicts	X			X		
		Different word orderings	X			X		
		Different aggregation levels	X	X			X	
		Temporal mismatch		X			X	X
		Different units	X				X	
		Different representations	X				X	
		Schema	Use of synonyms	X				
	Use of homonyms		X					
	Use of special characters		X					
	Different encoding formats		X				X	

3. Missing data with reference to data quality can be mapped to

- (a) Accessibility
- (b) Accuracy
- (c) Freshness
- (d) Consistency
- (e) Completeness

Correct Answer: b, e.

4. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to reflect on your teaching experience supported through data. You may reflect on:

- **Please think of one type of educational data as introduced in the previous section. How would this data have to be characterised on the different dimensions of data quality in order to be good source of information? Please explain your indicators to the dimensions and explain your ratings according to those indicators.**

4.5 Data Ethics and Privacy Principles for Teaching Analytics

4.5.1 *Ethical and Privacy Challenges Associated with the Application of Educational Data Analytics*

Educational institutions have always used a variety of data about students, teachers and the learning environment, such as socio-demographic information, grades on entrance qualifications, or pass and fail rates, to inform their curricular planning, academic decision-making as well as for resource allocation. Such data can help to successfully predict student's dropout rates and to enable the implementation of strategies for supporting learning and instruction as well as retaining students (Ifenthaler & Tracey, 2016). However, serious concerns and challenges are associated with the application of data analytics in educational settings:

1. **Not all educational data is relevant and equivalent. Therefore, the validity of data and its analyses is critical for generating useful summative, real-time, and predictive insights.**
2. **Limited access to educational data generates disadvantages for involved stakeholders. For example, invalid forecasts may lead to inefficient decisions and unforeseen problems.**

3. **Information from distributed networks and unstructured data cannot be directly linked to educational data collected within an institution's environment.**
4. **Ethical and privacy issues are associated with the use of educational data for learning analytics. That implies how personal data is collected and stored as well as how it is analysed and presented to different stakeholders.**

Consequently, educational institutions need to address ethics and privacy issues linked to educational data analytics: They need to define who has access to which data, where and how long the data will be stored, and which procedures and algorithms to implement for further use of the available educational data (Ifenthaler, 2015).

4.5.2 Privacy in the Digital World

Within the digital world, many individuals are willing to share personal information without being aware of who has access to the data, how and in what context the data will be used, or how to control ownership of the data. Accordingly, data are generated and provided automatically by online systems, which limits the control and ownership of personal information in the digital world (Slade & Prinsloo, 2013).

There are several reasons why learners would like to keep their information private: First, there are competitive reasons, for example, if a learner performs poorly, a fellow student shall not know about it. Second, there are personal reasons, for example a learner might not want to share information about him-/ herself. There are also country-specific differences who owns the personal data. In the United States the collected data belongs to the collectors. In Europe the personal data belongs to the individual (e.g., the learner).

Table 4.5 provides an overview of privacy theories in the digital age. The first two concepts (1, 2) emphasize requirements for reaching privacy in a certain situation and focus on protection and normative or descriptive privacy. Early privacy theories (3) are based on control or limitation: Control refers to the influence of individuals on the flow of their personal data, whereas limitation means the possibility to prevent others from accessing personal data. Contemporary privacy theories (4) incorporate these earlier theories as well as normative and descriptive privacy concepts but go beyond them in being more holistic and applicable to different contexts (Ifenthaler & Schumacher, 2016).

4.5.3 Ethical Principles

Ethical principles for educational data analytics have been developed to underpin decision-making processes and provide guidance in the application of ethics (West et al., 2016). The key principles, as outlined and used in healthcare settings, are also relevant to the discussion of educational data analytics:

Table 4.5 Overview of privacy concepts

1. Individual has privacy in a particular situation if they are offered three protections	Protection from interference	Protection from information access by others	Protection from intrusion
The rights of an individual to be left alone and free from intrusion and interference			
2. Two broad classifications of privacy situations	Normative privacy Zones of privacy	Descriptive privacy	
	Individuals are protected by cultural norms: Formal laws and informal policies	Privacy can be expected by natural means, such as physical barriers	
3. Early theories of privacy	Control theory	Limitation theory	
	Allowing individuals control over their personal information	Limitations on the persons who could gain access to personal information	
4. More recently proposed information privacy theories aimed at achieving necessary protections by building on earlier theories and normative and descriptive privacy	Floridi ontological theory of information privacy	Nissenbaum contextual integrity theory of informational privacy	Moor & Tavani hybrid RALC theory of privacy
		Contemporary privacy theories are more holistic and go beyond the early theories of privacy; they were developed to apply them to diverse contexts	

Ifenthaler and Schumacher (2016a, b)

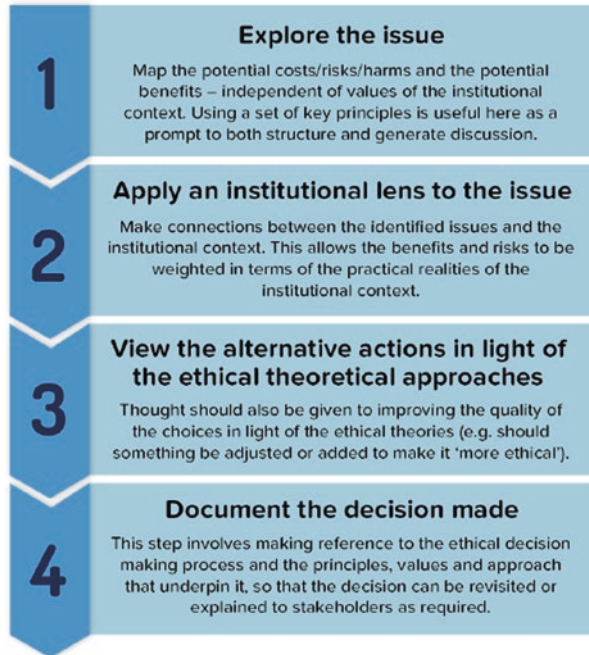
1. Respect for Autonomy generally translates to the idea of self-determination and the right of people to make their own decisions.
2. Non-maleficence essentially means that we should do no harm.
3. Beneficence means that in addition to doing no harm, we should also pursue good outcomes for others.
4. Justice translates into the concept of fairness and is often related to the distribution of resources based on equity, need, effort, merit and the market.

Figure 4.7 presents a four step framework that views ethical decision making as an operational process. The aim of this framework is to concisely model how a complex issue can be mapped, refined, decided on, and documented within a fairly linear process that would suit the busy operating environments of most institutions. There may be circumstances where reflection or new information means retracing earlier steps and the framework does not oppose doing so (West et al., 2016).

Questions and Teaching Materials

1. **Ethical key principles for educational data analytics include ...**
 - (a) **Respect for autonomy**
 - (b) **Building advantages over competitors**

Fig. 4.7 Ethical decision making process for learning analytics. (West et al., 2016)



- (c) Pursue good outcomes for all involved stakeholder
- (d) Doing no harm to every involved stakeholder

Correct Answer: a, c, d.

2. **Descriptive privacy is based on the assumption of natural means, e.g., physical barriers**
- (a) No
 - (b) Yes

Correct Answer: b.

3. **Reasons for learners to keep data private include ...**
- (a) Environmental reasons
 - (b) Competitive reasons

- (c) **Technical reasons**
- (d) **Personal reasons**

Correct Answer: b, d.

4. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to reflect on your teaching experience supported through data. You may reflect on:

- **Do you include your learners when designing a data analytics survey?**
- **Do you ask for consent to collect data from your learners?**

4.6 Identify Issues of Authorship, Ownership, Data Access and Data-Sharing

4.6.1 *Privacy Calculus*

To enhance the acceptance of educational data analytics, it is relevant to involve all stakeholders as early as possible. Students need to be considered in particular, as they take on two roles in the educational data analytics: (1) as producers of analytics data and (2) as recipients of the analyses derived from them (Slade & Prinsloo, 2013).

