

Studien zum Physik- und Chemielernen

H. Niedderer, H. Fischler, E. Sumfleth [Hrsg.]

209

Philipp Straube

Modellierung und Erfassung von Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung bei (Lehramts-) Studierenden im Fach Physik



λογος

Studien zum Physik- und Chemielernen

Herausgegeben von Hans Niedderer, Helmut Fischler und Elke Sumfleth

Diese Reihe im Logos-Verlag bietet ein Forum zur Veröffentlichung von wissenschaftlichen Studien zum Physik- und Chemielernen. In ihr werden Ergebnisse empirischer Untersuchungen zum Physik- und Chemielernen dargestellt, z. B. über Schülervorstellungen, Lehr-/Lernprozesse in Schule und Hochschule oder Evaluationsstudien. Von Bedeutung sind auch Arbeiten über Motivation und Einstellungen sowie Interessensgebiete im Physik- und Chemieunterricht. Die Reihe fühlt sich damit der Tradition der empirisch orientierten Forschung in den Fachdidaktiken verpflichtet. Die Herausgeber hoffen, durch die Herausgabe von Studien hoher Qualität einen Beitrag zur weiteren Stabilisierung der physik- und chemiedidaktischen Forschung und zur Förderung eines an den Ergebnissen fachdidaktischer Forschung orientierten Unterrichts in den beiden Fächern zu leisten.

Hans Niedderer

Helmut Fischler

Elke Sumfleth

Philipp Straube

Modellierung und Erfassung von
Kompetenzen naturwissenschaftlicher
Erkenntnisgewinnung bei (Lehramts-)
Studierenden im Fach Physik

Logos Verlag Berlin



Dissertation, Freie Universität Berlin, 2016

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

© Copyright Logos Verlag Berlin GmbH 2016

Alle Rechte vorbehalten.

978-3-8325-4351-8

Logos Verlag Berlin GmbH
Comeniushof, Gubener Str. 47,
D-10243 Berlin
Germany

Tel.: +49 (0)30 / 42 85 10 90

Fax: +49 (0)30 / 42 85 10 92

<http://www.logos-verlag.de>

Modellierung und Erfassung von Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung bei (Lehramts-) Studierenden im Fach Physik

Dissertation

zur Erlangung des Grades
eines Doktors der Naturwissenschaften

am Fachbereich Physik der
Freien Universität Berlin

vorgelegt von
Philipp Frederik Straube, M.Ed.
aus
Berlin

Berlin, 2016

1. Gutachter: Prof. Dr. Volkhard Nordmeier
2. Gutachter: Prof. Dr. Andreas Borowski

Tag der Disputation: 15. September 2016

"The assessment of professional competence is a
difficult and, in many ways, a very
frustrating task."

(Michael T. Kane)

Inhalt

1. Einleitung	5
1.1 Vorarbeiten und Anschlussfähigkeit	8
2. Theoretische Grundlagen.....	9
2.1 Definition von Kompetenz.....	9
2.1.1 Kompetenzmodelle	11
2.2 Professionelle Kompetenz von Lehrerinnen und Lehrern.....	12
2.3.1 Der Einfluss von Subject Matter (Content) Knowledge	15
2.3.2 Syntactic Knowledge (SK).....	16
2.3.3 Einfluss von SK auf den Unterricht.....	19
2.3.4 Zusammenfassung	20
2.4 Erkenntnisgewinnung aus normativer Sicht.....	21
2.5 Scientific Inquiry: Erkenntnisgewinnung aus wissenschaftstheoretischer Perspektive.....	22
2.5.1 Die Induktion	22
2.5.2 Poppers Falsifikationismus.....	23
2.5.3 Kuhn's wissenschaftliche Revolutionen	25
2.5.4 Lakatos Forschungsprogramme.....	27
2.5.5 Zusammenfassung.....	30
2.5.6 Folgerungen für ein Kompetenzmodell	32
2.6 Scientific Reasoning: Handlungsmodell der Erkenntnisgewinnung.....	33
3. Entwicklung eines Kompetenzstrukturmodells wissenschaftlicher Erkenntnisgewinnung	37
3.1 Scientific Reasoning nach Mayer.....	39
3.2 Modellkompetenz.....	40
3.3 Kompetenzstrukturmodell.....	40
3.4 Operationalisierung der Kompetenzen	41
3.4.1 Dimension: Naturwissenschaftliche Untersuchungen.....	41
3.4.2 Modelle Nutzen.....	44
4. Forschungsfragen und Hypothesen.....	47
4.1 Frage zur Dimensionierung des Kompetenzstrukturmodells	48
4.2 Fragen zum Kompetenzstand von Physikstudierenden.....	50
4.3 Fragen zum Kompetenzstand von Studierenden im Fach „Integrierte Naturwissenschaften“.....	53

4.4	Frage zum Kompetenzstand von Studierenden der Sozialkunde	54
5.	Methoden	55
5.1	Verfahren zur Testkonstruktion.....	55
5.1.1	Formate zur Erfassung von Kompetenzen.....	55
5.1.2	Entscheidung für ein Erhebungsverfahren	56
5.1.3	Itemkonstruktion	57
5.2	Statistische Verfahren zur Auswertung	59
5.2.1	Das RASCH-Modell.....	60
5.2.2	Normierung	62
5.2.3	Personenparameterschätzung.....	63
5.2.4	Latentes Regressionsmodell	64
5.2.5	Dimensionsanalyse	65
5.2.6	Differential Item Functioning (DIF).....	65
5.2.7	Mittelwertsvergleiche.....	66
5.2.8	Prüfung von Normalverteilung und Varianzhomogenität	67
5.2.9	Effektstärken	68
6.	Testkonstruktion	69
6.1	Vorgehen bei der Aufgabenkonstruktion.....	69
6.1.1	Prä-Pilotierung und Pilotierung	71
6.2	Konstruktionsbeispiele	72
6.2.1	Aufgabenbeispiel 1: „Hypothese“	72
6.2.2	Aufgabenbeispiel 2: „Planung und Durchführung“	76
6.2.3	Aufgabenbeispiel 3: „Auswertung und Interpretation“	81
6.2.4	Aufgabenbeispiel 4: „Zweck von Modellen“	86
6.3	Beschreibung des fertigen Tests	90
6.3.1	Testgüte	92
7.	Auswertung	95
7.1	Beschreibung der Stichprobe	95
7.2	Fragestellung 1: Dimensionierung des Kompetenzstrukturmodells	97
7.3	Fragestellungen zum Kompetenzstand von Physikstudierenden	100
7.3.1	Vorbereitung der Daten	100
7.3.2	Differential Item Functioning (DIF).....	102
7.3.3	Normierung	103

7.3.4	Bearbeitung der Fragestellungen 2-5.....	104
7.4	Fragestellungen 7 und 8: Kompetenzstand von Studierenden im Fach „Integrierte Naturwissenschaften“	115
7.5	Fragestellung 9: Kompetenzstand von Studierenden der Sozialkunde	119
8.	Interpretation der Ergebnisse.....	123
9.	Diskussion.....	127
10.	Literaturverzeichnis.....	134
11.	Anlagen.....	147
12.	Abbildungsverzeichnis.....	169
13.	Tabellenverzeichnis.....	170
14.	Kurzfassung	172
15.	Abstract	173
16.	Vorveröffentlichungen aus dieser Dissertation.....	174
17.	Lebenslauf	175

1. Einleitung

Die Modellierung und Erfassung von Kompetenzen ist ein Thema, das die deutsche Forschungslandschaft Anfang der 2000er Jahre erfasste. Ausgelöst durch den sogenannten "PISA-Schock", dem verglichen mit anderen Ländern schlechten Abschneiden der deutschen Schülerinnen und Schüler im Jahr 2000, wurde die Einführung nationaler Bildungsstandards beschlossen. Hierdurch sollte eine Einheitlichkeit und damit Vergleichbarkeit der Bildungssysteme der einzelnen Bundesländer hergestellt werden. Ausgehend von diesen Standards wurden in den einzelnen Bundesländern Lehrpläne entwickelt, in deren Fokus die Kompetenzentwicklung der Schülerinnen und Schüler stehen soll (Drieschner, 2008). Wichtiges Merkmal dieser Umstellung ist auch ein Paradigmenwechsel; weg von einer Inputorientierung, hin zu einer Outputorientierung der Lehrpläne (Huber, 2008). Das Institut für Qualitätsentwicklung im Bildungswesen (IQB), das an der Humboldt-Universität zu Berlin angesiedelt ist, sollte fortan die Qualität durch bundesweite Kompetenzerhebungen sichern. In diesem Rahmen ist auch das Projekt "Evaluation der Standards in der Naturwissenschaften" (EsNaS) entstanden, das ein Testinstrument zur Messung naturwissenschaftlicher Kompetenz im Sinne der Bildungsstandards entwickelt hat (Walpulski et al., 2010). Für die Hochschulen wurden im Rahmen des Bologna-Prozesses und der damit verbundenen Angleichung der europäischen Studiensysteme kompetenzorientierte Studienordnungen etabliert (Konferenz der europäischen Hochschulministerinnen und -minister, 2003).

Wenngleich diese Vorgaben auf nationaler Ebene durch die Kultusministerkonferenz (KMK) und durch die Hochschulen im Rahmen der Neufassungen der Studienordnungen implementiert wurden und im Rahmen von Akkreditierungen auch eingefordert werden, so wird dadurch allerdings noch keine Aussage über die tatsächliche Umsetzung innerhalb der universitären Lehre gemacht. Weiterhin bleibt auch der Erfolg universitärer Lehre bei der Kompetenzentwicklung der Studierenden offen, da die Anzahl der Abschlüsse oder der Abschlussnoten sicherlich kein objektives Maß für den Kompetenzstand der Absolvierenden sind.

Die in den letzten Jahren in Deutschland durchgeführten Schulleistungsstudien implizieren ihrerseits auch Schlüsse für die Ausbildung von Naturwissenschaftslehrkräften. So folgert Köller (2014):

„Glaubt man also an die hohe Bedeutung des Professionswissens für gelingenden Unterricht – und alles spricht empirisch dafür – so ergeben sich vor allem Nachhol- bzw. Verbesserungsbedarfe in der nichtgymnasialen Lehreraus- und -weiterbildung. Zu fordern ist hier eine Absicherung hoher fachlicher und fachdidaktischer Kompetenzen in allen Phasen der Lehrerbildung.“

(Köller, 2014, S. 8).

Damit einher geht die Frage, ob die Studierenden die notwendigen „*arbeitsmarktrelevanten Qualifikationen*“ (Sippel, 2009, S. 3) erwerben. Daher wird unter anderem vom Wissenschaftsrat gefordert, die „*[...] zentralen Leistungen in der Lehre zum Gegenstand einer systematischen Erfassung und Bewertung sowie wissenschaftlichen Betrachtung zu machen.*“ (Wissenschaftsrat, 2008, S. 78). Erst auf Basis dieser Forschung kann „*[...] eine gezielte qualitätsorientierte Binnensteuerung und Leistungskontrolle sowie Verbesserung der Studienergebnisse [...]*“ vorgenommen werden (Ebd.). Neben anderen Feedbackformaten ist die objektive Kompetenzmessung ein wesentliches Instrument um einer derartigen Forderung gerecht zu werden (Sippel, 2009). Im Gegensatz zum primären und sekundären Bildungssektor blieben vergleichbare Forschungsvorhaben im tertiären Bildungssektor bislang aber größtenteils aus, so dass hier eine Forschungslücke konstatiert wird (Blömeke, Zlatkin-Troitschanskaia, Kuhn & Fege, 2013; Zlatkin-Troitschanskaia & Kuhn, 2010).

„Betrachtet man das Referenzkriterium einer „large-scale angelegten objektiven Messung von akademisch vermittelten (domänenspezifischen und generischen) Fertigkeiten und Kenntnissen von Studierenden und Hochschulabsolventen unterschiedlicher Fachdisziplinen“, so muss auf nationaler Ebene eine erhebliche Forschungslücke konstatiert werden. Ansätze, die den Kriterien nationaler oder internationaler Vergleichsstudien aus dem Schulbereich (wie z.B. VERA oder PISA) entsprechen, gibt es zum heutigen Zeitpunkt im nationalen Hochschulbereich nicht.“

(Zlatkin-Troitschanskaia & Kuhn, 2010, S. 3)

Die vorliegende Arbeit ist im Rahmen des Kooperationsprojekts "Ko-WADiS"¹ entstanden. In diesem Projekt waren die Fachdidaktiken der Biologie, der Chemie und der Physik beteiligt. Die forschungsleitende Idee war das Postulat einer domänenübergreifenden und fachwissensunabhängigen naturwissenschaftlichen Erkenntnisgewinnungskompetenz. In dieser Arbeit wird dargestellt, wie sich physikalische Erkenntnisgewinnung in diese Annahme einbettet und inwiefern sich bei Studierenden eine Kompetenzentwicklung nachweisen lässt. Zu diesem Zweck wurde ein theoretisches Kompetenzstrukturmodell entwickelt, welches die Bereiche *Untersuchungen* und *Modelle* abdecken soll. Davon ausgehend

¹ Kompetenzmodellierung und -erfassung zum Wissenschaftsverständnis über naturwissenschaftliche Denk- und Arbeitsweisen bei Studierenden (Lehramt) in den drei Naturwissenschaften Biologie, Chemie und Physik

wurden Aufgaben entwickelt, die die entsprechenden Zellen des Kompetenzstrukturmodells valide abbilden sollen. Die Entwicklung dieser Aufgaben, sowie die Auswertung der Ergebnisse werden im zweiten Teil der Arbeit dargestellt. Die Arbeit soll einen Beitrag zur Qualitätssicherung der Lehrerbildung im Bereich der naturwissenschaftlichen Erkenntnisgewinnung leisten.

Der erste Teil dieser Arbeit beschäftigt sich mit dem derzeitigen Forschungsstand zum Professionswissen von Lehrerinnen und Lehrern unter besonderer Berücksichtigung der Erkenntnisgewinnung. Im Anschluss daran wird aus wissenschaftstheoretischer Sicht der Prozess der Erkenntnisgewinnung dargestellt und Schlussfolgerungen für ein Kompetenzstrukturmodell gezogen. Die anschließende Darstellung kognitionswissenschaftlicher Handlungsmodelle zum Prozess der Erkenntnisgewinnung mündet in ein Kompetenzstrukturmodell, welches die Grundlage der weiteren Forschung darstellt. Im zweiten Teil werden die forschungsleitenden Fragestellungen und Hypothesen formuliert. Im dritten Teil werden die methodischen Grundlagen zur Testkonstruktion und zu den statistischen Verfahren vorgestellt. Die Darstellung der Konstruktion des Kompetenztest, sowie die Auswertung der erhobenen Daten erfolgt im vierten Teil. Daran anschließend werden die Ergebnisse interpretiert und diskutiert.