Figure 4.8 shows the deliberation process for disclosing information for educational data analytics. Students assess their concern over privacy on the basis of the specific information required for the learning analytics system (e.g., name, learning history, learning path, assessment results, etc.). This decision can be influenced by risk-minimizing factors (e.g., trust in the learning analytics systems and/or institution, control over data through self-administration) and risk-maximizing factors (e.g., non-transparency, negative reputation of the learning analytics system and/or institution). Concerns over privacy are then weighed against the expected benefits of the learning analytics system. The probability that the students will disclose required information is higher if they expect the benefits to be greater than the risk. Hence, the decision to divulge information on learning analytics systems is a cost-benefit analysis based on available information to the student.

4.6.2 *Educational Data Analytics Benefits*

Table 4.6 provides a matrix outlining the benefits of educational data analytics for stakeholders including three perspectives: (1) summative, (2) real-time/formative, and (3) predictive/prescriptive. The summative perspective provides detailed

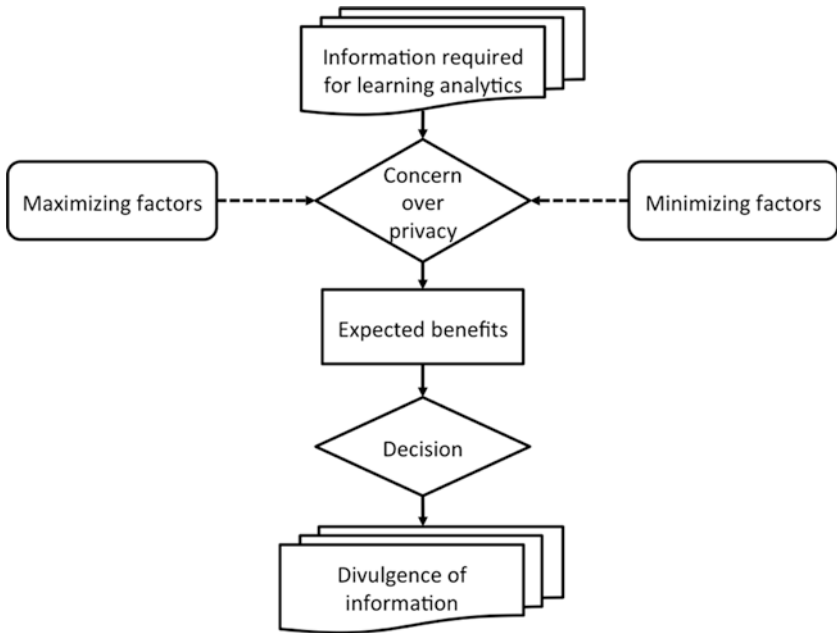


Fig. 4.8 Deliberation process for sharing information for learning analytics systems. (Ifenthaler & Schumacher, 2016a, b)

insights after completion of a learning phase (e.g., study period, semester, final degree), often compared against previously defined reference points or benchmarks. The real-time or formative perspective uses ongoing information for improving processes through direct interventions. The predictive or prescriptive perspective is applied for forecasting the probability of outcomes in order to plan for future strategies and actions (Ifenthaler, 2015).

Each cell of the educational data analytics benefits matrix includes examples to be implemented at different phases of the learning process as well as for different purposes. When choosing a specific benefit of educational data analytics, the teacher, e-Tutor or instructional designer needs to understand:

- (a) who has access?
- (b) to what data?
- (c) to do what?
- (d) for what reason?

In sum, data ownership refers to the possession of, control of, and responsibility for information. Questions surrounding the ownership of data include considerations of who determines what data is collected, who has the right to claim possession over that data, who decides how any analytics applied to the data are created, used and shared, and who is responsible for the effective use of data. Ownership of data also

Table 4.6 Educational data analytics benefits matrix

Stakeholder	Perspective		
	Summative	Real-time/Formative	Predictive/Prescriptive
Governance	Apply cross-institutional comparisons Develop benchmarks Inform policy making Inform quality assurance processes	Increase productivity Apply rapid response to critical incidents Analyse performance	Model impact of organizational decision-making Plan for change management
Institution	Analyse processes Optimize resource allocation Meet institutional standards Compare units across programs and faculties	Monitor processes Evaluate resources Track enrolments Analyse churn	Forecast processes Project attrition Model retention rates Identify gaps
Learning design	Analyse pedagogical models Measure impact of interventions Increase quality of curriculum	Compare learning designs Evaluate learning materials Adjust difficulty levels Provide resources required by learners	Identify learning preferences Plan for future interventions Model difficulty levels Model pathways
Facilitator/ teacher	Compare learners, cohorts and courses Analyse teaching practices Increase quality of teaching	Monitor learning progression Create meaningful interventions Increase interaction Modify content to meet cohorts' needs	Identify learners at risk Forecast learning progression Plan interventions Model success rates
Learner	Understand learning habits Compare learning paths Analyse learning outcomes Track progress towards goals	Receive automated interventions and scaffolds Take assessments including just-in-time feedback	Optimize learning paths Adapt to recommendations Increase engagement Increase success rates

Ifenthaler (2015)

relates to the outsourcing and transfer of data to third parties. A number of scholars point to the lack of legal clarity with respect to data ownership (Corrin et al., 2019). With the absence of legal systems in place to address this issue, the default position has been that the “data belongs to the owner of the data collection tool [who is], typically also the data client and beneficiary” (Greller & Drachler, 2012, p.50).

4.6.3 Data for Instructional Support

Personalised learning is the notion of customising learning resources and activities to fit the interests and needs of individual learners. As with many educational technologies, personalised learning has a long history. However, with the growth of the Internet and ICTs and the advancement of intelligent systems, it is possible to use learning analytics as the basis for automated recommendation engines that drive individualised e-learning. This technology has been promised by several emerging LMSs, but has not yet become a sustainable reality on any scale. However, personalised learning technology can significantly change how instruction occurs and transform the notion of a learning place dramatically (Spector & Ren, 2015). Hence, data is a critical tool that makes this personalised learning possible. When students, parents, and teachers are empowered with access to timely, useful, safeguarded data, there are so many ways to support students on their path to success.

4.6.4 Data for Instructional Support

Corrin et al. (2019, p. 11) provide a well-informed overview on issues of educational data analytics focussing on (a) consent and (b) anonymity.

Consent is referred to as entering into a contract with data subjects in order to obtain their permission for their data to be gathered and analyzed. Consent must be informed in order to be valid; consequently, people should be given clear and transparent information about the purposes for data collection so that they may give informed consent. They should have the ability to opt out of having their data gathered at any time. Consent is not always a simple matter because it is not always a legal requirement, such as when data gathering is judged required for an organization's 'legitimate interests.' (Corrin et al., 2019, p. 32). An example referring to the issues of students not being able to opt out of having their data collected is given in the JISC code of practice (http://repository.jisc.ac.uk/6985/1/Code_of_Practice_for_learning_analytics.pdf).

A more challenging ethical practice is informed consent in the context of learning analytics, which has been critically debated in recent learning analytics research. West et al., 2016 refer to the problematic relationship between 'consent' and 'informed consent' noting that these concepts are often conflated in higher education digital environments. For example, students are frequently asked to agree for their data to be collected, however, the purposes for which the data will be used is hidden or is not communicated clearly (West et al., 2016, p. 914). Cormack (2016) adds that it is not always clear prior to the collection and analysis of data what correlations will emerge or what the impact on individuals will be. This fact seems to make it difficult for educational organizations to communicate clear and transparent information about the use and purposes of data being collected and for of obtaining informed consent.

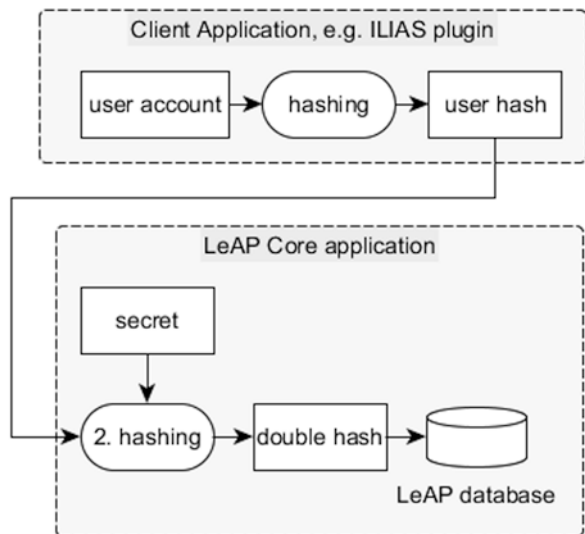
Individuals are given the option of concealing or revealing their identity and any identifying information about themselves through anonymity. Individuals' identities may be de-identified before data is shared or analyzed in the field of learning analytics. Despite the fact that it is widely recognized that institutions should make every attempt to anonymize data, experts have claimed that anonymity cannot always be guaranteed. "Anonymized data can relatively readily be de-anonymized when they are integrated with other information sources," according to Drachler and Greller (2016, p. 94). Anonymity also limits the possible applications of learning analytics because it hinders or precludes meaningful bilateral communication, as well as the capacity for student intervention, feedback, and assistance.

4.6.5 Data Privacy in Productive Systems

One of the main concerns of educational data analytics is the handling of data privacy issues. As almost every learning analytics feature collects and processes user data by default, it is inevitable to consider this topic, particularly in regard of the country's data privacy act. It is even more important when the decision is to work within the running, productive environment of the educational institution as soon as possible.

As shown in the Fig. 4.9, the educational institution decided to use a pseudonymisation in two steps. Whenever a direct touch with students' activities occurs, a 32-bit hash value as an identifier is used. All tracking events and prompting requests use this hash value to communicate with the core application. The core API then takes this hash, enriches it with a secret phrase (a so-called pepper) and hashes

Fig. 4.9 Concept of the encryption of student's identity. (Klasen & Ifenthaler, 2019)



Content	Info	Members	Learning Progress	... ▾
---------	------	---------	-------------------	-------

Within this course, information about the used objects is being tracked. Thereby, only a pseudo-anonym hash of your account is tracked, which allows no direct conclusion to individual students. The data is used in an active research project at the Chair of Business Economics V. Please support this research by allowing us to track the object id and timestamp of your clicks in ILIAS. If you have any questions, please do not hesitate to contact us.
Thank you,

LA-Profile:

LA-profile aktive

LA-profile anonymous

Tracking is anonymous. But some LA-Features are not available.

LA-profile not active

Tracking is disabled. No LA-Features are available.

Fig. 4.10 Individual setting for data collection and analytics. (Klasen & Ifenthaler, 2019)

it again. The doubled hash is then stored within the core's database. As a result, a match with new student generated data can be made to already existing data without being directly traceable back to a specific student by a given date within the database.

Another important issue for implementing educational data analytics in productive systems is the setting of data collection and data analytics functionalities. Figure 4.10 shows an example implemented in a productive Learning Management System allowing the student to change the setting for data collection and data analytics anytime. In addition, the student may request to delete the data stored or download all stored data for self-inspection. Hence, compliance with EU GDPR is given in this case.

Given the examples how to implement data privacy settings in productive systems, think about your own institution and how you may implement similar features in order to be compliant with the EU GDPR.

4.6.6 Case Study: Curtin Challenge I

This case study demonstrates how the analysis of navigation patterns and network graph analysis informs the learning design of self-guided digital learning experiences.

The Curtin Challenge digital learning platform (<http://challenge.curtin.edu.au>) supports individual and team-based learning via gamified, challenge-based, open-ended, inquiry-based learning experiences that integrate automated feedback and rubric-driven assessment capabilities. The Challenge platform is an integral component of Curtin University’s digital learning environment along with the Blackboard learning management system and the edX MOOCs platform. The Challenge development team at Curtin Learning and Teaching are working towards an integrated authoring system across all three digital learning environments with the view of creating reusable and extensible digital learning experiences (Ifenthaler et al., 2018).

Curtin Challenge includes several content modules, for example Leadership, Careers and English Language Challenge. Since 2015, over 2600 badges have been awarded for the completion of a challenge. The design features of each module contain approximately five activities that might include one to three different learner interactions.

Educational analytics data for the presented case study includes 2,753,142 database rows. Overall, 3550 unique users registered and completed a total of 14,587 navigation events within a period of 17 months. Figure 4.11 provides an overview of modules started (M = 3427, SD = 2880) and completed (M = 2903, SD = 2303)

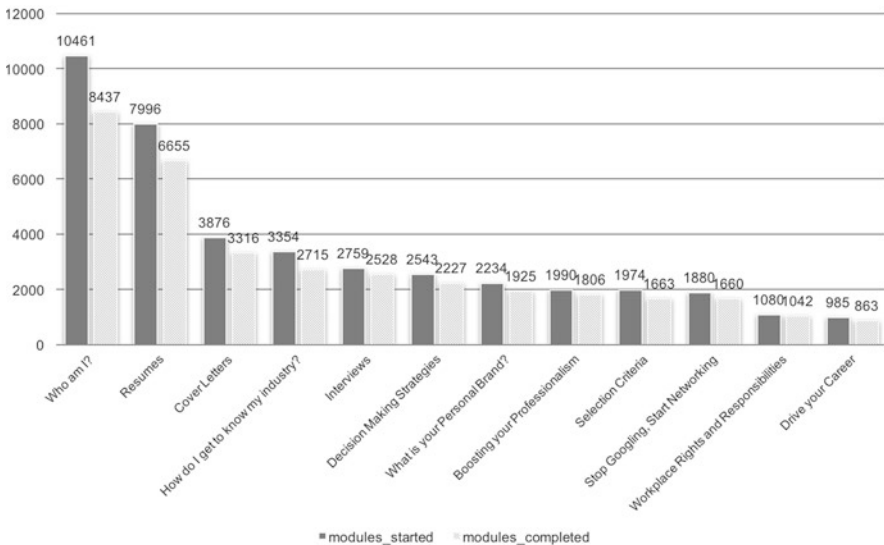


Fig. 4.11 Module completion of Curtin Careers Challenge. (Ifenthaler et al., 2018)

for the Curtin Careers Challenge. The average completion rate for the Curtin Careers Challenge was 87%. The most frequently started module was “Who am I?” (10,461) followed by the module “Resumes” (7996). The module “Workplace Rights and Responsibilities” showed the highest completion rate of 96%, followed by the module “Interviews” (92%).

4.6.7 Case Study: Curtin Challenge II

The network analysis identifies user paths within the learning environment and visualises them as a network graph on the fly. The dashboard visualisations help the learning designer to identify specific patterns of learners and may reveal problematic learning instances. The nodes of the network graph represent individual interactions. The edges of the network graph represent directed paths from one interaction to another. The indicator on the edges represent the frequency of users taking the path from one interaction to another and in parenthesis the percentage of users who took the path. An aggregated network graph shows the overall navigation patterns of all users. A network graph can be created for each individual user, for selected groups of users (e.g., with specific characteristics), or for all users of the learning environment.

The aggregation of all individual network graphs provides detailed insights into the navigation patterns of all users. Figure 4.12 shows the aggregated network graph including paths taken by all 3550 users showing 14,587 navigation events. The five modules are highlighted using different colours.

Provided the case study above, the following questions arise:

- Who is the author of the data presented?
- Who holds ownership of the data presented?
- Who can access the data presented?
- Who can share the data presented (and to what purpose)?

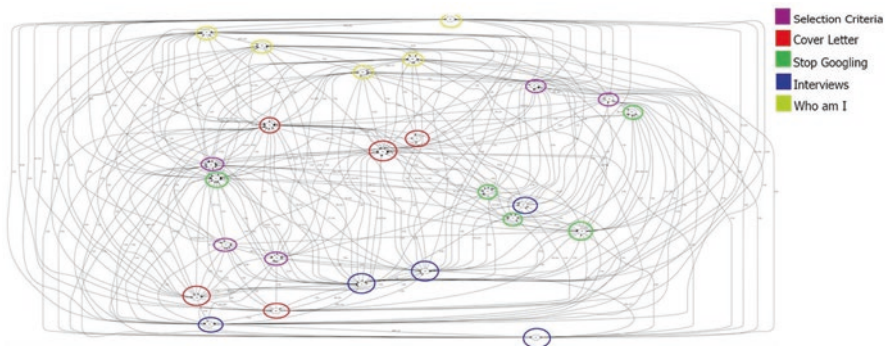


Fig. 4.12 Aggregated network graph. (Ifenthaler et al., 2018)

Questions and Teaching Materials

1. **The educational data analytics benefits matrix includes references to the location of the institution**

- (a) **False**
- (b) **True**

Correct Answer: a.