1.1 Vorarbeiten und Anschlussfähigkeit

In den letzten Jahren haben sich zahlreiche Forschungsarbeiten in den Naturwissenschaftsdidaktiken mit dem Lehrerprofessionswissen befasst. Riese und Reinhold (2012) fokussieren in ihren Arbeiten auf das Professionswissen Physiklehramtsstudierender mit verschiedenen Ausbildungszielen. In einer Anschlussstudie untersuchen sie auch explizit die Kompetenzen im Bereich der naturwissenschaftlichen Erkenntnisgewinnung (Woitkowski, 2015; Woitkowski, Riese & Reinhold, 2014). Die Forschung baut auf dem SDDS-Modell von Klahr und Dunbahr (1988) auf.

Im Projekt ProwiN² (Borowski et al., 2010) wird der Zusammenhang der drei Facetten des Lehrerprofessionswissen (*Fachwissen, Fachdidaktisches Wissen und Pädagogisches Wissen*) von Lehrkräften in den drei Naturwissenschaften untersucht. Außerdem soll der Einfluss der drei Facetten auf das Unterrichtshandeln erforscht werden. Erkenntnisgewinnung wird in diesem Projekt auch thematisiert, allerdings unter der Facette des Fachdidaktischen Wissens (Tepner et al., 2012). Damit wird insofern eher auf das Wissen um den Einsatz im Unterricht und dessen Nutzen für die Kompetenzförderung bei Schülerinnen und Schülern fokussiert.

Im Gegensatz zu ProwiN bezieht sich das Projekt KiL³ auf die Entwicklung professioneller Kompetenzen in den Lehramtsstudiengängen (Kröger, Neumann & Petersen, 2014). Auch hier wird auf die gesamte inhaltliche breite des Lehrerprofessionswissens eingegangen.

Die vorliegende Forschungsarbeit ergänzt die vorhandenen Arbeiten dahingehend, dass es die vorhandenen domänenspezifischen Erkenntnisse und Testinstrumente um ein domänenübergreifendes Instrument ergänzt. Durch den Fokus auf die naturwissenschaftlichen Erkenntnismethoden wird es außerdem möglich, diesen Bereich detaillierter zu erfassen, was in anderen Projekten auf Grund ihrer inhaltlichen Breite bislang nicht möglich war. Das im Weiteren genutzte Kompetenzmodell schließt außerdem an die im Schulbereich genutzten Modelle an (z.B. Pant et al., 2013).

² Professionswissen von Lehrkräften in den Naturwissenschaften

³ Messung professioneller Kompetenzen in mathematischen und naturwissenschaftlichen Lehramtsstudiengängen

2. Theoretische Grundlagen

2.1 Definition von Kompetenz

Bei dem Begriff *Kompetenz* handelt es sich um ein Konstrukt, das je nach Definition unterschiedlich konnotiert ist. Dieses Konstrukt ist nicht direkt beobachtbar. Der Kompetenzstand einer Person kann nur indirekt über deren Performanz bei bestimmten Aufgabenstellungen beobachtet werden (Shavelson, 2013). „*Kompetenzen lassen sich nur auf der Basis einer Palette von Einzelbeobachtungen bei unterschiedlichen Aufgaben bzw. in variierenden Situationen abschätzen.*“ (Klieme & Hartig, 2008, S. 24).

In der deutschen Forschungslandschaft hat sich als Definition von Kompetenz diejenige von Weinert (2001a) durchgesetzt (Schecker & Parchmann, 2006). Demnach ist Kompetenz die

„[...] bei Individuen verfügbaren oder von ihnen erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen, motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können.“

(Weinert, 2001a; zitiert nach Weinert, 2001b, S. 27f.).

Diese Definition liegt auch den nationalen Bildungsstandards für den mittleren Schulabschluss zu Grunde (z. B. Kultusministerkonferenz, 2005).

Diese Grundlage eignet sich jedoch nicht für die Large-Scale Kompetenzmessung mittels *Paper & Pencil-Test*, da die Erfassung von motivationalen und volitionalen Bereitschaften bei größeren Proband/inn/engruppen nicht möglich ist (Schecker & Parchmann, 2006). Zumal generell fraglich ist, ob ein/e Proband/in, die/der an einem Tag unmotiviert ist und daher keine Leistung zeigt, weniger kompetent als an einem anderen Tag ist (Hartig, 2008). Dementsprechend müsste dieser Aspekt der Weinert'schen Definition ausgeklammert werden. Ein weiteres Problem ergibt sich mit der Erfassung der „*verantwortungsvollen Nutzung*“. Diese kann durch einen Paper & Pencil-Test ebenso wenig erfasst werden.

Neben dieser Definition existieren andere Definitionen, die sich teilweise nur in Details unterscheiden. So wird in PISA Kompetenz definiert als „*prinzipiell erlernbare, mehr oder minder bereichsspezifische Kenntnisse, Fertigkeiten und Strategien*“ (Baumert, Stanat & Demmrich, 2001). Die Autoren betonen, dass es sich dabei auch um eine Form von Wissen handelt, die sich aber von einem bloßen Faktenwissen unterscheidet. Vielmehr gehe es um ein wirkliches Verstehen des Wissens. Auch Baumert et al. (2001) zählen motivationale und

volitionale Aspekte zu ihrem Kompetenzbegriff, räumen aber auch ein, dass man sich bei der Erfassung von Handlungskompetenzen auf Teilaspekte beschränken muss. Explizit für die universitäre Lehramtsausbildung definiert die Kultusministerkonferenz (2010) Kompetenz als „*Kenntnisse, Fähigkeiten, Fertigkeiten und Einstellungen, über die eine Lehrkraft zur Bewältigung ihrer Aufgaben im Hinblick auf das jeweilige Lehramt verfügen muss*“.

Alternativ bietet sich eine Definition an, die im Rahmen des DFG Schwerpunktprogramms "*Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen*" entwickelt wurde. Demnach sind Kompetenzen „*kontextspezifische kognitive Leistungsdispositionen, die sich funktional auf Situationen und Anforderungen in bestimmten Domänen beziehen*“ und erlern- und vermittelbar sind (Klieme & Leutner, 2006, S. 4). In diesem Zusammenhang weisen die Autoren auf die besondere Bedeutung der Kontextabhängigkeit hin.

Bei der Betrachtung von verschiedenen Kompetenzdefinitionen identifiziert Kane (1992) zwei Komponenten:

“One component is the domain of possible encounters that the professional is expected to manage effectively and the other includes the knowledge, skills, and judgment that the professional is expected to use in managing these encounters.”

(Kane, 1992, S. 166)

Unterschieden werden also zum einen die Umstände, in denen die Kompetenz angewendet werden muss und zum anderen die eigentlichen Fähigkeiten. Diese zwei Komponenten zeigen sich auch in den oben genannten Definitionen, wobei vom Detailreichtum, dem zweiten Teil der Kane'schen Unterscheidung deutlich mehr Raum gewährt wird. Die von Kane (1992) genannten Umstände werden in der Weinertschen Definition als „*variable Situationen*“ berücksichtigt. Bei Pisa wird Kompetenz als „*mehr oder minder bereichsspezifisch*[.]“ und bei Klieme & Leutner als „*kontextspezifisch*“ bezeichnet. In der Definition der KMK (2010) für das Lehramt wird eine Einschränkung auf die bestimmten Umstände nicht vorgenommen. Es wird lediglich darauf verwiesen, dass diese „*auf das jeweilige Lehramt*“ bezogen sind.

Die Frage, wie ähnlich die genannten Kontexte zu fassen sind, ist ein kritischer Punkt bei der Modellierung von Kompetenzen (Hartig, 2008). Hierbei können einerseits zu enge Definitionen dazu führen, dass sämtliche einzelnen Handlungen zu einer Kompetenz werden (z.B. Geräteeinschalt-Kompetenz). Andererseits kann ein zu breites Konstrukt

sowohl zu Problemen bei der Definition, als auch beim Rückschluss auf bestimmte Bildungsmaßnahmen führen (Hartig, 2008). Hartig und Klieme (2006, S. 129) betonen dazu, dass Kompetenzen „in einem gewissen Maß über ähnliche Situationen generalisierbar sind“ und Hartig (2008, S. 21) definiert den Begriff Kontext als *„eine Menge hinreichend ähnlicher realer Situationen, in denen bestimmte, ähnliche Anforderungen bewältigt werden müssen“*.

In Zusammenhang mit der Definition von Kompetenz wird auch immer ihr Unterschied zu generellen kognitiven Fähigkeiten betont (Hartig, 2008; Klieme, Maag-Merki & Hartig, 2007). Der Kompetenzbegriff grenzt sich gleich durch mehrere Faktoren von einer generellen kognitiven Fähigkeit wie Intelligenz ab. Zunächst werden Kompetenzen als erlernbar betrachtet, im Gegensatz zur Intelligenz, die als stabil und vererbbar betrachtet wird (Schweizer 2006, zitiert nach Hartig, 2008). Weiterhin ist Intelligenz losgelöst von inhaltlichem Vorwissen und daher auf neue Probleme anwendbar. Für Kompetenzen gilt dies nur für einen speziellen Kontext (Hartig, 2008).

Aufgrund der Einschränkung auf kognitive Leistungsdispositionen, als auch der klaren Definition der Kontextspezifität wird in dieser Arbeit der Definition von Klieme und Leutner (2006) gefolgt.

2.1.1 Kompetenzmodelle

Die Erfassung von Kompetenzen setzt ein theoretisches Modell voraus, anhand dessen die entsprechenden Messinstrumente konstruiert werden können. Grundsätzlich werden hierbei Kompetenzstrukturmodelle, die sich mit der Dimensionalität von Kompetenzen beschäftigen, sowie Kompetenzniveaumodelle, die verschiedene Ausprägungen einer Kompetenz fokussieren, unterschieden (Hartig & Klieme, 2006). Modelle erster Art strukturieren demnach die zu untersuchende Kompetenz. Die Unterteilung wird durch die zu bewältigenden Anforderungen vorgegeben (Hartig & Klieme, 2006). Teilkompetenzen *„werden v. a. nach den Inhalten der interessierenden Situationen, der relevanten Aufgaben und den zur Lösung dieser Aufgaben zu bewältigenden Anforderungen definiert.“* (Hartig & Klieme, 2006, S. 131). Bei der Frage, wie detailliert man bestimmte Teilkompetenzen unterscheidet, sollte das Modell herangezogen werden, für das es auch fächerübergreifend den größten Konsens gibt (Mayer & Wellnitz, 2014). Ausgehend von der im Kompetenzmodell postulierten Struktur kann mit statistischen Methoden überprüft werden, inwiefern sich diese auch in den empirisch erhobenen Daten widerspiegelt (Hartig & Höhler, 2010). Einerseits werden dadurch die theoretischen Annahmen geprüft, auf denen das Kompe-

tenzmodell fußt, andererseits wird es möglich, bei hohen Korrelationen zwischen einzelnen Teilkompetenzen auf eine einzelne Erfassung zu verzichten und diese als eine Skala zu behandeln (Hartig & Klieme, 2006). Zur Konstruktion von Kompetenzstrukturmodellen eignen sich zwei Verfahren:

- zum einen die *deduktive Methode*, bei der aus der Theorie Handlungsmuster abgeleitet und beschrieben werden, die durch Expertenurteile abgesichert werden.
- zum anderen das *Kriteriumssampling*, bei dem „[...] eine repräsentative Stichprobe von Verhaltensweisen ausgewählt [...]“ und in Testaufgaben überführt wird (Hartig & Jude, 2007, S. 28). Kompetenzniveaumodelle sollen über eine reine normorientierte Beschreibung der Kompetenz von Personen hinausgehen. Innerhalb eines solchen Modells soll genau beschrieben werden, welche Anforderungen Personen mit einem bestimmten Kompetenzstand bewältigen können. Da es nicht praktikabel ist, dies für jeden Kompetenzstand einzeln zu beschreiben, werden mehrere Kompetenzstufen definiert, die einen bestimmten Bereich der Skala abdecken. Für diese Stufen werden dann die Anforderungen beschrieben (Hartig & Klieme, 2006, S. 133).

2.2 Professionelle Kompetenz von Lehrerinnen und Lehrern

Die *professionelle Handlungskompetenz* von Lehrerinnen und Lehrern wird im Wesentlichen durch deklaratives, prozedurales und strategisches Wissen bestimmt (Baumert & Kunter, 2006). Für die Differenzierung der verschiedenen Wissens- (aber auch Könnens-) Bereiche gibt es verschiedene Modelle (für einen Überblick, siehe: Baumert & Kunter, 2006). Shulman (1987b) definiert verschiedene Kategorien von *Teacher Knowledge*, die er als *Knowledge Base* für Lehrerinnen und Lehrer bezeichnet. Neben allgemeinem *Pädagogischen Wissen (PK)* führt Shulman (1986a) drei zentrale Kategorien des Lehrerprofessionswissens an: Das *Subject Matter Content Knowledge (SMK)*⁴, das *Pedagogical Content Knowledge (PCK)* und das *Curricular Knowledge*.

Grossman, Wilson und Shulman (1989) unterteilen das *Subject Matter (Content) Knowledge*, welches die eigentlichen Fachinhalte umfasst, in weitere Unterkategorien:

- das *Content Knowledge (CK)*, was die reinen fachlichen Inhalte umfasst,

⁴ Teilweise auch nur als *Content Knowledge* (Shulman, 1987b) oder *Subject matter Knowledge* (Grossman, Wilson und Shulman, 1989) bezeichnet.

- das *Substantive Knowledge*, welches sich auf die in einer Disziplin geltenden Annahmen und Paradigmen bezieht⁵,
- die *Beliefs about Subject Matter*, die sich auf Vorstellungen über das Fachwissen beziehen,
- sowie das *Syntactic Knowledge (SK)*.

Das *Syntactic Knowledge* befasst sich damit, wie in einer Wissenschaft neues Wissen generiert wird (Grossman et al., 1989, S. 27ff.):

„The teacher need not only understand that something is so; the teacher must further understand why it is so, on what grounds its warrant can be asserted, and under what circumstances our belief in its justification can be weakened and even denied.”

(Shulman, 1986a, S. 9).

Neben einem positiven Einfluss des *Syntactic Knowledge* auf den Unterricht wird auch die Bedeutung für die spätere Aneignung von (neuem) Fachwissen für den Lehrenden selbst diskutiert (Grossman et al., 1989, S. 30).

Der Bereich des *Pedagogical Content Knowledge* umfasst Wissen darüber, wie die entsprechenden Fachinhalte gelehrt werden können. *Curriculum Content Knowledge* betrifft Wissen, über die unterschiedlichen Wege curricular zu unterrichten (Shulman, 1986a, S. 10). In seinen späteren Ausführungen ergänzt er noch drei weitere Kategorien (Shulman, 1987b).

Im deutschsprachigen Raum wird in der Forschung derzeit oftmals auf Shulmans Modell zurückgegriffen, wobei sich die drei Domänen *Fachwissen*, *Fachdidaktisches Wissen* und *Pädagogisches Wissen* durchgesetzt haben (Baumert & Kunter, 2006, S. 482). Auch Baumert und Kunter (2006) beziehen sich auf diese drei Bereiche, ergänzen ihr Modell aber noch um die zwei Aspekte Organisations- und Beratungswissen. In ihrer Begründung des Modells zeigen sie, dass die beiden wesentlichen Aspekte des Modells das Fachwissen und das Fachdidaktische Wissen sind (Baumert & Kunter, 2006, S. 496). Beide Autoren betonen die besondere Bedeutung des Fachwissens als *„notwendige aber nicht hinreichende Bedingung für qualitätvollen Unterricht und Lernfortschritte der Schülerinnen und Schüler [...] Fachwissen ist die Grundlage, auf der fachdidaktische Beweglichkeit entstehen kann“* (Baumert & Kunter, 2006, S. 496; siehe auch van Driel, Berry & Meirink, 2014, S. 861). Sie übersetzen den Bereich Fachwissen mit dem Shulman'schen Aspekt des Subject Matter Content Knowledge (Baumert & Kunter, 2006, S. 482). Viele Studien verwenden aber den Dreiklang PK, PCK

⁵ Vgl. Kuhn (1976b), siehe unten.

und CK und übersetzen die fachwissenschaftliche Dimension damit als *Content Knowledge* (CK) (z.B. Cautet, Borowski & Fischer, 2015; Kirschner, 2013).

Knowledge ist in diesem Zusammenhang aber nicht nur in der direkten deutschen Übersetzung als „(Deklaratives)Wissen“ zu sehen, sondern vor allem auch als prozedurales Wissen. So spricht Shulman (1987b) auch von „*Skills*“ und über „what teachers should know and know how to do“ (Shulman, 1987b, S. 19). Damit wird deutlich, dass auch schon Shulman davon ausging, dass Wissen auch angewendet werden muss und demnach im Sinne einer *Kompetenz* zu verstehen ist. An anderer Stelle spricht er im Zusammenhang mit der Fundierung seiner Forschung ausdrücklich von „*teaching competence*“ (Shulman, 1986a, S. 4). Auch Baumert und Kunter sprechen zwar von Professionswissen, ordnen dieses aber in ihr Modell der *professionellen Handlungskompetenz* ein.

2.3 Forschungsstand des Professionswissens von Lehrkräften

Der Thematisierung naturwissenschaftlicher Erkenntnismethoden im Unterricht wird bislang nicht ausreichend Raum eingeräumt. Als Grund dafür wurden in den letzten Jahren eine unzureichende explizite Thematisierung in der Ausbildung von Lehrerinnen und Lehrern und eine daraus resultierende geringe Vertrautheit mit dem Thema diskutiert (Kircher & Dittmer, 2004, S. 4). Auch Leisner-Bodenthin (2006, S. 107) berichtet davon, dass die Lehrkräfte in ihren Fortbildungen zunächst von der Notwendigkeit eines derartigen Unterrichts⁶ überzeugt werden mussten.

2.3.1 Der Einfluss von Subject Matter (Content) Knowledge

Subject Matter (Content) Knowledge umfasst, wie oben dargestellt, sowohl die fachlichen Inhalte einer Disziplin (CK) als auch die Prozesse, mit denen neues Wissen generiert wird (SK) (Grossman et al., 1989). Einige der im Folgenden zitierten Studien fokussieren explizit das SMK. Andere Studien verwenden den Terminus Fachwissen in Bezug auf CK, womit aber nicht zwangsläufig SK ausgeschlossen wird⁷. Es wird immer gekennzeichnet, welches Konzept in der jeweiligen Studie fokussiert wurde.

Der Zusammenhang zwischen dem Lernerfolg von Schülerinnen und Schülern und dem Inhaltlichen Fachwissen (SMK) wurde in verschiedenen Studien thematisiert. In Ihrem Beitrag zum Subject Matter (Content) Knowledge kommen Grossman et al. (1989, S. 25) zu dem Ergebnis, dass es keinen einfachen Zusammenhang zwischen beiden Punkten gibt. Auch Lipowsky (2006, S. 52) konnte in seinem Überblicksartikel Indizien für einen Zusammenhang nur für die Mathematik finden. Für die Naturwissenschaften konnte er in den von ihm einbezogenen Studien keine eindeutige Befundlage ermitteln. In ihrem großen Überblicksartikel über das Lehrerprofessionswissen findet Abell (2007, S. 1120) Hinweise für einen positiven Zusammenhang, auch wenn sie selber auf die Unterschiedlichkeit der Studien hinweist. Sie vermutet einen über andere Variablen moderierten Zusammenhang. In der neueren Version des Überblicksartikels kommen van Driel et al. (2014, S. 861) zu dem Schluss, dass Fachwissen (SMK) eine notwendige aber nicht hinreichende Voraussetzung von PCK ist. Bezüglich des Fachdidaktischen Wissens von Physiklehrerstudierenden kommen Riese und Reinhold (2012, S. 135) zu einem ähnlichen Ergebnis: Hohes Fachdidaktisches Wissen erfordert ein hohes Fachwissen. Abell (2007, S. 1119) findet in den von ihr durchgesehen Studien deutliche Hinweise auf einen Einfluss

⁶ In diesem Fall zu Modellen.

⁷ Die Begriffe werden vielmehr synonym verwendet (s. o.).

eines ausgeprägten Fachwissens (SMK) auf die Unterrichtsqualität. Auch in aktuelleren Studien zeigt sich, dass profundes Fachwissen: „*is typically associated with more confidence and more interactive and adventurous ways of teaching.*“ (van Driel et al., 2014, S. 854). Cauet et al. (2015) untersuchen in ihrer Studie den Einfluss des Professionswissens von Gymnasiallehrkräften auf den Lernzuwachs ihrer Schülerinnen und Schüler. Auch sie kommen zu dem Ergebnis, dass sich ein Einfluss von CK auf einen kognitiv aktivierenden Unterricht zumindest andeutet. Ein Einfluss auf den Leistungszuwachs der Lernenden konnte nicht festgestellt werden.

Auch Vogelsang (2014) kommt in seiner Analyse des Paderborner Instruments (Riese, 2009) zu dem Ergebnis, dass „*kein direkter Zusammenhang zwischen dem erfassten Professionswissen und der Performanz im Physikunterricht besteht*“ (Vogelsang, 2014, S. 502). Professionswissen kann demnach nicht unmittelbar in Handeln umgesetzt werden. Vielmehr bedürfe es zum Unterrichten mit hoher Performanz einer zusätzlichen Ressource, die möglicherweise im Sinne einer ‚Erfahrung‘ im Vorbereitungsdienst erworben werde.

Bezüglich der konkreten Entwicklung des CK deutet sich an, dass sich dieses erwartungsgemäß im Verlauf des Lehramtsstudiums kontinuierlich entwickelt. Innerhalb des ersten Studienjahres erfolgt der größte Kompetenzzuwachs. Im letzten Studienjahr erfolgt ein Abfall der gemessenen Kompetenzen, was die Autoren der Studie unter anderem durch träges Wissen und Vergessenheitseffekte erklären (Kröger, Neumann & Petersen, 2015, S. 108). Schulrelevantes Fachwissen ist bei im Dienst befindlichen Physiklehrkräften höher als bei Studierenden oder Referendaren. Es verändert sich aber über die Dienstjahre kaum. Dieses Wissen wird demnach vermutlich direkt am Anfang des Lehrerberufs ausgebildet (Borowski, Kirschner, Liedtke & Fischer, 2011). Kirschner (2013, S. 83ff.) findet in ihrer Studie, die das Professionswissen von Physiklehrkräften in der Ausbildung und im Schuldienst untersucht, eine doppelt so große Korrelation von CK und PCK im Vergleich von PK zu CK ($r=.453$ zu $r=.265$). Sie stellt außerdem eine Korrelation mit dem Geschlecht fest: Sowohl in CK, als auch in PCK schneiden männliche Lehrkräfte besser ab.

2.3.2 Syntactic Knowledge (SK)

Während die vorher genannten Studien sich auf Fachwissen (SMK) im Allgemeinen beziehen, so wurde auch das *Syntactic Knowledge* (SK) näher untersucht⁸.

Emereole (2009) kommt in seiner Studie zu dem Ergebnis, dass die untersuchten Lehrkräfte nicht ausreichend mit dem Thema der Erkenntnisprozesse in den Naturwissen-

⁸ Einige der hier genannten Studien wurden auch in Abell (2007) zitiert.

schaften vertraut sind, um die Lerner bei der Entwicklung eines angemessenen Verständnisses der Erkenntnisprozesse zu unterstützen. Auch schienen die untersuchten Lehrkräfte nicht in der Lage zu sein, ihren eigenen Wissenstand in Bezug auf Erkenntnisprozesse angemessen einzuschätzen (Emereole, 2009). In einer qualitativen Studie (Windschitl, 2003, S. 138) mit sechs Lehramtsstudierenden zeigten die Proband/inn/en Probleme bei der Aufstellung von passenden Fragestellungen. Die Generalisierbarkeit scheint aber aufgrund der geringen Stichprobe nicht gegeben.

Bei der Untersuchung von Biologie-Lehrkräften kommt Yip (2001) zu dem Ergebnis, dass die meisten Proband/inn/en die Rolle von Hypothesen („Concepts of Assumptions“) im Erkenntnisprozess nur mangelhaft verstehen würden. Auch zur Beurteilung der Evidenz eines Ergebnisses in Bezug auf ihre Annahmen seien diese nicht fähig. Lawson (2002) kommt in seiner Studie zu dem Ergebnis, dass Lehramtsstudierende der Biologie Probleme im Testen von Hypothesen haben, die nicht direkt zugängliche (beobachtbare) Variablen beinhalten.

Taylor und Dana (2003) untersuchten in einer Fallstudie drei Physiklehrkräfte mit verschiedener Berufserfahrung bezüglich ihrer Fähigkeiten, unkontrollierte Variablen in Experimenten zu identifizieren und selber kontrollierte Experimente zu planen. Sie kommen zu dem Ergebnis, dass eine festgestellte ausgeprägte Fähigkeit im Erkennen unkontrollierter Experimente nicht automatisch auch dazu führt, selbständig richtig kontrollierte Experimente planen zu können. So wurden in der Planung der Experimente bestimmte Variablen ignoriert und nicht weiter kontrolliert. Sie schlussfolgern, dass es zwei Kompetenz-Ebenen bei der Kontrolle von Experimenten gibt. So erfordere die niedrigere Ebene, das Erkennen von unkontrollierten Experimenten, geringere Fähigkeiten als die höhere Ebene, das Planen eigener kontrollierter Experimente.

Roth, McGinn und Bowen (1998) untersuchen in ihrer Studie die Aufbereitung von Daten in Form von Repräsentationen (Graphen, u.ä.). Sie stellen fest, dass sich die Nutzung von Repräsentationen durch Lehramtsstudierende auf demselben Niveau wie von 8-Klässlern bewegt. Die durch andere befragte Forschende genutzten Strategien wurden nicht eingesetzt.

In einer deutschlandweit durchgeführten Studie mit Physiklehramtsstudierenden (Woitkowski, 2015; Woitkowski et al., 2014) konnten das Geschlecht, die Abiturnote und Praktikumszeit als signifikante Einflussfaktoren auf den Kompetenzstand im Bereich Er-

kenntnisgewinnung identifiziert werden⁹. Außerdem wurden keine signifikanten Studiengangseffekte festgestellt. Der Kompetenzstand im Bereich der Erkenntnisgewinnung wird demnach nicht von der Studienart (Lehramt, Fachphysik) beeinflusst (Woitkowski, 2015). Leisner-Bodenthin (2006) stellte in ihrer Studie zur Förderung von Modellkompetenz im Schulunterricht fest, dass bei den Lehrkräften ein Missverständnis auftrat, bezüglich eines *Lernen über Modelle*. Die Lehrerinnen und Lehrer gingen vielmehr von einem *Lernen mit Modellen* aus. Als Ergebnis der Studie folgert sie, dass „Für die Entwicklung domänenspezifischer Modellkompetenz [...] die Kompetenz der Lehrkraft (Fachwissen, Wissenschaftsverständnis, [...], Modellbildung, [...]) ausschlaggebend“ seien. (Leisner-Bodenthin, 2006, S. 106).

In einer Studie mit Niederländischen Naturwissenschaftslehrenden zeigte sich, dass diese im Zweck von Modellen hauptsächlich eine Repräsentations- und Erklärungsfunktion sahen. Nur wenige Lehrende gaben an, dass Modelle auch zum Aufstellen von Voraussagen genutzt werden können (Van Driel, Jan H. & Verloop, 1999, S. 1150). Auch Justi und Gilbert (2003) bescheinigen dem Großteil der von ihnen untersuchten Lehrkräften eine naive Vorstellung von Modellen. Die meisten der Proband/inn/en sahen in Modellen ein Abbild von etwas und ein Idealbild an dem man sich orientieren muss. Nahezu 90% der Proband/inn/en gaben außerdem an, dass Modelle zur Visualisierung oder Erklärung dienen. Dennoch gaben auch 82% der Proband/inn/en an, dass Modelle genutzt werden können, um Voraussagen zu tätigen. Bei einer detaillierteren Analyse nach Fächern zeigte sich, dass gerade Physik- und Chemielehrkräfte ein Bild von Modellen hatten, was einem wissenschaftlichen Bild am nächsten kam. Sämtlichen der untersuchten Physiklehrkräfte war außerdem bewusst, dass Modelle zum Tätigen von Voraussagen genutzt werden können (Justi & Gilbert, 2003). Oh und Oh (2011) fassen die Ergebnisse folgendermaßen zusammen:

„teachers’ perceptions of models were complex and sometimes inconsistent and [...] they adopted different approaches to using models in their instructions, depending on their knowledge, beliefs and experiences“

(Oh & Oh, 2011, S. 1110)

⁹ Die Abiturnote verlor ihre Signifikanz, sobald die Mathematiknote aus der Schule, sowie das Absolvieren einer Abiturprüfung in Physik mit einbezogen wurden. Stattdessen wurden diese beiden Punkte signifikant.