2. **Examples of analytics benefits for teaching purposes can be related to different perspectives of data processing. Which of the following benefits can be related to predictive analytics?**

- (a) **Conduct cross-institutional comparisons**
- (b) **Track enrolments**
- (c) **Allocate financial resources.**
- (d) **Plan for interventions.**

Correct Answer: d.

3. **Within the deliberation process of sharing information, risk-maximizing factors include**

- (a) **Non-transparency**
- (b) **Positive reputation**
- (c) **Holistic marketing of data**
- (d) **Established data regulations**

Correct Answer: a, c.

4. **ACTIVITY/PRACTICE QUESTION (Reflect on)**

We encourage you to reflect on your teaching experience supported through data. You may reflect on:

- **Are you able to provide your students all the data collected about them when they may request it?**

4.7 Applying and Communicating Educational Data and Analytics Findings

4.7.1 Adaptive Learning Technologies

Adaptive learning and teaching are an alternative to the traditional “one-size-fits-all” approach in the development of digital learning environments. Adaptive learning systems build a model of the goals, preferences and knowledge of each individual learner, and use this model throughout the interaction with the learner, in order to adapt to the needs of that learner (Brusilovsky, 1996). Educational data analytics provides the key element for designing and implementing adaptive learning experiences. In sum, adaptive learning and teaching are referred to as customised learning experiences that focus on the just-in-time need of an individual learning by providing meaningful interventions, feedback or support.

Learning management systems (LMSs) are most commonly used in technology-enhanced learning, typically present identical courses and content for every learner without consideration of the learner’s individual characteristics, situation, and needs (Graf & Kinshuk, 2014). As seen in Massive Open Online Courses, such a one-size-fits-all strategy frequently leads to frustration, learning challenges, and a high drop-out rate (MOOCs).

Adaptive learning technologies aim to solve this problem by allowing learning systems to automatically modify the learning environment and/or learning activities to the learners’ unique situation, traits, and needs, resulting in individualized learning experiences. The system must represent the student and the learning setting in order to create adaptive interventions. This is where data and analytics are required. According to Graf and Kinshuk (2014), adaptive interventions can be based on the following areas:

- Learning styles
- Cognitive abilities
- Affective states
- Context and environment

Other common approaches besides “adaptive learning system” include “personalized learning system” which emphasizes the aim of the system to consider a learner’s individual differences. “Intelligent learning (or tutoring) system” focus on the use of techniques from the field of artificial intelligence to provide learning support.

The phrase “adaptive learning system,” on the other hand, emphasizes a learning system’s ability to provide different courses, learning materials, or learning activities for different learners automatically. Adaptive, personalized, and intelligent learning systems are those that use learning analytics to tailor instruction to learners’ traits and requirements. In their framework of personalization in technology enhanced learning, FitzGerald et al. (2018) characterized learning analytics systems as follows (see Table 4.7):

Table 4.7 Personalization dimensions and learning analytics

Dimension 1: What is being personalised?	Dimension 2: Type of learning	Dimension 3: Personal characteristics of the learner	Dimension 4: Who/what is doing the personalisation	Dimension 5: How is personalisation carried out?	Dimension 6: Impact/beneficiaries
Content, navigation, links and visual design	Formal	Emphasis on prior knowledge (e.g., based on recent assessment scores)	Carried out by computer software, sometimes based on information inputted by the learner e.g. response to a questionnaire	Tends to be cognitive-based personalisation	Learner (most direct impact) but could also be the teacher if savings can be made in terms of time and costs devoted to developing differentiated teaching materials

FitzGerald et al. (2018)

4.7.2 Automated and Semi-Automated Interventions

Closely linked to the demand of new approaches for designing and developing up-to-date adaptive learning environments is the necessity of enhancing the design and delivery of assessment systems and automated computer-based diagnostics (Almond et al., 2002; Ifenthaler et al., 2010). These systems need to accomplish specific requirements, such as:

- (a) adaptability to different subject domains,
- (b) flexibility for experimental and instructional settings,
- (c) management of huge amounts of data,
- (d) rapid analysis of specific data,
- (e) immediate feedback for learners and educators, and
- (f) generation of automated reports of results (Pirnay-Dummer et al., 2012, b).

Recently, promising methodologies have been developed which provide a strong basis for applications in learning and instruction in order to follow up with the demands that come with better theoretical understanding of the phenomena that are a prerequisite or an integral part or go along with the learning process.

Several possible solutions to the assessment and analysis problems of knowledge representations have been discussed (Ifenthaler & Pirnay-Dummer, 2014). Therefore, it is worthwhile to compare the model-based assessment and analysis approaches in order to illustrate their advantages and disadvantages, strengths and limitations (see Table below). Yet, there is no ideal solution for the automated assessment of knowledge. However, within the last five years, strong progress has been made in the development of model-based tools for knowledge assessment. Still, Table 4.8 highlights necessary further development of the available tools, especially for everyday classroom application.

Table 4.8 Comparison of model-based assessment tools

	Pathfinder	ALA-Reader	jMAP	HIMATT	AKOVIA
Description	Converting estimates of relatedness of pairs of concepts into a network representation	Scoring open-ended concept maps and essays	Workbench to map concepts onto a pre-defined structure	Experimental toolset to elicit and analyze graphical or text-based artifacts	Automated researcher tool to analyze existing graphical or text-based artifacts
Measures	Graph theory-based measures Network representation	Graph theory-based measures Scoring algorithm	Adjacency matrix of links	Quantitative structural measures Semantic measures Graphical representation as qualitative measure	Quantitative structural Semantic measures Graphical representation as qualitative measure
Objectivity	Model building process depends on the interpretation by the subjects	Model building process depends on observers	Model building process depends on observers	Automated analysis	Automated analysis
Reliability	Tested ()	Tested (Clariana, 2010)	Not tested	Tested (Ifenthaler et al., 2010)	Tested (Ifenthaler et al., 2010)
Validity	Tested ()	Tested (Clariana, 2010)	Not tested	Tested (Ifenthaler et al., 2010)	Tested (Ifenthaler et al., 2010)
Auto-matization	Partly	Partly	Analysis only	Elicitation & analysis	Model-elicitation (text) & analysis
Strength	Well established research approach	Instant analysis	Off-line availability Instant analysis	Complete experimental setup Server-based for both the elicitation and the analysis	Large datasets Fast analysis Scripting server & online access Data can be assessed by any means outside of the system
Limitations	Connectivity to other learning environments is rather weak	Connectivity to other learning environments is rather weak	Model construction objectivity	Connectivity to other learning environments is rather weak	No elicitation module available

4.7.3 *Instructional Design Principles for Adaptivity*

Leutner (2004) has summarized ten instructional design principles for fostering adaptivity in open learning environments. These principles highlight various instructional elements that can be designed for adaptivity and personalized learning. The principles are:

Adapting ...

- P 1: ... the amount of instruction
- P 2: ... the sequence of instructional units
- P 3: ... the content of information
- P 4: ... the presentation format of information
- P 5: ... task difficulty
- P 6: ... concept definitions
- P 7: ... the system response time
- P 8: ... advice in exploratory learning
- P 9: ... the menu structure of computer software in software training programs
- P10: ... system control versus learner control.

Questions and Teaching Materials

1. **Based on which data features can adaptive interventions be implemented?**
 - (a) **Features such as need for financial study support help to build adaptive interventions**
 - (b) **Features related to the social environment can help to build adaptive interventions**
 - (c) **Features related to cognitive processing can help to build adaptive interventions**
 - (d) **Features such as need for social collaboration help to build adaptive interventions**
 - (e) **Plan for interventions**

Correct Answer: c.

2. **Design principles for adaptive learning environments include ...**
 - (a) **Adapting the speed of algorithms for data processing**
 - (b) **Adapting the presentation format of learning artefacts**
 - (c) **Adapting the task difficulty**
 - (d) **Adapting the sequence of instructional units**

Correct Answer: b, c, d.