2.3.3 Einfluss von SK auf den Unterricht

Der Zusammenhang mit dem Lernerfolg und der Gestaltung von Unterricht zeigt sich in verschiedenen Studien. So konnte gezeigt werden, dass die Fähigkeit der Lehrkraft in der Variablenkontrolle und das Verständnis der Erkenntnisprozesse positiv mit den gemessenen Leistungen der Lernenden verknüpft sind. Dabei kam der Fähigkeit zur Variablenkontrolle der größte Einfluss zu (Aiello-Nicosia, Sperandeo-Mineo & Valenza, 1984). Auch ein Einfluss auf die Gestaltung von Unterricht wird diskutiert und in Studien genauer untersucht: „*Teachers who do not understand the role played by inquiry in their disciplines are not capable of adequately [...] teaching that subject matter to their students*“ (Grossman et al., 1989, S. 30). Ähnlich argumentieren auch Taylor und Dana (2003, S. 723): „[...] *it is likely that science teachers must first possess appropriate conceptions of scientific evidence themselves before they can help their students develop similar conceptions.*“

Empirisch zeigten sich einige dieser vermuteten Zusammenhänge. So führt Yip (2001) eine unkritische Haltung von Schülerinnen und Schülern gegenüber den Ergebnissen von Experimenten auf ein mangelhaftes Verständnis der Erkenntnisprozesse (s.o.) durch die Lehrenden zurück. Folge auf der einen und Ursache auf der anderen Seite seien ein fast ausschließlicher Einsatz von stark gelenkten Experimenten im Unterricht („Kochrezeptexperimente“).

Windschitl (2003, S. 138f.) konnte in einer Fallstudie mit sechs Lehramtsstudierenden den Einfluss von Erfahrungen mit authentischen Forschungsprojekten während ihrer Ausbildungsphase zeigen. Diese Studierenden setzten eher offene Formen von Inquiry based science Learning (IBSL) in ihrem eigenen Praktikumsunterricht ein. Auch wenn diese Ergebnisse auf Grund der kleinen Stichprobe und anderer Limitationen der Studie nur bedingt aussagekräftig sind, so decken sich diese Ergebnisse mit anderen Studien. So zeigt sich dort, dass Lehrende, die im Rahmen von Fortbildungen selbstständig Forschungsvorhaben durchgeführt haben, eine erhöhte Selbstwirksamkeit in Bezug auf das Durchführen von IBSL basierendem Unterricht aufwiesen (Morrison, 2014).

Windschitl (2003) kommt zu dem Schluss, dass

„of all the factors examined, preservice teachers' use of inquiry in the classroom is most strongly associated with previous research experience. Knowing then how potentially powerful these experiences can be, it suggests that teacher education programs should promote some authentic science research experiences either in conjunction with methods classes or within other areas of the preservice program.“

(Windschitl, 2003, S. 140).

2.3.4 Zusammenfassung

Die Betrachtung der Forschung zum Lehrberufswissen mit besonderer Berücksichtigung des *Syntactic Knowledge* und der Modellbildung gibt insgesamt ein unvollständiges Bild ab. Die Untersuchungen zum SMK zeigen, dass dieses einen Einfluss – zumindest mittelbar – auf die Unterrichtsqualität, insbesondere den Einsatz von offenen Unterrichtsformen und den Lernerfolg der Schülerinnen und Schüler haben kann. Selbiges gilt auch für das *Syntactic Knowledge*.

Die Forschungslage bezüglich des Kompetenzstandes von Lehrkräften im Bereich der *Syntactic Knowledge* gibt ein ähnlich unvollständiges Bild ab. Viele Studien untersuchen nur Teilaspekte, sind aufgrund der geringen Proband/inn/enzahl nur bedingt aussagekräftig und berücksichtigen möglicherweise lokale Besonderheiten, die sich nicht zwangsläufig auch in Deutschland wiederfinden müssen. Die Studien deuten allesamt darauf hin, dass es hier noch Optimierungsbedarf gibt, auch wenn gerade in der Modellkompetenz Physiklehrkräfte eine adäquate Vorstellung zu haben scheinen.

Diese Ergebnisse können aber allenfalls als Hinweise gedeutet werden, da unter dem Aspekt SMK verschiedene Konstrukte zusammengefasst werden und in einigen Studien zudem die Proband/inn/enzahl recht klein war. Es bleibt außerdem fraglich, inwiefern die internationalen Ergebnisse auf Deutschland zu übertragen sind und ob diese Erkenntnisse auch für die Ausbildungsphase in der Hochschule gelten. Betrachtet man den Forschungsstand insgesamt, so zeigt sich, dass eine vollständige fächerübergreifende Modellierung von Kompetenzen im Bereich der Erkenntnisgewinnung bislang noch nicht stattfand.

2.4 Erkenntnisgewinnung aus normativer Sicht

Eine gute Physiklehrerin oder ein guter Physiklehrer muss mit der wissenschaftstheoretischen Fundierung ihres bzw. seines Faches vertraut sein (vgl. die Forderung von Shulman, 1986a, dass Lehrende auch mit den methodischen Grundlagen ihres Fachwissens vertraut sein müssen). Die wissenschaftstheoretische Ausbildung muss daher gerade in der Lehrerinnen- und Lehrerbildung fokussiert werden. Die für die deutschen Schulen verbindlichen Standards der Kultusministerkonferenz für den mittleren Schulabschluss führen als einen von vier grundlegenden Kompetenzbereichen explizit den Bereich „*Erkenntnisgewinnung*“ an (Kultusministerkonferenz, 2005, S. 7). Dieser Punkt umfasst „*Experimentelle und andere Untersuchungsmethoden sowie Modelle nutzen*“ (Kultusministerkonferenz, 2005, S. 7). Konsequenterweise wird in den „*Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung*“ eine Ausbildung in „*[...] den Arbeits- und Erkenntnismethoden der Physik und über Kenntnisse und Fertigkeiten im Experimentieren [...]*“ (Kultusministerkonferenz, 2010, S. 38) gefordert.

Auch die „*Thesen für ein modernes Lehramtsstudium im Fach Physik*“ der Deutschen Physikalischen Gesellschaft fordern die explizite Berücksichtigung naturwissenschaftlicher Erkenntnismethoden im Lehramtsstudium (Deutsche physikalische Gesellschaft, 2006, S. 11).

Weiterhin wird ein Grundverständnis der Erkenntnisprozesse und -methoden als Voraussetzung für ein Leben in einer Welt gesehen, in der Wissenschaften und Informationen eine immer größere Rolle spielen. Scientific Reasoning (s.u.) wird als grundlegende Fähigkeit zum Verständnis und zur Bewertung von wissenschaftlichen Erkenntnissen angesehen (Giere, Bickle & Mauldin, 2006, S. 2ff.). Dies steht im Einklang mit der Forderung nach einer naturwissenschaftlichen Grundbildung (*Scientific Literacy*), die auch Konzepte und Prozesse der Naturwissenschaften umfasst (Bybee, 2002, S. 31).

2.5 Scientific Inquiry: Erkenntnisgewinnung aus wissenschaftstheoretischer Perspektive

Wissenschaftliche Erkenntnisgewinnung, international auch unter dem Label *Scientific Inquiry* bekannt, „refers to characteristics of the scientific enterprise and processes through which scientific knowledge is acquired, including the conventions and ethics involved in the development, acceptance, and utility of scientific knowledge.“ (Schwartz, Lederman & Crawford, 2004, S. 611). Sie beschäftigt sich damit, wie wissenschaftliche Erkenntnisse gewonnen werden und welche Maßstäbe für die Ableitung von Folgerungen aus den beobachteten Tatsachen anlegt werden (vgl. Chalmers, 2007, S. 35).

Diese Fragestellungen werden innerhalb der Wissenschaftstheorie behandelt. Diese hat die Aufgabe, „allgemeine Maßstäbe für theoretische Begründungszusammenhänge einer wissenschaftlichen Praxis zu formulieren“ (Mittelstrass, 1973, S. 1). Dies geschieht nicht nur deskriptiv, sondern soll die Begründungs- und Rechtfertigungsbasis für das wissenschaftliche Handeln bieten. Sie bietet demnach die normativen Grundlagen für das Verhältnis von Theorie und Praxis (Mittelstrass, 1973, S. 2). „Die Frage, wie sich die Absicht, Physik zu treiben, rechtfertigen lässt, bliebe belanglos, wenn nicht zuvor schon die methodischen Mittel zur Begründung von Sätzen und zur Rechtfertigung von Zielen bereitstünden [...]“ (Mittelstrass, 1973, S. 41). Beide Punkte sind eine fundamentale Voraussetzung von Wissenschaft (Mittelstrass, 1973, S. 57).

Im Folgenden werden die wichtigsten Ansätze der Wissenschaftstheorie des 20. Jahrhunderts in Bezug zur Erkenntnisgewinnung dargestellt und diskutiert, welche Ableitungen sich daraus für ein Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung ziehen lassen.

2.5.1 Die Induktion

Ein einfacher, aber nicht unproblematischer Ansatz ist das induktive Schließen, also die Ableitung eines allgemeinen Gesetzes auf Grundlage von Einzelbeobachtungen. So könnte man zum Beispiel aus der Beobachtung von fliegenden Tauben schließen, dass alle Vögel fliegen können. Wir wissen, dass dies nicht der Fall ist. Es wird versucht, dieses Problem durch Regeln zu umgehen, die festlegen sollen, ab wann ein induktiver Schluss möglich ist. So muss die Induktion auf einer Vielzahl von Beobachtungen unter unterschiedlichen Bedingungen stattfinden, wobei keine einzige Beobachtung im Widerspruch zu den anderen stehen darf (Chalmers, 2007, S. 39). Diese Regeln erzeugen aber ihrerseits wieder Probleme, so zum Beispiel die Frage danach, was eine Vielzahl von Beobachtungen und

unterschiedlichen Bedingungen ist. Diese und andere Probleme der Induktion werden ausführlich bei Chalmers (2007, S. 39-45) dargestellt.

2.5.2 Poppers Falsifikationismus

Ein Ansatz, die Probleme der Induktion zu umgehen geht auf Karl Popper (Popper, 1984) zurück. Er nennt seine Methode „*Die Lehre von der deduktiven Methodik der Nachprüfung*“ (Popper, 2010, S. 118). Induktive Schlüsse sind seiner Ansicht nach in der Wissenschaft nicht zulässig. Im Kern von Poppers Lehre steht die Bedingung, dass Theorien falsifizierbar sein müssen. Theorien, die nicht falsifizierbar sind, sind immunisiert und bieten keinen empirischen Nutzen (Menne, 1980, S. 115).

Falsifizierbare wissenschaftliche Theorien müssen sich einer strengen Prüfung unterziehen. Dieser Weg der Wissenschaft beginnt mit einer Hypothese. Diese wird nicht völlig beliebig gebildet, sondern baut auf „[...] *ein Gebäude, eine organisierte Struktur der Wissenschaft, die die allgemein anerkannte Problemsituation dem Wissenschaftler zur Verfügung stellt*“¹⁰ (Popper, 1974, S. 51). Die Wissenschaft bedarf demnach eines gewissen Dogmatismus (Popper, 1974, S. 55). Daraus werden auf „*logisch deduktivem Weg Folgerungen abgeleitet*“, die leicht nachprüfbar sein müssen (Popper, 2010, S. 120). Diese Folgerungen sollten mit dem Ziel formuliert werden, die Hypothese zu falsifizieren. Geht die Prüfung positiv aus, so hat die Theorie die Prüfung vorläufig bestanden. Es gibt keinen Grund, die Theorie zu verwerfen. Geht die Prüfung negativ aus, so betrifft dies auch die Theorie, aus der die Folgerung abgeleitet wurde. Die Theorie muss dann verworfen werden. Der Weg der Wissenschaft besteht demnach aus „*Versuch und Irrtum, Vermutung und Widerlegung*“ (Chalmers, 2007, S. 52). Die geeignetste Theorie besteht alle diese Falsifikationsversuche. Kennzeichen einer guten Theorie ist, dass diese breit formuliert ist und trotzdem allen Falsifikationsversuchen widersteht. Chalmers (2007, S. 56) fasst dies folgendermaßen zusammen: „*Eine sehr gute Theorie ist eine Theorie, die umfassende Aussagen über die Welt macht, die folglich in hohem Maße falsifizierbar ist und die stets einer Falsifizierung standhält.*“

Die Tatsache, dass bei Popper nur Falsifikation wissenschaftlichen Fortschritt bringt, führt unweigerlich zu der Frage, wie wissenschaftliche Erkenntnis überhaupt entstehen kann. Popper selbst führt dazu aus:

¹⁰ Diese deutliche Formulierung stammt aus einer Antwort auf die Kritik von T. Kuhn. In seinem ursprünglichen Werk wird dieser Aspekt weniger stark beleuchtet, wie er selbst einräumt. An dieser Stelle stimmt er Kuhn zu, dass eine Normalwissenschaft (s. u.) existiert.

„Was wir gewöhnlich „wissenschaftliche Erkenntnis“ nennen, ist in der Regel nicht ein Wissen in diesem Sinn, sondern eine Information über die verschiedenen rivalisierenden Hypothesen und über die Weise, in der sie sich in verschiedenen Prüfungen bewährt haben“

(Popper, 2010, S. 74).

Die Prüfung der Hypothese, die im Kern von Poppers Lehre steht, bringt demnach auch bei positivem Ausgang des prüfenden Experimentes eine Erkenntnis, aber nur insoweit, dass man sagen kann, dass das Experiment der Hypothese nicht widersprochen hat. Besteht also die Hypothese ein prüfendes Experiment, so hat sie die Prüfung vorläufig bestanden (Popper, 2010, S. 120). Besteht die Hypothese verschiedene Prüfungen, so hat sich die Hypothese bewährt (Popper, 2010, S. 121). Endgültig beweisen kann man eine Hypothese nicht, was Popper zu der Feststellung führt, dass es kein endgültiges Wissen gibt. Damit gibt es in der Wissenschaft auch keine Beweise (Popper, 2010, S. 74). Zusammenfassend stellt Popper dazu fest¹¹:

„Obwohl wir in der Wissenschaft unser Bestes tun, die Wahrheit aufzufinden, sind wir uns doch des Umstandes wohl bewusst, dass wir nie sicher sein können, ob wir sie gefunden haben. Wir haben in der Vergangenheit aus vielen Enttäuschungen gelernt, dass wir niemals Endgültigkeit erwarten dürfen; und wir haben gelernt, nicht mehr enttäuscht zu sein, wenn unsere wissenschaftlichen Theorien widerlegt werden; denn wir können in den meisten Fällen mit großer Zuverlässigkeit feststellen, welche von zwei vorgelegten Theorien die bessere ist.“

(Popper, 2010, S. 73).

Auch wenn der Falsifikationismus dem Induktivismus überlegen scheint, so krankt auch dieser an einigen Problemen, die deutlich werden, wenn man sich an echte naturwissenschaftliche Probleme heran wagt. Während das oben genannte Beispiel mit den Tauben einfach durch die Beobachtung eines nicht fliegenden Vogels zu falsifizieren ist, führt dies bei komplexeren Theorien zu Problemen. Diese entstehen dadurch, dass – gerade in der modernen Physik – die Beobachtung, die zur Falsifikation einer Hypothese durchgeführt wird, selber auf Theorien beruht, die ihrerseits falsch sein können (Chalmers, 2007, S. 74). So stellte sich die vermeintliche Falsifikation von Einsteins Theorie am CERN, wonach die Lichtgeschwindigkeit die höchste Geschwindigkeit im Universum ist, als Fehler heraus. Die Messung basierte auf der Annahme, dass durch den verwendeten Messaufbau tatsächlich die Geschwindigkeit der Neutrinos gemessen werden kann. Der negative Ausgang eines Experiments führt also nicht zwangsläufig zur Falsifikation der getesteten Hypothese, vielmehr kann einfach eine der Nebenannahmen falsifiziert worden sein (Chalmers, 2007, S. 74). Chalmers (2007, S. 76-77) führt einige Beispiele von Theorien auf, die

¹¹ Teile dieses Abschnitts wurden schon veröffentlicht in: Strelow, Straube und Nordmeier (2014).

genau an diesem Punkt gescheitert wären, so Newtons Gravitationsgesetz, welches trotz der vermeintlichen Falsifikation beibehalten wurden.

„Eine für den Falsifikationisten etwas peinliche historische Tatsache ist, dass gerade jene Theorien, die allgemein zu den besten wissenschaftlichen Theorien gezählt werden, niemals entwickelt worden wären, wenn sich Wissenschaftler strikt an die falsifikationistische Methodologie gehalten hätten“
(Chalmers, 2007, S. 76).

2.5.3 Kuhn's wissenschaftliche Revolutionen

Eine Theorie, die auch die wissenschaftshistorische Entwicklung berücksichtigen soll, geht auf Thomas Kuhn (1976b) zurück. Ihm zufolge gibt es Zeiten der *Normalwissenschaft*, also der *„Entstehung und Fortdauer einer bestimmten Forschungstradition“* (Kuhn, 1976b, S. 26). In diesen Zeiten herrscht ein grundsätzliches wissenschaftliches Paradigma vor, zum Beispiel die Korpuskel-Theorie des Lichts von Newton. Forschende, die einem gemeinsamen Paradigma folgen, unterwerfen sich denselben Regeln und Normen. *„Um als Paradigma angenommen zu werden, muss eine Theorie besser erscheinen als die mit ihr im Wettstreit liegenden, sie braucht aber nicht [...] alle Tatsachen, mit denen sie konfrontiert wird, zu erklären“* (Kuhn, 1976b, S. 32). Forschung zu Zeiten der Normalwissenschaft beschäftigt sich damit, die Natur in Einklang mit dem durch das Paradigma vorgegebenen *„Korsett“* zu bringen (Kuhn, 1976b, S. 38). Diese Forschung lässt sich in drei Klassen einteilen: Zunächst solche Forschungsvorhaben, die die (immer genauere) Bestimmung von Fakten über die Natur zum Ziel haben, so zum Beispiel die Messung von Wellenlängen, Dichten, Leitfähigkeiten. Die zweite Klasse umfasst den Vergleich der Natur mit den Voraussagen der Theorie des Paradigmas. Die dritte Art von Forschungsarbeiten betrifft die Ausschärfung der Paradigmatheorie, wie die Bestimmung von Naturkonstanten oder die Entwicklung von quantitativen Gesetzen (Kuhn, 1976b, S. 39–43). Bei der Bearbeitung dieser Forschungsvorhaben lösen die Forschenden *„Rätsel“*, die instrumenteller, begrifflicher oder mathematischer Form sein können (Kuhn, 1976b, S. 50). Die Forschungsvorhaben sind einerseits so gewählt, dass sie das vorherrschende Paradigma berücksichtigen, andererseits wird durch das Paradigma auch die Wahrnehmung beeinflusst. Auch die Operationen und Messungen werden durch das Paradigma beeinflusst, womit auch die Beobachtungen paradigma-geleitet sind (Kuhn, 1976b, S. 138). Weiterhin entstehen in Zeiten der Normalwissenschaft Regeln, die dem Forschungsprozess zu Grunde liegen. Diese betreffen beispielsweise Methoden, Probleme und Lösungsgrundsätze. Ein Paradigma kann aber auch unabhängig von Regeln existieren. Häufig wird erst in dem Moment, indem das Paradigma

angegriffen wird, über die Regeln diskutiert (Kuhn, 1976b, S. 62). Dieses Angreifen eines Paradigmas – Kuhn nennt es Krise – beginnt mit der Entdeckung einer Anomalie, also *„mit der Erkenntnis, dass die Natur in irgendeiner Weise die von einem Paradigma erzeugten, die normale Wissenschaft beherrschenden Erwartungen nicht erfüllt hat.“* (Kuhn, 1976b, S. 66). Gleichzeitig muss diese Anomalie empirisch und theoretisch anerkannt werden, ihr Anomaliecharakter muss also von der Wissenschaftsgemeinde akzeptiert werden. Daraufhin verändern sich die Paradigmakategorien und –verfahren (Kuhn, 1976b, S. 75). Neben solchen Anomalien kommen auch das Auftauchen neuer Theorien, die die Natur umfassender erklären können, oder theoretische Konflikte in Frage. Letztere werden sichtbar, wenn es zu *„Wucherungen von Versionen einer Theorie“* kommt, die Forschenden also mit allen Mitteln versuchen, das bislang vorherrschende Paradigma mit wenigen Korrekturen auf die Natur anzupassen (Kuhn, 1976b, S. 81–83).

In Folge der Krise kommt es zur wissenschaftlichen Revolution. Damit werden *„jene nicht kumulativen Entwicklungsepisoden angesehen [...], in denen ein älteres Paradigma ganz oder teilweise durch ein nicht mit ihm vereinbares Neues ersetzt wird“* (Kuhn, 1976b, S. 104). Revolutionen müssen aber nicht die komplette Wissenschaft betreffen, selbst wenn sie demselben Paradigma folgen (Kuhn, 1976b, S. 64). Die Frage, welches von zwei konkurrierenden Paradigmen gewählt werden sollte, lässt sich laut Kuhn nicht mit den Kriterien der normalen Wissenschaft lösen, weil beide Paradigmen unter dem Gesichtspunkt ihrer selbst gestellten Werte und Normen diskutiert werden. Dadurch werden die Diskussionen grundsätzlich aneinander vorbei geführt. Da ein Paradigma nicht alle Probleme lösen kann und zwei konkurrierende Paradigmen unterschiedliche Probleme lösen, ist die Frage nach der Bedeutung der gelösten Probleme entscheidend (Kuhn, 1976b, S. 122). Ein vorherrschendes Paradigma wird aber erst abgelöst, wenn ein neues seinen Platz einnehmen kann (Kuhn, 1976b, S. 90). An dieser Stelle betont Kuhn auch seine Ablehnung der Falsifikation. Durch einen Vergleich der Theorie mit der Natur wird eine Theorie nicht falsifiziert. Erst wenn ein anderes Paradigma existiert, das diesen Platz einnehmen kann, kommt es zu einem Paradigmenwechsel.

„[...] das Urteil das zu dieser Entscheidung führt, beinhaltet den Vergleich beider Paradigmata mit der Natur und untereinander.“
(Kuhn, 1976b, S. 90).

Das Ziel der Wissenschaft ist *„Theorie und Tatsachen in bessere Übereinstimmung zu bringen, und diese Tätigkeit kann leicht als ein Prüfen oder als ein Suchen nach Bestätigung oder Falsifikation“*

tion angesehen werden. Hingegen ist ihr Ziel die Lösung eines Rätsels, für dessen bloße Existenz die Gültigkeit des Paradigmas vorausgesetzt werden muss.“

(Kuhn, 1976b, S. 91).

Das Scheitern bei der Lösung eines Rätsels führt demnach nicht zur Falsifikation. Vielmehr wird dies als Anomalie oder als Scheitern des Forschenden angesehen (Chalmers, 2007, S. 92).

„Die normale Forschung ist nicht kritisch, sondern traditions- und paradigmagesteuert. Sie unternimmt keine Versuche, eine grundlegende und paradigmatische Theorie zu falsifizieren; [...] Die „normale Forschung“ versucht nicht, etwas zu entdecken, weder neue Phänomene noch neue Theorien.“

(Andersson, 1988, S. 44).

Ein Kritikpunkt, der Kuhn's Ansichten betrifft, ist der Übergang zwischen zwei Paradigmen. Hierbei wird kritisiert, dass es keine Kriterien gibt, woran festgemacht werden könnte, welches Paradigma vorzuziehen ist. Dies begründet sich zum einen darin, dass es viele verschiedene Faktoren gibt, nach denen eine solche Entscheidung gefällt werden müsste, und zum anderen darin, dass in den verschiedenen Paradigmen unterschiedliche Standards existieren, die dabei herangezogen werden. Die Vertreter der jeweiligen Paradigmen erkennen aber die Standards des jeweils anderen nicht an (Chalmers, 2007, S. 96–97, 101).

2.5.4 Lakatos Forschungsprogramme

Lakatos (1974) versucht, mit seinem Ansatz die wissenschaftsgeschichtliche Analyse Kuhns mit der Logik der Forschung Poppers zu begegnen (Andersson, 1988, S. 5). Als Hauptunterschied zwischen den Theorien von Popper und Kuhn identifiziert Lakatos (1974, S. 91), dass für *„Popper Wissenschaft Revolution in Permanenz‘ und Kritik der Kern des wissenschaftlichen Unternehmens ist“*, während Kuhn eine Revolution als Ausnahme darstellt und Kritik kein Thema sei. Als weiteren Unterschied identifiziert Lakatos (1974, S. 91), dass der Wandel in den Wissenschaften bei Popper rational zu erklären ist, während bei Kuhn der Wechsel zwischen zwei Paradigmen keinen Vernunftregeln folgt. Damit würde die Frage der Annahme eines Paradigmas zu einer Machtfrage, statt dass sie rational zu erklären ist.

Lakatos unterscheidet in seiner Analyse den dogmatischen, den methodologischen, den naiven und den raffinierten Falsifikationismus. Der dogmatische Falsifikationismus geht davon aus, dass alle Theorien nur Vermutungen sind, die nicht bewiesen, jedoch widerlegt werden können. Die Kriterien, die eintreten können, nach denen eine Theorie als falsifi-

ziert gelten muss, müssen im Vorhinein angegeben werden (Lakatos, 1974, S. 94). Lakatos hält diese Form des Falsifikationismus für unhaltbar, unter anderem deswegen, weil sämtliche Theorien beobachtbare Sachverhalte nicht verbieten. Daher kann eine Theorie durch Hilfhypothesen immer weiter gestützt werden (Lakatos, 1974, S. 98-99).

Im methodologischen Falsifikationismus (zu dem Lakatos auch Popper zählt) wird berücksichtigt, dass die experimentellen Aufbauten, mit denen Theorien geprüft werden sollen, ihrerseits auf fehlbaren Theorien beruhen (Lakatos, 1974, S. 104). Eine Theorie darf aber nicht auf Grundlage eines einzelnen Experiments fallen gelassen werden. Vielmehr muss die Falsifikation auf Grundlage mehrerer Wiederholungen basieren und die Nebenannahmen der Falsifikation müssen sich ihrerseits selber bewährt haben (Lakatos, 1974, S. 105). Problematisch ist hierbei, dass die Möglichkeit besteht, dass eine wahre Theorie verworfen wird und eine falsche angenommen, z.B. weil die Nebenannahmen falsch waren oder die Messung auf Irrtümern beruhte (Lakatos, 1974, S. 106). Jedoch kommt nur durch den Mut, diese Entscheidungen zu fällen (auch mit dem Risiko die richtige Theorie verworfen zu haben), wissenschaftlicher Fortschritt zustande (Lakatos, 1974, S. 110).

Eine Weiterentwicklung des Falsifikationismus nennt Lakatos (1974, S. 113) den *raffinierten Falsifikationismus* („sophisticated falsificationism“). Die wesentliche Erweiterung ist, dass es für eine wissenschaftliche Theorie nicht mehr ausreicht, dass sie falsifizierbar ist. Sie muss gleichzeitig auch mehr erklären als ihre vorangegangene Theorie. Einerseits bezieht sich dies auf bereits bekanntes, also der Frage, ob die neue Theorie Umstände eher erklären kann als ihr Vorgänger. Andererseits können in der folgenden empirischen Überprüfung weitere Voraussagen der Theorie überprüft werden. Daraus folgt, dass die konkurrierende Theorie mehr Falsifikationsmöglichkeiten bieten muss, als ihre Vorgängerin. Da es für sämtliche Hypothesen natürlich unendliche Falsifikationsmöglichkeiten gibt, ist dieser Punkt damit zu übersetzen, ob die Hypothese umfassendere Aussagen über die Wirklichkeit macht als eine konkurrierende Vorgängerhypothese. Diese Einschränkung vermeidet ad-hoc Modifikationen, um Hypothesen zu retten, z. B. *X hängt von Z ab, außer wenn Y eintritt* (Chalmers, 2007, S. 63–64). In dieser Spielart des Falsifikationismus gilt eine Theorie nur dann als falsifiziert, wenn neben der Falsifikation der etablierten Theorie zusätzlich eine neue Theorie existiert, die in der Lage ist, neuartige Tatsachen zu erklären, gleichzeitig aber die bewährten Tatsachen aus der alten Theorie weiterhin erklärt (Lakatos, 1974, S. 114).