3. Informing teaching through data requires realistic technological and personal support

- (a) **False**
- (b) **True**

Correct Answer: b.

4. ACTIVITY/PRACTICE QUESTION (Reflect on)

We encourage you to reflect on your teaching experience supported through data. You may reflect on:

- **When interacting with an adaptive learning system, do you trust the recommendations the system provides for your own learning?**
- **Have you designed or developed an adaptive tool for implementing in your learning environments?**

4.8 Methodologies for Improving Learning and Teaching Processes as Well as Curricula

4.8.1 Creating Interventions in Classroom Settings

Following Ann L. Brown's (1992) article, the effective methodology for improving learning and teaching processes as well as curricula is the combination of creating innovative educational environments and conducting experimental studies of those innovations. The so called design experiment is illustrated in the Fig. 4.13. Brown (1992) explains, that a functional classroom is central to the design experiment before an investigation can be implemented. Hence, classroom life is synergistic: Aspects of it that are often treated independently, such as teacher training, curriculum selection, testing, and so forth actually form part of a systemic whole. Just as it is impossible to change one aspect of the system without creating perturbations in others, so too it is difficult to study any one aspect independently from the whole operating system. Brown (1992) suggests that we must operate always under the constraint that an effective intervention should be able to migrate from our experimental classroom to average classrooms operated by and for average students and teachers, supported by realistic technological and personal support.

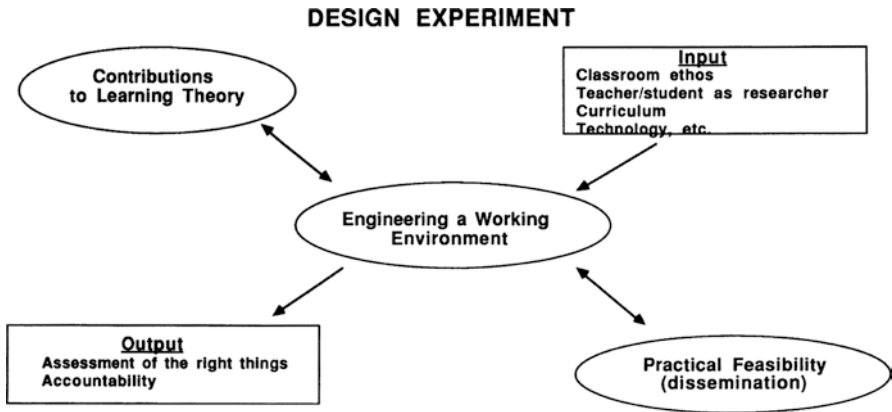


Fig. 4.13 Features of design experiments. (Brown, 1992)

4.8.2 Educational Design Research at a Glance

Educational Design Research (EDR) or Design-Based Research (DBR) – the terms are mostly used synonymously – is a meta-methodology in educational research. It represents a genre of applied research in which the iterative development of solutions to practical and complex educational problems provides the setting for scientific inquiry. The solutions can be educational products, processes, programs, or policies. EDR not only targets solving significant problems educational practitioners face but at the same time seeks to discover new knowledge that can inform the work of others with similar problems. EDR distinguishes itself from other forms of inquiry by attending to both solving problems by putting knowledge to use, and through that process, generating new knowledge (McKenney & Reeves, 2014). EDR projects seek to establish collaborations among researchers and practitioners in real-world settings in order to avoid the widespread theory vs. practice dilemmata. EDR is closely related to research-based educational design as conducted with teaching and learning analytics, yet entails a bit more. Both concepts are shaped by iterative, data-driven processes to reach successive approximations of a desired intervention. However, research-based educational design focuses solely on intervention development, whereas design research strives explicitly to make a ‘transferable’ scientific contribution in form of design principles (McKenney & Reeves, 2014). Major characteristics of Educational Design Research are shown in Table 4.9:

McKenney and Reeves (2014) described a process model for conducting educational design research. Figure 4.14 shows the model which has three main features (Huang et al., 2019):

- **Three core phases in a flexible, interactive structure: analysis, design, and evaluation.**

Table 4.9 Characteristics of EDR/DBR

Characteristics	Explanations
Pragmatic grounded	Design-based research refines both theory and practice The value of theory is appraised by the extent to which principles inform and improve practice
Interactive, iterative, and flexible	Design is theory-driven and grounded in relevant research, theory and practice Design is conducted in real-world settings and the design process is embedded in, and studied through, design-based research
Integrative	Designers are involved in the design processes and work together with participants Processes are iterative cycle of analysis, design, implementation, and redesign Initial plan is usually insufficiently detailed so that designers can make deliberate changes when necessary
Contextual	Mixed research methods are used to maximize the credibility of ongoing research Methods vary during different phases as new needs and issues emerge and the focus of the research evolves Rigor is purposefully maintained and discipline applied appropriate to the development phase
Pragmatic grounded	The research process, research findings, and changes from the initial plan are documented Research results are connected with the design process and the setting The content and depth of generated design principles varies Guidance for applying generated principles is needed

Wang and Hannafin (2005)

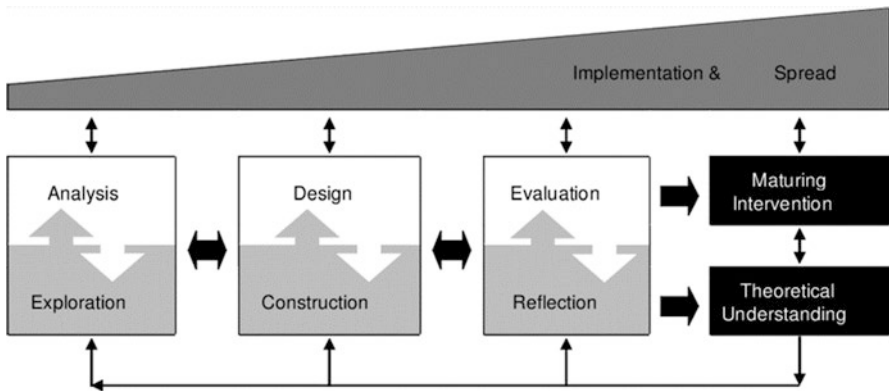


Fig. 4.14 Generic model for conduction Educational Design Research. (McKenney & Reeves, 2014)

- **Dual focus on theory and practice; integrated research and design processes; theoretical and practical outcome**
- **Indications of being use-inspired: planning for implementation and spread; interaction with practice; contextually responsive**

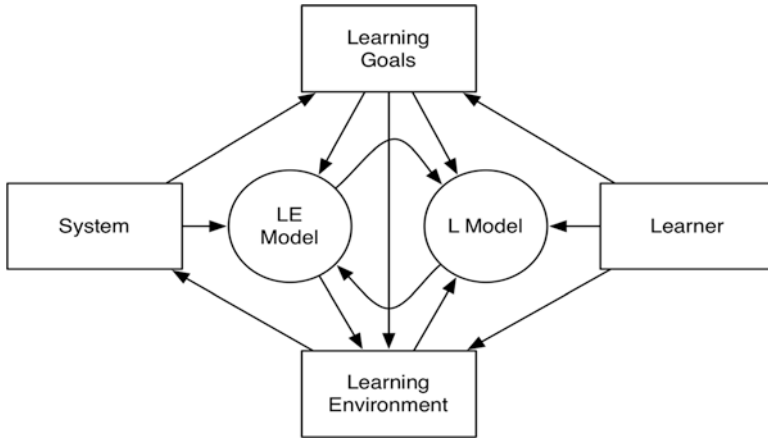


Fig. 4.15 Interdependences of system, learning goals, learner, and learning environment. (Pirnay-Dummer et al., 2012, b)

4.8.3 Designing Model-Based Learning Environments

In model-based and model-oriented learning environments two kinds of models need to be considered: (1) the model of the learning goal, which represents the expertise, set of skills, or, in general, the things to be learned and (2) the model within the learner that is constructed and retained in dependence on the learning environment and on the basis of the current epistemic beliefs active within the learner, i.e., whether and how the learner usually explains parts of the world. We will abbreviate the first type as the LE model (model of the learning environment) and the L model (model of the learner), always assuming that the two types are closely intertwined, especially in well-designed learning environments (Pirnay-Dummer et al., 2012, b).

As shown in Fig. 4.15 above, the educational system (meso- and exo-system) and the learners have different influences on the learning goals at different times. The learning goals constitute the constraints for the learning environment. The learning environment is a manifestation (a derivative) of the LE model. Possible and available learning environments (technology and/or best practices) influence the system by setting the boundaries for what is possible – and decidable as regards educational planning. The learner has influence on the learning environment (as more or less pre-structured by its design). Learning takes place as soon as the LE model and the L model interact. During that time, the learning goal influences and guides the interaction between the two models. LE model-oriented technologies usually focus on the L model while model-centered technologies concentrate more on the LE model. It is our understanding that the two (very similar) approaches will always go hand in hand and influence each other (Pirnay-Dummer et al., 2012, b).

Questions and Teaching Materials

1. **Educational Design Research (EDR) has several characteristic. Which of the following does not belong to EDR?**
- (a) **EDR is well grounded**
 - (b) **EDR is following a single set of statistical procedures**
 - (c) **EDR is related to contextual issues.**
 - (d) **EDR is integrating various methods and approaches**

Correct Answer: b.

2. **The generic model of Educational Design Research includes the following main features ...**
- (a) **core phase management**
 - (b) **core phase analysis**
 - (c) **core phase design**
 - (d) **core phase transformation**
 - (e) **core phase evaluation**

Correct Answer: b, c, e.

3. **Model-based and model-oriented learning environments consider five different models**
- (a) **No**
 - (b) **Yes**

Correct Answer: a.

4. **ACTIVITY/PRACTICE QUESTION (Reflect on)**

We encourage you to reflect on your teaching experience supported through data. You may reflect on:

- **Do you always have sufficient information about the educational system before you design a learning environment?**
- **Do you use evidence from different stakeholders when revising a curriculum?**

4.9 Concluding Self-Assessed Assignment

4.9.1 Introduction

You are requested to complete a concluding self-assessed assignment. This self-assessed assignment is a real-life scenario activity (based on the use case of the instructional designer David), using a rubric across three proficiency levels and an exemplary solution rating. When you have completed this assignment, you will assess it yourself, following the rubric which will list the criteria required and give guidelines for the assessment.

This self-assessed assignment procedure consists of 5 steps:

- **Step 1.** Real life scenario
- **Step 2.** Prepare your answer
- **Step 3.** Exemplary Sample Solution
- **Step 4.** Rubrics for assessing your work
- **Step 5.** Self-evaluate your answer

4.9.2 Step 1. Real Life Scenario

David is an instructional designer. He recently got involved in a newly funded European research project which focusses on the implementation of teaching analytics for a workplace learning environment. The workplace learning environment includes data collection capabilities for students and teachers. All relevant data is securely stored. Data protection rights have been recognised and are fully in place, following EU-GDPR. In addition to the implementation part of the project, all project partners agreed to follow an educational design research approach.

While David started to better understand the key features of teaching analytics and how to conduct educational design research, he knows that you just recently learned about these topics as well. Can you help David to create a strategy for implementing robust teaching analytics capabilities following the **learning analytics profiles** (student, learning, curriculum) approach?

Another challenge, for which David asks for your help, focusses on the benefits of **learning analytics design**, i.e., using available data from the workplace learning environment to provide dynamic perspectives including design decisions during the course of learning. Can you point out three benefits David may use for his project?

4.9.3 Step 2. Prepare Your Answer

The implementation of robust teaching analytics capabilities is crucial for the design, implementation and development of digital-enhanced learning environment. Think about your own educational institution and the current implementation strategy.

1. Describe your implementation strategy and share available cases or evidence as well as guidelines in your educational institution.
2. Provide tips for other learners when reflecting on their own experiences and institutional practice

4.9.4 Step 3. Exemplary Sample Solution

Learning Analytics Profiles

The strategy for implementing robust teaching analytics capabilities in the workplace learning environment require at least the following key issues while following the three profiles (1) student profile, (2) learning profile, (3) curriculum profile:

The student profile includes static and dynamic indicators. Static indicators include gender, age, education level and history, work experience, current employment status, etc. Dynamic indicators include interest, motivation, response to reactive inventories (e.g., learning strategies, achievement motivation, emotions), computer and social media competencies, enrolments, drop-outs, pass/fail rate, academic performance, etc.

The learning profile includes indicators reflecting the current behaviour and performance within the learning environment (e.g., learning management system). Dynamic indicators include trace data such as time specific information (e.g., time spent on learning environment, time per session, time on task, time on assessment). Other indicators of the learning profile include login frequency, task completion rate, assessment activity, assessment outcome, learning material activity (upload/download), discussion activity, support access, ratings of learning material, assessment, support, effort, etc.

The curriculum profile includes indicators reflecting the expected and required performance defined by the learning designer and course creator. Static indicators include course information such as facilitator, title, level of study, and prerequisites. Individual learning outcomes are defined including information about knowledge type (e.g., content, procedural, causal, meta cognitive), sequencing of materials and assessments, as well as required and expected learning activities.

The available data from all data profiles are analysed using pre-defined analytic models allowing summative, real-time, and predictive comparisons. The results of the comparisons are used for specifically designed interventions which are returned to the corresponding profiles. The (semi-) automated interventions include reports, dash-boards, prompts, and scaffolds for teachers. Additionally, teachers can send customised messages for following up with critical incidents (e.g., students at risk, assessments not passed, satisfaction not acceptable, etc.).

Learning Analytics Design

The traditional perspective on learning design is rather static and does not include changes to the learning environment within a short timeframe or while learning processes. In contrast, learning analytics design provides a dynamic perspective

including design decisions on the fly. Especially for learning environments with a large number of learners, the benefits of learning analytics design are obvious:

- Teachers using navigation sequence analysis can identify areas of dropout and change the related materials and instructions accordingly.
- Identifying alignment or misalignment of optimal learning design with actual behaviour of the learners enables the teacher to build adequate interventions when needed.
- The teacher may provide assistance, scaffolds, or feedback to learners off the track.
- The teacher may identify learning materials and activities which need revisions to improve the overall quality of the learning environment.

4.9.5 Step 4. Rubrics for Assessing Your Work

	Unacceptable (1)	Good/solid (3)	Exemplary (5)
Student profile data	It is not clear what data is related to the student profile.	Data related to the student profile are clearly identified. Analytics perspectives are not fully developed.	Data related to the student profile are clearly identified and examples are provided. Analytics perspectives are linked with benefits for teaching.
Learning profile data	It is not clear what data is related to the learning profile.	Data related to the learning profile are clearly identified. Analytics perspectives are not fully developed.	Data related to the learning profile are clearly identified and examples are provided. Analytics perspectives are linked with benefits for teaching.
Curriculum profile data	It is not clear what data is related to the curriculum profile.	Data related to the curriculum profile are clearly identified. Analytics perspectives are not fully developed.	Data related to the curriculum profile are clearly identified and examples are provided. Analytics perspectives are linked with benefits for teaching.
Learning analytics design	The examples do not relate to the basic assumptions of learning analytics design.	The examples are related to teaching practice.	The examples are clearly related to teaching practice and provide reasonable benefits for learning and teaching.

4.9.6 Step 5. Self-Evaluate Your Answer

Now that you have seen the exemplary solution, please rate your own work using the criteria in the rubrics for assessing your work.

Calculate your overall score based on the rubrics for assessing your work.

	Unacceptable (1)	Good/solid (3)	Exemplary (5)
Student profile data			
Learning profile data			
Curriculum profile data			
Learning analytics design			

For each of the criteria in the rubric assign to your solution:

- **1 point** if the option “**Unacceptable**” applies,
- **3 points** if the option “**Good / solid**” applies,
- **5 points** if the option “**Exemplary**” applies.

Then add up the individual points to calculate your overall score.

My overall score is:

Please mark the applicable answer.