Lakatos (1974, S. 129) merkt an, dass es sich bei den häufig als Theorie bezeichneten Annahmen (z.B: der Newton'schen Mechanik) nicht um eine einzelne Theorie, sondern vielmehr eine Theorienreihe handeln würde. Diese Reihe verbindet eine gewisse Kontinuität. Diese Kontinuität bezeichnet er als *Forschungsprogramm* und weist selber auf die Ähnlichkeit zur Kuhn'schen Normalwissenschaft hin. Seiner Ansicht nach ist es ein Zeichen reifer Wissenschaft, das Forschen nach Versuch und Irrtum überwunden zu haben und stattdessen in Forschungsprogrammen zu arbeiten (Lakatos, 1974, S. 169). Wichtigster Teil eines Forschungsprogramms ist ein harter Kern von allgemeinen Hypothesen, die die Basis für weitere Forschung innerhalb des Programms bieten (Chalmers, 2007, S. 108). Dazu entstehen Regeln, die den Kern vor Angriffen schützen sollen. Diese geben vor, welche wissenschaftlichen Wege man beschreiten (*positive Heuristik*) und welche man unbeschritten lassen sollte (*negative Heuristik*). Heuristik ist bei Lakatos als „Menge von Regeln oder Hinweisen zur Unterstützung von Entdeckungen bzw. Erfindungen“ (Chalmers, 2007, S. 109) zu verstehen. Im Rahmen der negativen Heuristik entstehen Hypothesen, die die Theorie vor widerlegenden Experimenten schützen sollen (Lakatos, 1974, S. 131). Die positive Heuristik gibt Forschungsstrategien oder Forschungsordnungen vor, die Ansätze dafür liefern, wie die Theorie weiter entwickelt werden kann. Sie

„besteht aus einer partiell artikulierten Reihe von Vorschlägen oder Hinweisen, wie man die ‚widerlegbaren Fassungen‘ des Forschungsprogramms verändern und entwickeln soll und wie der ‚widerlegbare‘ Schutzgürtel [um den Kern der Theorie, Anm. des Verfassers], modifiziert und raffinierter gestaltet werden kann.“

(Lakatos, 1974, S. 131).

Solche Heuristiken können ganz bewusst eingesetzt werden, um neue Theorien zu etablieren, die anderen bewährten Theorien widersprechen, so zum Beispiel Bohr's Atomtheorie, die die Voraussagen Maxwells über das Abstrahlen von Energie per Definition außer Kraft setzte (Lakatos, 1974, S. 137 ff.). „Die Arbeit innerhalb eines einzelnen Programms beinhaltet [also] die Erweiterung und Modifikation des Schutzgürtels durch die Hinzunahme und Benennung verschiedener Hypothesen“ (Chalmers, 2007, S. 111). Die Bedingung für derartige Hypothesen ist aber, dass diese tatsächlich überprüfbar sind (Chalmers, 2007, S. 111)!

Auch in den Forschungsprogrammen kann es aber zu einem Punkt kommen, an denen die Anomalien überhand nehmen, genau wie die zu deren Lösung aufgestellten Adhoc-Hypothesen. Es setzt die Phase der Degeneration¹² eines Programms ein, welche durch

¹² Auch hier benennt Lakatos die Äquivalenz zu den Kuhn'schen Krisen eines Paradigmas

das Auftauchen einer neuen Theorie, die mehr erklärt als die alte, noch beschleunigt werden kann (Lakatos, 1974, S. 149). Lakatos (1974, S. 150) betont aber, dass neue Forschungsprogramme nicht erst dann auftreten (und auftreten sollten) wenn die degenerative Phase eines Programms eingesetzt hat. Vielmehr ist wichtig, dass es einen Wettstreit zwischen konkurrierenden Programmen gibt.

„[...] die Kontinuität der Wissenschaft, die Zähigkeit gewisser Theorien, die Rationalität eines gewissen Ausmaßes an Dogmatismus, [lässt] sich nur dann erklären [...], wenn wir die Wissenschaft als ein Schlachtfeld von Forschungsprogrammen und nicht von isolierten Theorien auffassen.“

(Lakatos, 1974, S. 168).

Hierbei grenzt er sich von Kuhn ab, bei dem ein Paradigma (oder Forschungsprogramm) eine Monopolstellung beanspruche. Dies entspreche weder der Historie noch sei es wünschenswert. Der Wettstreit zwischen konkurrierenden Forschungsprogrammen kann durch Experimente entschieden werden, wenn die Forschungsprogramme unterschiedliche Voraussagen machen. Eine Niederlage eines Programms führt allerdings noch nicht zur Aufgabe des Programms in Gänze. Erst wenn auch in folgenden Versuchen das Programm weiterhin unterliegt, muss es aufgegeben werden (Lakatos, 1974, S. 150). Die Aufgabe eines Programms erfolgt also nicht auf Grundlage eines entscheidenden Experiments (*experimentum crucis*), sondern dieses Experiment wird erst rückblickend zu diesem ernannt (Lakatos, 1974, S. 157, 167).

Auch Lakatos Position wird kritisiert. So liefere die Theorie keinen Ansatz für die Bewertung aktueller Forschungsprogramme. Es bleibe weiterhin nützlich, auch an degenerierenden Forschungsprogrammen festzuhalten, da diese ein Comeback erleben könnten. Erst in der Rückschau sei es möglich, die Wissenschaftsgeschichte im Sinne Lakatos zu interpretieren. Ein weiterer Punkt, der kritisiert wird, ist, dass nicht bei allen Forschungsprogrammen tatsächlich ein *harter Kern* zu identifizieren sei. Vielmehr würde es Programme geben, in denen Forschende zur Lösung von Problemen so weit gingen, dass harte Kerne nicht mehr auszumachen seien (Chalmers, 2007, S. 118 ff.).

2.5.5 Zusammenfassung

Mit Popper, Kuhn und Lakatos wurden die Ansätze von dreien der vier einflussreichsten Wissenschaftstheoretiker des 20. Jahrhunderts vorgestellt. Wie dargestellt wurde, haben alle drei Ansätze auch Kritik erfahren. Teilweise wird aber der Ansatz von Lakatos, der sowohl in der Tradition Poppers steht, aber auch Elemente von Kuhn enthält, als der am weitesten fortgeschrittene bezeichnet (z.B. Worral, 1980). Der bislang unerwähnte vierte

Wissenschaftstheoretiker ist Paul Feyerabend, der mit seinem Buch „Wider den Methodenzwang“ (1983), die Ansicht vertrat, dass „*Wissenschaft keinerlei Merkmale aufweise, die sie notwendiger Weise anderen Erkenntnisformen überlegen mache.*“ (Chalmers, 2007, S. 122) und damit sämtliche Ansätze von Popper, Kuhn und Lakatos als falsch beurteilt. Das einzige Prinzip, das der Wissenschaft zu Grunde liegen könnte, sei der Ansatz, dass alles gehen würde ('anything goes', Feyerabend, 1983, S. 32). Feyerabend kommt zu dem Schluss, dass sämtliche Versuche, die Wissenschaft von anderen Erkenntnisformen abzugrenzen, gescheitert seien. Er sieht gar die Gefahr der Repression der Menschheit durch die Wissenschaft (Chalmers, 2007, S. 126). Die dargestellte Kritik wird von heutigen Wissenschaftsphilosophen teilweise bejaht, jedoch nur soweit, dass es unrealistisch erscheint, dass es eine universelle Methode von Wissenschaft gibt. So vertritt Chalmers (2007, S. 132) die Ansicht, dass es zwar keine universelle Methode gebe, sehr wohl aber „*historisch kontingente Methoden und Standards*“ die erfolgreiche Wissenschaftsdisziplinen enthalten (Chalmers, 2007, S. 132).

Die historische Entwicklung wissenschaftlicher Erkenntnis wird eher in den Ansätzen von Lakatos oder Kuhn beschrieben. Während ein Paradigmenwechsel bei Popper lediglich zum Austausch einiger Sätze führt, so berücksichtigt Kuhn auch „*eine Änderung der Wahrnehmung von Realität und eine Veränderung der Maßstäbe, die an die Bewertung einer Theorie angelegt werden.*“ (Chalmers, 2007, S. 100). Problematisch an Kuhns Theorie sei aber, dass es nach Kuhn keine neutralen Regeln gibt, nach denen entschieden werden kann, ob eine Theorie fortschrittlich ist (Chalmers, 2007, S. 101). Der Wechsel zwischen zwei Paradigmen sei vielmehr eine Sache der Mode und nicht rational zu beschreiben (Lakatos, 1974, S. 172). Lakatos Theorie der Forschungsprogramme vermeidet dieses Problem. Der Wechsel zwischen zwei Forschungsprogrammen lässt sich durch die Überlegenheit eines Programms klar rational begründen (s.o.) (Chalmers, 2007, S. 113).

In der Rolle des theoretischen Rahmens liegt ein Hauptunterschied zwischen den drei genannten Ansätzen: Bei Popper als Gebäude, bei Kuhn als Paradigma und bei Lakatos als Forschungsprogramm bezeichnet. Popper selbst scheint mit den Lakatos'schen Forschungsprogrammen mehr zu sympathisieren als mit der Normalwissenschaft mit ihrem Paradigma. Forschungsprogramme können auch parallel existieren, während das vorherrschende Paradigma über allen Konkurrenten stehen würde (Popper, 1974, S. 54). Die normalen Forschenden, wie Kuhn sie beschreibt, existieren zwar, sie seien aber „*bemitleidenswert*“, weil ihnen eine kritische Grundhaltung fehle (Popper, 1974, S. 53). Es wird aber

auch für die Normalwissenschaft argumentiert, weil „eine Disziplin, in der Grundsätzliches immer wieder hinterfragt wird, [...] kaum bemerkenswerte Fortschritte machen [wird], weil Prinzipien nicht lange genug verschont bleiben, um fachwissenschaftliche Arbeit zu leisten“ (Chalmers, 2007, S. 99).

2.5.6 Folgerungen für ein Kompetenzmodell

Kuhn (1974a, S. 6) merkt an, dass die „Fachleute“ für die Zeit der Normalwissenschaft ausgebildet werden und nicht für die Zeiten der Krisen oder Degeneration. Die Übersetzung der dargestellten wissenschaftstheoretischen Ansätze in ein Kompetenzmodell kann also auf eine Berücksichtigung von Krisen verzichten. Viele der zukünftigen Wissenschaftlerinnen und Wissenschaftler werden möglicherweise nie den Wechsel eines Forschungsprogrammes oder eines Paradigmas erleben. Damit entfällt auch die unter Wissenschaftsphilosophen immer noch umstrittene Frage danach, wann eine Theorie abgelehnt werden muss. Guten Gewissens kann man den Forschungsprozess auf das reduzieren, was Kuhn als Rätsellösen bezeichnet oder Lakatos als Ausschärfung der Theorie. Der Wechsel von Forschungsprogrammen (oder -paradigmen) findet ohnehin statt. Ein Kompetenzmodell der Erkenntnisgewinnung muss also vor allem auf die alltägliche Arbeit von Forscherinnen und Forschern fokussieren und ihre alltäglichen Arbeits- und Denkweisen berücksichtigen. Innerhalb eines Forschungsprogramms muss diesem Programm gefolgt werden. Hypothesen müssen also im Sinne Lakatos erlaubt sein, oder, um es mit Popper (1984, S. 26) zu sagen: „Einmal aufgestellte und bewährte Hypothesen dürfen nicht ‚ohne Grund‘ fallengelassen werden.“

Die genauere Betrachtung naturwissenschaftlicher Forschung, wie dies an einigen Beispielen bereits oben geschah, offenbart noch einen anderen Aspekt: So entziehen sich in der Physik (aber auch der Chemie und Biologie) viele der untersuchten Phänomene unserer direkten Wahrnehmung. Die erwähnte Korpuskel-Theorie des Lichts oder das Bohr'sche Atommodell sind nur zwei Beispiele, bei denen dies der Fall ist. An die Stelle der direkten Untersuchung an einem Gegenstand kann dann das Denken in Modellen treten. Aus theoretischen Überlegungen und Modellvorstellungen heraus können Vorhersagen über die Natur getroffen werden. Die oben beschriebenen Erkenntnisprozesse können demnach auch auf Modelle angewendet werden. Sie können so zur Bildung von Hypothesen genutzt werden, die anschließend in der Realität getestet werden (vgl. Bleichroth et al., 1999, S. 27).

2.6 Scientific Reasoning: Handlungsmodell der Erkenntnisgewinnung

Neben wissenschaftstheoretischen Überlegungen wurden auch kognitionswissenschaftliche Handlungsmodelle zum Prozess der Erkenntnisgewinnung entwickelt. Diese zielen stärker auf die mentalen Prozesse bei der Erkenntnisgewinnung ab und geben sehr viel konkretere Ablaufschemen für mögliche Handlungsschritte oder -komponenten beim Erkenntnisprozess an.

Der Prozess bzw. das Handeln von Wissenschaftlerinnen oder Wissenschaftlern bei der wissenschaftlichen Erkenntnisgewinnung wird international unter den Begriffen *Scientific Thinking* (Kuhn, Amsel & O'Loughlin, 1988) oder **Scientific Reasoning** (Klahr, 2000; Koslowski, 1996) zusammengefasst (Mayer, 2007). Scientific Thinking ist nach D. Kuhn et al. (1988, S. 219) ein Zusammenspiel von Theorie und Evidenz. Kompetente Forschende sind nach ihrer Definition Personen, die bewusst mit einer Theorie arbeiten, anstatt sie nur anzuwenden. Dabei müssen sie in der Lage sein, Theorie und Evidenz zu unterscheiden und separat mit ihnen umzugehen. Der korrekte Umgang mit der Variablenkontrollstrategie ist ein wesentliches Element (Kuhn et al., 1988, S. 15–17). Die von Kuhn mit diesem Ansatz durchgeführten Studien legen nahe, dass die hauptsächliche Kompetenzentwicklung vom Jugend- zum Erwachsenenalter stattfindet. Im Erwachsenenalter bleiben die Kompetenzen aber weit unter einem optimalen Level (Kuhn et al., 1988, S. 219 f.).

Koslowski (1996, S. 6) baut auf den Studien von Kuhn et al. (1988) auf, merkt aber an, dass wissenschaftliche Erkenntnisgewinnung nicht vom Vorwissen loslösbar ist. Die Entscheidung, ob zwischen zwei korrelierten Variablen tatsächlich ein Wirkmechanismus bestehe, lasse sich nur durch das Wissen über eine entsprechende Theorie entscheiden, die einen Zusammenhang vorraussagt: „[...] *the principles of scientific inquiry are used in conjunction with (not independently of) knowledge about the world.*“ (Koslowski, 1996, S. 13) Der Erfolg wissenschaftlicher Erkenntnismethoden hängt demnach auch von der Genauigkeit und der Richtigkeit der Theorien ab. Dieser Ansatz des *Scientific Reasoning* erfordert von den Proband/inn/en, dass ihre wissenschaftliche Bewertung auf theoretisch korrekten Annahmen beruht, aber auch, dass ein methodologisch legitimer Weg gewählt wurde (Koslowski, 1996, S. 15).

Der Prozess des Scientific Reasoning ist nach Klahr (2000, S. 14f.) ein Problemlöseprozess, der zwei Problemräume umfasst, den Hypothesenraum und den Experimentalraum

(SDDS = Scientific discovery as Dual-search, Klahr & Dunbahr, 1988, S. 32ff.). Grob lässt sich der Ablauf in drei Schritte unterteilen: die *Suche im Hypothesenraum*, die *Suche im Experimentierraum* und die *Analyse von Evidenz*: „[...] Science [involves] [...] the generation of hypotheses, the formulation and execution of experiments, and the revision of hypotheses on the basis of empirical evidence“ (Klahr, 2000, S. 15). Bei der Hypothesensuche muss eine Hypothese zur Erklärung eines Phänomens gebildet werden. Diese kann entweder auf Grundlage vorhandenen Wissens oder anhand experimenteller oder beobachteter Ergebnisse gebildet werden. Im Experimentierraum muss dann ein Experiment gefunden werden, was die entsprechende Hypothese prüfen kann. Bei der Analyse von Evidenz muss entschieden werden, ob auf Grundlage der erhobenen Daten eine Hypothese verworfen werden muss oder ob sie durch das Experiment gestützt wurde (Klahr, 2000, S. 30–39).¹³

Im Mittelpunkt des Modells der „concepts of evidence“ (Gott & Duggan, 1995) steht auch die Lösung eines Problems. Innerhalb dieses Konzepts liegt der Hauptfokus auf dem Entwurf eines Experiments, der Messung, dem Umgang mit Daten und der anschließenden kritischen Analyse (Evaluation) der erhaltenen Evidenz. Diese Schritte sind aber nicht unbedingt in zeitlicher Reihenfolge zu verstehen. Sie tragen insgesamt aber zur Absicherung erhaltener Evidenz mittels Daten bei (Gott & Duggan, 1995, S. 30).

Fischer et al. (2014, S. 39) unterscheiden acht Komponenten von scientific reasoning:

1. *Problem identification*: hierbei wird die Situation analysiert und das Problem identifiziert.
2. *Questioning*: Aufbauend auf dem identifizierten Problem wird eine handlungsleitende Fragestellung formuliert.
3. *Hypothesis generation*: Dieser Schritt umfasst die Formulierung einer möglichen Antwort auf die Fragestellung. Diese muss theoriegeleitet entwickelt werden und wissenschaftlichen Standards entsprechen.
4. *Construction and redesign of artefacts*: Unter diesen Punkt fällt die kritische Analyse eventueller Prototypen – sowohl Objekte als auch mathematische Systeme.
5. *Evidence generation*: Hierunter fällt die Überprüfung der Hypothese z.B. durch Experimente, bei denen die abhängige Variable beobachtet, die unabhängige verändert und sämtliche anderen beteiligten Variablen kontrolliert werden.

¹³ Das vollständige SDDS-Modell umfasst weitere Unterkategorien, die an dieser Stelle nicht weiter thematisiert werden sollen. Es wird bei Klahr (2000, S. 30–39) umfassend dargestellt.

6. *Evidence evaluation*: Hierbei wird entschieden, inwiefern die Evidenz die Hypothese stützt oder widerlegt.
7. *Drawing conclusions*: Dieser Punkt umfasst die kritische Analyse der erhaltenen Ergebnisse und den Einbezug anderer Experimente, um zu einem Ergebnis zu gelangen.
8. *Communicating and scrutinizing*: Dieser Punkt umfasst die kritische Diskussion der erhaltenen Ergebnisse mit anderen Forschenden
(Fischer et al., 2014, S. 33–35)

Insgesamt stellt sich aus kognitionswissenschaftlicher Sicht der Prozess der Erkenntnisgewinnung als “[.] *a complex activity that requires the coordination of several higher level cognitive skills, including heuristic search through problem spaces, inductive reasoning, and deductive logic*” dar. (Zimmerman, 2000, S. 109). Innerhalb der Erkenntnisgewinnung lassen sich domänen-spezifisches Wissen (starke Methoden) und domänenübergreifende Strategien (schwache Methoden) unterscheiden, die in allen Wissenschaftsdisziplinen anwendbar sind (Klahr & Simon, 2001, S. 78; Zimmerman, 2000, S. 102). Die starken Methoden sind domänen-spezifische Methoden und umfassen beispielweise den Umgang mit Instrumenten, die Prozeduren im Forschungsprozess und experimentelle Paradigmen. Die starken Methoden unterscheiden das alltägliche Denken vom wissenschaftlichen Denken. Schwache Methoden des wissenschaftlichen Denkens bestehen sowohl aus universellen *Problemlöseprozessen* als auch aus *reasoning* (Klahr, 2000, S. 12f.). Problemlösen besteht darin, einen Weg von einem Anfangszustand in einen Zielzustand zu überführen (Mayer, 2012, S. 181): „[...] *the problem-solving process can be characterized as a search for a path that links initial state to goal state.*” (Klahr & Simon, 2001, S. 76). “*A problem consists of an initial state, a goal state, and a set of operators for transforming the initial state into the goal state by a series of intermediate steps.*” (Klahr & Simon, 2001, S. 76). Die Lösung eines Problems besteht im Planen, Durchführen und Überwachen einer Problemlösung (Mayer, 2012, S. 182). Reasoning ist „*the form of thinking that involves inference. Inference can be either inductive or deductive [...]*”. Inference “*is the process where one proposition (i.e., knowledge claim) is arrived at and accepted on the basis of other propositions that were originally accepted*” (Ricco & Overton, 2012, S. 257). Die Unterscheidung zwischen Reasoning und Problemlösen ist häufig nicht eindeutig, unter anderem deswegen, weil beide Konzepte die Idee des Erreichens eines Ziels beinhalten, für das es keine bekannten Lösungswege gibt (Zimmerman, 2000, S. 107).

Neben der Erkenntnisgewinnung, wie sie hier dargestellt wurde, setzt sich in einigen Ansätzen zum Scientific Reasoning auch die Erkenntnis durch, dass „[...] *developing scientific*

knowledge and implementing scientific enquiry are most often accompanied with the construction and testing of models.“ (Oh & Oh, 2011, S. 1110). So betonen Giere et al. (2006, S. 21): „*Understanding scientific reasoning requires understanding something about models and how they are used in science.*“ Wichtige Aspekte in der Arbeit mit Modellen sind der *Abgleich mit der Natur* (Giere et al., 2006, S. 25) und die *Ableitung (reasoning) von Vorrausagen* über die Natur aus einem Modell (Giere et al., 2006, S. 28f.). Oh und Oh (2011, S. 1111) unterscheiden die Aspekte *meaning of a model, purposes of modelling, multiplicity of scientific models, change in scientific models, uses of models in the science classroom.*

Folgerungen für ein Kompetenzmodell:

Die hier vorgestellten Ansätze der Erkenntnisgewinnung unterscheiden sich im Detail, stellen aber insgesamt die wissenschaftliche Erkenntnisgewinnung als Problemlöseprozess dar (Mayer, 2007, S. 177). Dieser Definition wird innerhalb der naturwissenschaftsdidaktischen Forschung auch mehrheitlich zugestimmt (Mayer, 2007, S. 184). Ein Kompetenzmodell zur Naturwissenschaftlichen Erkenntnisgewinnung muss demnach den Erkenntnisprozess als Problemlöseprozess darstellen. Weiterhin zeigt sich, dass der Einsatz von Modellen im Rahmen von *Scientific Reasoning* ebenfalls eine herausragende Rolle einnimmt.

3. Entwicklung eines Kompetenzstrukturmodells wissenschaftlicher Erkenntnisgewinnung

In den drei Naturwissenschaften Biologie, Chemie und Physik sind das Durchführen von Untersuchungen und das Modellieren zwei zentrale Methoden der Erkenntnisgewinnung, die sich nur durch den konkreten Fachinhalt unterscheiden. Die Erkenntnismethoden an sich unterscheiden sich nicht. Bezogen auf den in der Definition von Kompetenz auftretenden Kontextbezug, den Hartig (2008, S. 21) als „eine Menge hinreichend ähnlicher realer Situationen, in denen bestimmte, ähnliche Anforderungen bewältigt werden müssen“ beschreibt, bedeutet dies, dass diese Erkenntnisprozesse *ähnlichen reale Situationen* darstellen, in denen *ähnliche Anforderungen* bewältigt werden müssen. Daraus resultiert die Einschränkung, dass ausschließlich die schwachen, domänenübergreifenden wissenschaftlichen Erkenntnisprozesse (Klahr, 2000) berücksichtigt werden können. Die starken, domänenspezifischen Methoden sind spezifisch für eine Naturwissenschaft und müssen daher hier unberücksichtigt bleiben. Kompetenzen im Bereich der naturwissenschaftlichen Erkenntnisgewinnung werden häufig so modelliert, „*dass der Prozess zur Aufstellung und Begründung von Hypothesen und Theorien als forschungslogischer Ablauf mit verschiedenen Teilschritten dargestellt wird.*“ (Mayer & Wellnitz, 2014, S. 22–23). Wellnitz et al. (2012) stellen in ihrer Übersicht verschiedene Ansätze für die Modellierung naturwissenschaftlicher Erkenntnisgewinnung dar (siehe Tabelle 1).

Tabelle 1: Übersicht über Strukturmodelle wissenschaftlicher Erkenntnisgewinnung (verändert und erweitert nach Wellnitz et al., 2012)

<p>Hamman, 2007; Klahr, 2000; Klos, Henke, Kieren, Walpulski & Sumfleth, 2008; Woitkowski, Riese & Reinhold, 2011¹⁴;</p>	<p>Hofstein, Navon, Kipnis & Mamlouk-Naaman, 2005</p>	<p>Grube, 2010; Mayer, 2007; Mayer, Grube & Möller, 2008</p>	<p>Nowak, Nehring, Tiemann & Upmeyer zu Belzen, 2013¹⁵</p>
<p>Hypothesenbildung und Hypothesenerweiterung</p>	<p>Fragen und Hypothesen aufstellen</p>	<p>Fragstellungen formulieren</p>	<p>Fragstellungen und Hypothesen bilden</p>
		<p>Hypothesen generieren</p>	
<p>Testen von Hypothesen</p>	<p>Experimente planen</p>	<p>Untersuchungen planen</p>	<p>Untersuchungen durchführen</p>
	<p>Experimente durchführen und beobachten</p>		
<p>Analyse von Evidenzen</p>	<p>Ergebnisse analysieren und Schlussfolgerungen ziehen</p>	<p>Daten analysieren und Schlussfolgerungen ziehen</p>	<p>Daten analysieren und Schlussfolgerungen ziehen</p>

¹⁴ Sie beziehen hierbei auch den Umgang mit Modellen mit ein

¹⁵ Sie beziehen verschiedene Untersuchungsmethoden (inquiry methods) mit ein, auf welche diese Schritte angewendet werden können (Beobachten, Experimentieren, Modellieren)

Die in Tabelle 1 dargestellten Kompetenzstrukturmodelle berücksichtigen mit unterschiedlicher Auflösung einen forschungslogischen Ablauf im Sinne des Scientific Reasoning (siehe Abschnitt 2.6). Auch die Verwendung naturwissenschaftlicher Modelle zur Erkenntnisgewinnung wird in den Kompetenzstrukturmodellen von Woitkowski et al. (2011) und Nowak et al. (2013) berücksichtigt. Diesen kommt im Scientific Reasoning ebenfalls eine herausragende Rolle zu (s.o., Giere et al., 2006; Oh & Oh, 2011). Das dieser Studie zugrunde liegende Kompetenzstrukturmodell nutzt für den Bereich der wissenschaftlichen Untersuchungen das Modell von Mayer (2007) und ergänzt dieses um den Bereich naturwissenschaftlicher Modelle (Upmeier zu Belzen & Krüger, 2010).

3.1 Scientific Reasoning nach Mayer

Mayer (2007) definiert den Prozess der naturwissenschaftlichen Erkenntnisgewinnung als „relativ komplexe[n], kognitive[n], wissensbasierte[n] Problemlöseprozess [...], der durch spezifische Prozeduren charakterisiert ist.“ (Mayer, 2007, S. 181). Er unterscheidet innerhalb dieses Prozesses vier eigenständige Prozeduren (siehe Grafik). Einfluss auf die Problemlösung haben außerdem Personen- und Situationsvariablen.