- **0–4 points**
- **5–8 points**
- **9–11 points**
- **12–16 points**
- **17–20 points**

References

- Agostinho, S. (2006). The use of visual learning design representation to document and communicate teaching ideas. In *Proceedings of ASCILTE 2006*, Sydney, NSW. Retrieved from http://www.ascilite.org.au/conferences/sydney06/proceeding/pdf_papers/p173.pdf
- Akoka, J., Berti-Equille, L., Boucelma, O., Bouzeghoub, M., Comyn-Wattiau, I., Cosquer, M., Goasdoué-Thion, V., Kedad, Z., Nugier, S., Peralta, V. & Sisaid-Cherfi, S. (2007). A framework for quality evaluation in data integration systems. In *9th International Conference on Enterprise Information Systems (ICEIS)* (p. 10).
- Almond, R. G., Steinberg, L. S., & Mislevy, R. J. (2002). Enhancing the design and delivery of assessment systems: A four process architecture. *Journal of Technology, Learning, and Assessment*, 1(5), 3–63.
- Biggs, J., Kember, D., & Leung, D. Y. P. (2001). The revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology*, 71, 133–149. <https://doi.org/10.1348/000709901158433>
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, 2(2), 141–178.
- Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 6(2–3), 87–129.
- Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal*, 14(2), 1–10. <https://doi.org/10.5334/dsj-2015-002>
- Conole, G. (2008). The role of mediating artefacts in learning design. In L. Lockyer, S. Bennett, S. Agostinho, & B. Harper (Eds.), *Handbook of research on learning design and learning objects: Issues, applications and technologies* (pp. 188–208). Hershey.

- Conole, G. (2013). *Designing for learning in an open world*. Springer.
- Cormack, A. (2016). A data protection framework for learning analytics. *Journal of Learning Analytics*, 3(1), 91–106. <https://doi.org/10.18608/jla.2016.31.6>
- Corrin, L., Kennedy, G., French, S., Buckingham Shum, S., Kitto, K., Pardo, A., & Colvin, C. (2019). *The ethics of learning analytics in Australian higher education. A discussion paper*. Retrieved from <https://melbournecshe.unimelb.edu.au/research/research-projects/edutech/the-ethical-use-of-learning-analytics>.
- D’Mello, S. (2017). Emotional learning analytics. *Handbook of learning analytics*, 115–127. <https://doi.org/10.18608/hla17.010>
- Dalziel, J. (2008). Learning design: Sharing pedagogical know-how. In T. Iiyoshi & M. Kumar (Eds.), *Opening up education: The collective advancement of education through open technology, open content and open knowledge* (pp. 375–388). MIT Press.
- Dobozy, E. (2013). Learning design research: Advancing pedagogies in the digital age. *Educational Media International*, 50(1), 63–76. <https://doi.org/10.1080/09523987.2013.777181>
- Doyle, W. (1985). Recent research on classroom management: Implications for teacher education. *Journal of Teacher Education*, 36(3), 31–35. <https://doi.org/10.1177/002248718503600307>
- Drachler, H., & Greller, W. (2016). Privacy and analytics - it’s a DELICATE issue. A checklist for trusted learning analytics. In *Paper presented at the Sixth International Conference on Learning Analytics & Knowledge*. Edinburgh, UK.
- FitzGerald, E., Kucirkova, N., Jones, A., Cross, S., Ferguson, R., Herodotou, C., Hillaire, G., & Scanlon, E. (2018). Dimensions of personalisation in technology-enhanced learning: A framework and implications for design. *British Journal of Educational Technology*, 49(1), 165–181. <https://doi.org/10.1111/bjet.12534>
- Gertz, M., Özsu, M. T., Saake, G., & Sattler, K.-U. (2004). Report on the Dagstuhl seminar: “data quality on the web”. *ACM SIGMOD Record*, 33, 127–132.
- Goasdoué, V., Nugier, S., Duquennoy, D., & Laboisse, B. (2007). An evaluation framework for data quality tools. In *Proceedings of the 2007 MIT International Conference on Information Quality (ICIQ)* (pp. 280–294).
- Graf, S., & Kinshuk, S. (2014). Adaptive technologies. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 771–779). Springer. https://doi.org/10.1007/978-1-4614-3185-5_62
- Greller, W., & Drachler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, 15(3), 42–57.
- Gustafson, K. L., & Branch, R. M. (2002). *Survey of instructional development models* (4th ed.). ERIC Clearinghouse on Information & Technology.
- Hale, F. (2016). *Learning design: Details of the developing Edinburgh learning design roadmap (ELDeR) at the University of Edinburgh*. University of Edinburgh.
- Huang, R., Spector, J. M., & Yang, J. (2019). Design-based research. In R. Huang, J. M. Spector, & J. Yang (Eds.), *Educational technology. A primer for the 21st century* (pp. 179–188). Springer.
- Ifenthaler, D. (2012). Design of learning environments. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (Vol. 4, pp. 929–931). Springer.
- Ifenthaler, D. (2015). Learning analytics. In J. M. Spector (Ed.), *The SAGE encyclopedia of educational technology* (Vol. 2, pp. 447–451). Sage.
- Ifenthaler, D., & Pirnay-Dummer, P. (2014). Model-based tools for knowledge assessment. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4th ed., pp. 289–301). Springer.
- Ifenthaler, D., & Schumacher, C. (2016a). Student perceptions of privacy principles for learning analytics. *Educational Technology Research and Development*, 64(5), 923–938. <https://doi.org/10.1007/s11423-016-9477-y>
- Ifenthaler, D., & Schumacher, C. (2016b). Connectivism. In S. Danver (Ed.), *The SAGE encyclopedia of online education* (pp. 409–411). Sage.
- Ifenthaler, D., & Tracey, M. W. (2016). Exploring the relationship of ethics and privacy in learning analytics and design: Implications for the field of educational technology. *Educational*

- Technology Research and Development*, 64(5), 877–880. <https://doi.org/10.1007/s11423-016-9480-3>
- Ifenthaler, D., & Widanapathirana, C. (2014). Development and validation of a learning analytics framework: Two case studies using support vector machines. *Technology, Knowledge and Learning*, 19, 221–240. <https://doi.org/10.1007/s10758-014-9226-4>
- Ifenthaler, D., Pirnay-Dummer, P., & Seel, N. M. (Eds.). (2010). *Computer-based diagnostics and systematic analysis of knowledge*. Springer.
- Ifenthaler, D., Gibson, D. C., & Dobozy, E. (2018). Informing learning design through analytics: Applying network graph analysis. *Australasian Journal of Educational Technology*, 34(2), 117–132. <https://doi.org/10.14742/ajet.3767>
- Katal, A., Wazid, M., & Goudar, R. (2013). Big data: Issues, challenges, tools and good practices. In *Procedures of the 2013 Sixth International Conference on Contemporary Computing* (pp. 404–409). IEEE.
- Klasen, D., & Ifenthaler, D. (2019). Implementing learning analytics into existing higher education legacy systems. In D. Ifenthaler, J. Y.-K. Yau, & D.-K. Mah (Eds.), *Utilizing learning analytics to support study success* (pp. 61–72). Springer.
- Koper, R. (2006). Current research in learning design. *Journal of Educational Technology & Society*, 9(1), 13–22. Retrieved from http://www.ifets.info/journals/9_1/3.pdf
- Laranjeiro, N., Soydemir, S. N., & Bernardino, J. (2015). A survey on data quality: Classifying poor data. In *2015 IEEE 21st Pacific rim international symposium on dependable computing (PRDC)* (pp. 179–188). IEEE. <https://doi.org/10.1109/PRDC.2015.41>
- Leutner, D. (2004). Instructional-design principles for adaptivity in open learning environments. In N. M. Seel & S. Dijkstra (Eds.), *Curriculum, plans, and processes in instructional design: International perspectives* (pp. 289–307). Erlbaum.
- 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* (4th ed., pp. 131–140). Springer.
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054.
- Mor, Y., & Craft, B. (2012). Learning design: Reflections upon the current landscape. *Research in Learning Technology*, 20(Supplement), 85–94. <https://doi.org/10.3402/rlt.v20i0.19196>
- Papadakis, S. (2012). Enabling creative blended learning for adults through learning design. In P. Anastasiades (Ed.), *Blended learning environments for adults: Evaluations and frameworks* (pp. 257–273). IGI Global.
- Peralta, V. (2006). *Data freshness and data accuracy: A state of the art*.
- Pirnay-Dummer, P., Ifenthaler, D., & Seel, N. M. (2012). Designing model-based learning environments to support mental models for learning. In D. H. Jonassen & S. Land (Eds.), *Theoretical foundations of learning environments* (2nd ed., pp. 66–94). Routledge.
- Scheerens, J., Luyten, H., Steen, R., & Luyten-de Thouars, Y. (2007). *Review and meta-analyses of school and teaching effectiveness*. Universiteit Twente.
- Seel, N. M., & Ifenthaler, D. (2009). *Online lernen und lehren. [Online learning and teaching]*. Ernst Reinhardt.
- Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510–1529. <https://doi.org/10.1177/0002764213479366>
- Spector, J. M., & Ren, Y. (2015). History of educational technology. In J. M. Spector, D. Ifenthaler, T. E. Johnson, W. C. Savenye, & M. M. Wang (Eds.), *Encyclopedia of educational technology*. Sage.
- Spector, J. M., Johnson, T. E., & Young, P. A. (2014). An editorial on research and development in and with educational technology. *Education Technology Research and Development*, 62(1), 1–12. <https://doi.org/10.1007/s11423-014-9331-z.2>
- Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, 53(4), 5–23.
- West, D., Huijser, H., & Heath, D. (2016). Putting an ethical lens on learning analytics. *Educational Technology Research and Development*, 64(5), 903–922. <https://doi.org/10.1007/s11423-016-9464-3>

Useful Video Resources

- Merrill on Instructional Design [5:41].
 Enhancing educational data quality in heterogeneous learning contexts using Pentaho Data Integration [1:09:06].
 Privacy and learning analytics [03:00].
 GDPR for Schools & Universities | Is Your University GDPR Ready? [04:23].
 GDPR Guidance for Schools [06:20].
 Build your data eco-system [04:36].
 You Need Data to Personalize Learning [03:10].
 Adaptive e-Learning – considerations for lesson development You Need Data to Personalize Learning [02:49].
 Education Professor Wins Grant to Develop Digital Game-Based ‘stealth’ Assessments [02:49].
 What are the differences between the various design-based methods of inquiry? [03:49].

Further Readings

- Batini, C., & Scannapieca, M. (2006). *Data quality. Concepts, methodologies and techniques*. Springer.
- Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer.
- Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11, 87–110.
- Floridi, L. (2013). Information Quality. *Philosophy & Technology*, 26(1), 1–6. <https://doi.org/10.1007/s13347-013-0101-3>
- Fortenbacher, A., Pinkwart, N., & Yun, H. (2017). Learning analytics for sensor-based adaptive learning. *Paper presented at the LAK '17 Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. Vancouver, British Columbia, Canada, 13-03-2017.
- Fuchs, K., Henning, P. A., & Hartmann, M. (2016). INTUITEL and the hypercube model - developing adaptive learning environments. *Journal on Systemics, Cybernetics and Informatics*, 14(3), 7–11.
- Gosper, M., & Ifenthaler, D. (Eds.). (2014). *Curriculum models for the 21st century. Using learning technologies in higher education*. Springer.
- Heath, J. (2014). Contemporary privacy theory contributions to learning analytics. *Journal of Learning Analytics*, 1(1), 140–149.
- Heinich, R. (1970). *Technology and the management of instruction*. Association for Educational Communications and Technology.
- Ifenthaler, D. (2017). Learning analytics design. In L. Lin & J. M. Spector (Eds.), *The sciences of learning and instructional design. Constructive articulation between communities* (pp. 202–211). Routledge.
- Ifenthaler, D. (Ed.). (2018). *Digital workplace learning. Bridging formal and informal learning with digital technologies*. Springer.
- Ifenthaler, D., & Gosper, M. (2014). Guiding the design of lessons by using the MAPLET framework: Matching aims, processes, learner expertise and technologies. *Instructional Science*, 42(4), 561–578. <https://doi.org/10.1007/s11251-013-9301-6>
- Jones, M., & Winne, P. H. (Eds.). (1992). *Adaptive learning environments. Foundations and frontiers*. Springer.
- Koedinger, K. R., & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning sciences to the classroom. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 61–77). Cambridge University Press.

- Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist*, 57(10), 1439–1459. <https://doi.org/10.1177/0002764213479367>
- Lodge, J. M., Alhadad, S. S., Lewis, M. J., & Gašević, D. (2017). Inferring learning from big data: The importance of a transdisciplinary and multidimensional approach. *Technology, Knowledge and Learning*, 22(3), 385–400. <https://doi.org/10.1007/s10758-017-9330-3>
- McKenney, S., & Reeves, T. C. (2018). *Conducting educational design research* (2nd ed.). Routledge.
- Nissenbaum, H. (2011). A contextual approach to privacy online. *Deadalus*, 140(4), 32–48.
- Paramythis, A., & Loidl-Reisinger, S. (2004). Adaptive learning environments and e-learning standards. *Electric Journal of e-Learning*, 2(1), 181–194.
- Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450. <https://doi.org/10.1111/bjet.12152>
- Prinsloo, P., & Slade, S. (2014). Student data privacy and institutional accountability in an age of surveillance. In M. E. Menon, D. G. Terkla, & P. Gibbs (Eds.), *Using data to improve higher education. Research, policy and practice* (pp. 197–214). Sense Publishers.
- Prinsloo, P., & Slade, S. (2015). Student privacy self-management: Implications for learning analytics. Paper presented at the fifth international conference on learning analytics and knowledge.
- Richey, R. C., Klein, J. D., & Tracey, M. W. (2011). *The instructional design knowledge base*. Routledge.
- Scheerens, J. (1990). Review of school and instructional effectiveness research. *School Effectiveness and School Improvement*, 1(1), 61–80.
- Van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (Eds.). (2006). *Educational design research*. Routledge.
- Weippl, E. R., & Min Tjoa, A. (2005). Privacy in e-learning: Anonymity, pseudonyms and authenticated usage. *Interactive Technology and Smart Education*, 2(4), 247–256.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Appendix

Learn2 Analyse Educational Data Literacy Competence Framework

1. Data collection	1.1 Know – understand – be able to obtain, access and gather the appropriate data and/or data sources
	1.2 Know – understand – be able to apply data limitations and quality measures (e.g., validity, reliability, biases in the data, difficulty in collection, accuracy, completeness)
2. Data management	2.1 Know – understand – be able to apply data processing and handling methods (i.e., methods for cleaning and changing data to make it more organized – e.g., duplication, data structuring)
	2.2 Know – understand – be able to apply data description (i.e., metadata)
	2.3 Know – understand – be able to apply data curation processes (i.e., to ensure that data is reliably retrievable for future reuse, and to determine what data is worth saving and for how long)
	2.4 Know – understand – be able to apply the technologies to preserve data (i.e., store, persist, maintain, backup data), e.g., storage mediums/ services, tools, mechanisms
3. Data analysis	3.1 Know – understand – be able to apply data analysis and modeling methods (e.g. application of descriptive statistics, exploratory data analysis, data mining).
	3.2 Know – understand – be able to apply data presentation methods (e.g., pictorial visualisation of the data by using graphs, charts, maps and other data forms like textual or tabular representations)

(continued)

4. Data comprehension & interpretation	4.1 Know – understand – be able to interpret data properties (e.g., measurement error, outliers, discrepancies within data, key take-away points, data dependencies)
	4.2 Know – understand – be able to interpret statistics commonly used with educational data (e.g., randomness, central tendencies, mean, standard deviation, significance)
	4.3 Know – understand – be able to interpret insights from data analysis (e.g., explanations of patterns, identification of hypotheses, connection of multiple observations, underlying trends)
	4.4 Be able to elicit potential implications/links of the data analysis insights to instruction
5. Data application	5.1 Know – understand – be able to use data analysis results to make decisions to revise instruction
	5.2 Be able to evaluate the data-driven revision of instruction
6. Data ethics	6.1 Know – understand – be able to use the informed consent
	6.2 Know – understand – be able to protect individuals’ data privacy, confidentiality, integrity and security
	6.3 Know – understand – be able to apply authorship, ownership, data access (governance), re-negotiation and data-sharing