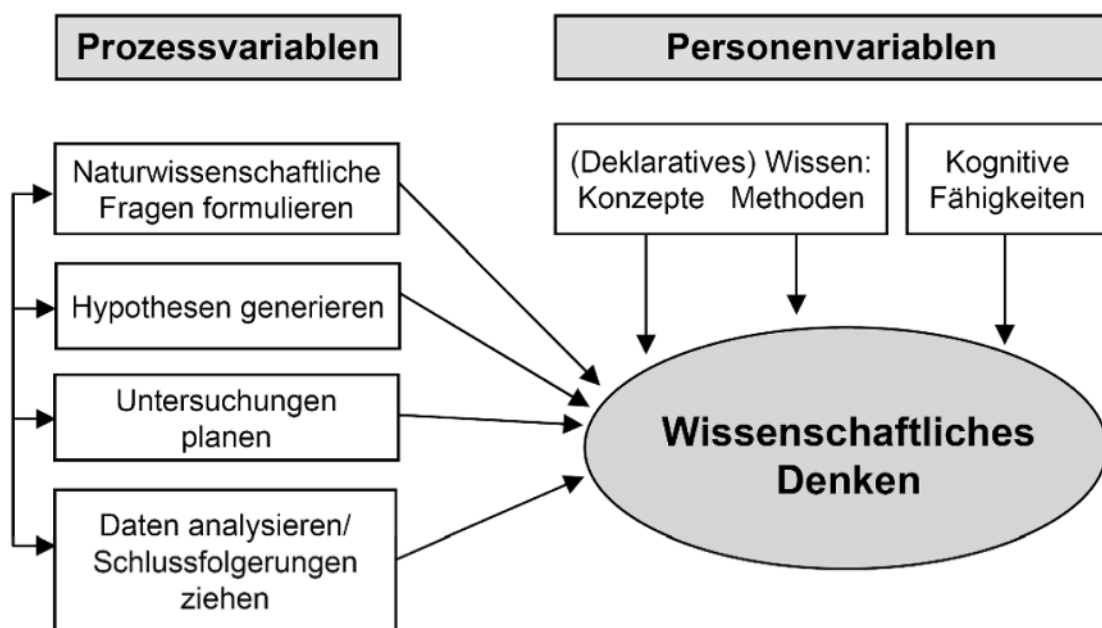


Abbildung 1: Strukturmodell zum Scientific Reasoning (aus Mayer, 2007, S. 181)

3.2 Modellkompetenz

Modellkompetenz nach Upmeier zu Belzen und Krüger (2010) umfasst

„[...] die Fähigkeiten, mit Modellen zweckbezogen Erkenntnisse gewinnen zu können und über Modelle mit Bezug auf ihren Zweck urteilen zu können, die Fähigkeiten, über den Prozess der Erkenntnisgewinnung durch Modelle und Modellierungen [...] zu reflektieren sowie die Bereitschaft, diese Fähigkeiten in problemhaltigen Situationen anzuwenden.“

(Upmeier zu Belzen & Krüger, 2010, S. 51).

Sie unterscheiden zwei Dimensionen von Modellkompetenz: *Kenntnisse über Modelle* und *Modellbildung*. Dabei sehen sie die Begründung der Dimension *Kenntnisse über Modelle* im Bereich des Wissenschaftsverständnisses. Die Dimension *Modellbildung* fokussiert auf den Umgang mit Modellen mit Bezug zum wissenschaftlichen Denken (Mayer, 2007; Upmeier zu Belzen & Krüger, 2010). Sie teilen die letztere Dimension in die drei Facetten *Zweck von Modellen*, *Testen von Modellen* und *Ändern von Modellen*.

3.3 Kompetenzstrukturmodell

Diese Definition mündet in ein Kompetenzstrukturmodell mit den beiden Dimensionen *Naturwissenschaftliche Untersuchungen* und *Modelle Nutzen*. Beide Dimensionen sind in weitere Kompetenzfacetten unterteilt (siehe Tabelle 2). Zusammenfassend stellt sich das hier genutzte Kompetenzstrukturmodell folgendermaßen dar:

Tabelle 2: Kompetenzstrukturmodell zur Naturwissenschaftlichen Erkenntnisgewinnung

Naturwissenschaftliche Erkenntnisgewinnung				
Naturwissen- schaftliche Untersuchungen	Frage- stellungen	Hypothesen	Planung und Durchführung	Auswertung und Inter- pretation
Modelle Nutzen	Zweck von Modellen	Testen von Modellen		Ändern von Modellen

Die durch das Modell fokussierten Kompetenzen lassen sich im Modell des Lehrerverfessionswissens im Bereich des SMK, genauer des Syntactic Knowledge (siehe Abschnitte 2.2 und 2.3.2) verorten. Damit wird ein weiterer Aspekt des Lehrerverfessionswissens

beleuchtet, der bislang nur aus fachdidaktischer Sicht (in ProWiN, Tepner et al., 2012) bzw. nur für die Physik (Woitkowski, 2015; Woitkowski et al., 2014) untersucht wurde.

3.4 Operationalisierung der Kompetenzen

Zur Entwicklung von Aufgaben im Bereich der Erkenntnisgewinnung ist es notwendig, die einzelnen Zellen¹⁶ des Kompetenzstrukturmodells genau zu operationalisieren. Dieses soll im Folgenden geschehen.

3.4.1 Dimension: Naturwissenschaftliche Untersuchungen

Die zentrale Untersuchungsmethode der Physik stellt das physikalische Experiment dar (z. B. Shamos, 1959). Mit Hilfe von Experimenten lassen sich *Fragen an die Natur* beantworten. Diese Frage wird aus einer Theorie heraus gewonnen, woraus eine Hypothese abgeleitet wird. Das Experiment führt zur Stützung oder Widerlegung der Hypothese (Hunger, 1964, S. 28). Experimente dienen demnach zur Prüfung von Hypothesen und Theorien, können aber auch Quelle dieser sein. Es „[...] handelt sich um eine Tätigkeit von Wissenschaftlern, die den Zweck, hat bestimmte Ergebnisse [...] zu liefern“ (Pietschmann, 1996, S. 83).

„Im engeren Sinne bezeichnet [der Begriff „Experiment“] spezielle apparative Anordnungen, die physikalische Vorgänge unter reproduzierbaren und variierbaren Bedingungen beobachtbar machen und die entwickelt werden, um Hypothesen oder Prognosen im Zusammenhang physikalischer Theorien zu prüfen“

(Bleichroth et al., 1999, S. 221)

Experimente müssen demnach verschiedene Bedingungen erfüllen, um als wissenschaftliche Experimente anerkannt zu werden. Sie müssen reproduzierbar sein, also unter genauer Wiederholung der äußeren Bedingungen durchgeführt werden und dabei die gleichen Ergebnisse liefern. Die äußeren Bedingungen müssen durch Messvorgänge kontrolliert werden (Pietschmann, 1996, S. 84). Ein weiterer Punkt ist die Veränderbarkeit von äußeren Parametern. (Majer, 1980, S. 208). So wird es möglich zu untersuchen, wie sich eine bestimmte abhängige Variable in Bezug auf die Veränderung der unabhängigen Variable verhält. Eine weitere Forderung an Experimente ist, dass die Ergebnisse auf irgendeine Weise quantifizierbar sein müssen. Erst dadurch kann eine objektive Überprüfung stattfinden. Andernfalls würde die Überprüfung subjektiv bleiben (Pietschmann, 1996, S. 86f.).

¹⁶ Die Begriffe ‚Zelle‘ des Kompetenzstrukturmodells und ‚Facette‘ der Kompetenz sind synonym zu verstehen.

Bei der Interpretation von den aus dem Experiment erhaltenen Evidenzen muss eine kritische Grundhaltung angenommen werden: „[...] *we must never do experiments in order to confirm our ideas but merely to check them.*“; (Duhem, 1998, S. 258). Weiterhin ist es problematisch, Experimente nur im Sinne seiner eigenen Theorie zu sehen: “[...] *it quite naturally happens that those who believe to much in their own theories do not sufficiently believe in the theories of others*” (Duhem, 1998, S. 258f.). Vielmehr muss man sämtliche Beobachtungen zur Kenntnis nehmen, nicht nur solche, die im Sinne der zu prüfenden Theorie stehen.

In der Physik spielen oft auch Wahrscheinlichkeitshypothesen eine Rolle (so zum Beispiel bei der Halbwertszeit von radioaktiven Elementen). Hierbei ist eine direkte Falsifikation schwierig festzustellen, da selbst dann, wenn ein als wahrscheinlich vorhergesagtes Ereignis nicht eintritt, ja auch noch eine Wahrscheinlichkeit bestand, dass das Ereignis nicht eintritt. Es muss daher vorher auf methodologischer Ebene durch die Forschenden festgelegt werden, ab welchem Grad der Unwahrscheinlichkeit ein Ereignis als falsifiziert gilt (Wiltsche, 2013, S. 93).

Experimente in der Wissenschaft sind häufig nicht einfach den Sinnen zugänglich. Während die Beobachtung zweier fallender Steine noch relativ einfach mit den Augen möglich ist, entzieht sich heutige Forschung in den meisten Fällen direkter Beobachtung (Chalmers, 2007, S. 28). Die Experimentatoren sind auf Messinstrumente angewiesen, deren Messungen sie dann interpretieren müssen. Hierbei tritt ein weiteres Problem auf. Die erhaltenen Messwerte müssen so weit wie möglich die intendierte Messgröße anzeigen. Dazu müssen sämtliche Störgrößen eliminiert werden. Dies setzt aber voraus, dass dem Experimentator diese auch bekannt sind (Chalmers, 2007, S. 28). Dieses ist einerseits vom Wissen des Experimentators abhängig, andererseits aber natürlich auch vom Stand der Wissenschaft überhaupt. Bestimmte Störgrößen sind während des Experiments möglicherweise noch gar nicht bekannt und können dementsprechend auch nicht berücksichtigt werden. Die Messungen basieren demnach auf Nebenhypothesen, die ihrerseits auch fehlbar sind. Dieser Aspekt des physikalischen Experimentierens ist auch als Duhem-Quine-These bekannt:

„The physicist can never subject an isolated hypothesis to experimental test, but only a whole group of hypotheses; when the experiment is in disagreement with his predictions, what he learns is that at least one of the hypotheses constituting this group is unacceptable and ought to be modified; but the experiment does not designate which one should be changed“
(Duhem, 1998, S. 263).

Die von Popper proklamierten „crucial experiments“ sind in der Physik demnach unmöglich (Duhem, 1998, S. 264).

a) Kompetenzfacette: Fragestellungen

Wissenschaftliche Fragestellungen müssen präzise und eindeutig formuliert sein. Uneindeutige Fragestellungen können zu unpräzisen oder mehrdeutigen Antworten führen (Menne, 1980, S. 119). Die Fragestellungen müssen den derzeitigen Stand der Forschung berücksichtigen (vgl. z.B. Kuhn, 1976b; Lakatos, 1974; Popper, 1974). Gleichzeitig dürfen aber auch akzeptierte Sätze in Frage gestellt werden (Menne, 1980, S. 119).

b) Kompetenzfacette: Hypothesen

„During hypothesis generation, students derive possible answers to the question from plausible models, available theoretical frameworks or empirical evidence they are aware of.“ (Fischer et al., 2014, S. 33; siehe auch Klahr & Dunbahr, 1988). Zur Beantwortung einer Fragestellung werden theoriegeleitet Hypothesen aufgestellt, welche den derzeitigen Stand der Forschung berücksichtigen. Um Probleme zu umgehen, die aus verschiedenen Deutungen des Ausgangs eines Experiment resultieren können – also der Frage, ob das Experiment die Hypothese nun stützt oder verwirft – ist es notwendig Hypothesen exakt und eindeutig zu formulieren (Chalmers, 2007, S. 58). Eine Hypothese muss so aufgestellt werden, dass sie „an der sinnlichen Erfahrung“ scheitern kann (Wiltsche, 2013, S. 81). Sie muss demnach falsifizierbar sein (Popper, 1984).

c) Kompetenzfacette: Planung und Durchführung

„A False scientific hypotheses is best refuted by showing that the „facts“ deduced from it are not true; that is by scientific test or experiment“ (Shamos, 1959, S. 8). Eine aufgestellte Hypothese wird durch ein Experiment geprüft. Dabei wird theoriegeleitet der Einfluss einer Variable auf eine andere beobachtet (Fischer et al., 2014). Die zu beobachtende Variable muss entsprechend isoliert werden, d.h. sämtliche anderen Variablen müssen kontrolliert werden (Pietzschmann, 1996). Dabei ist es notwendig, dass sämtliche störende Variablen auch bekannt sind (Chalmers, 2007).

d) Kompetenzfacette: Auswertung und Interpretation

Die Entscheidung über das Verwerfen oder die Akzeptanz einer Hypothese ist ein komplexer Prozess, der, wie dargestellt, auch von vorherrschenden Paradigmen beeinflusst wird und ebenfalls nicht auf Grundlage eines Experiments geschehen kann. Außerdem sind zur Auswertung ein theoretisches Wissen über Wirkungszusammenhänge und Alter-

nativhypothesen notwendig (Koslowski, 1996, s.o.). Im Fokus des Bereichs Auswertung und Interpretation steht daher die Frage, ob ein durchgeführtes Experiment eine vorliegende Hypothese stützt oder sie ablehnt: „*The aim of evidence evaluation is to assess the degree to which a certain piece of evidence supports a claim or theory.*“ (Fischer et al., 2014, S. 34).

3.4.2 Modelle Nutzen

„*Modelle sind Annäherungen (an die Wirklichkeit), aber nicht immer wegen einer Ungenauigkeit, sondern weil sie nur bestimmte Aspekte berücksichtigen.*“ (Kreisel, 1980, S. 438) Im Sinne des scientific reasoning werden zu Modellen skalierte Nachbildungen, Analogien und (die am häufigsten in der Wissenschaft genutzten) theoretischen Modelle gezählt (Giere et al., 2006, S. 21ff.). Skalierte Modelle sind in der Größe veränderte Nachbildungen. Ein theoretisches Modell „*[...] is part of an imagined world. It does not exist anywhere except in scientists' mind [...]*“ (Giere et al., 2006, S. 24).

“[...] a model is understood as a representation of a target. The targets represented by models can be various entities, including objects, phenomena, processes, ideas and their systems.”

(Oh & Oh, 2011, S. 114).

Der in diesem Zitat auftretende Begriff „*Target*“ (dt. Ziel) betrifft den repräsentierten Gegenstand oder Mechanismus.

In der Physikdidaktik umfasst der Begriff „Modell“ nach Stöckler (1995) materielle Nachbildungen und theoretische Modelle. Als weitere Gruppe von Modellen nennt er interpretierte Strukturen, die jedoch nur im Bereich der Mathematik auftauchen. Als materielle Nachbildungen gelten Modelle im Alltagssinn, zum Beispiel verkleinerte oder vergrößerte Abbildungen der Realität. Ein Modellauto oder eine Modelleisenbahn erfüllt – nicht nur durch den Namen – diese Definition. In der Wissenschaft werden verkleinerte Modelle zum Beispiel bei der Konstruktion von Schiffen oder Flugzeugen eingesetzt (Stöckler, 1995). Eine materielle Nachbildung muss aber nicht zwangsläufig eine in der Größe veränderte Abbildung der Realität sein. Vielmehr handelt es sich auch bei Gegenständen um Modelle, die das Verhalten eines anderen anschaulich erklären können. Eine Feder, die auf eine Schnur gespannt und angeregt wird und dabei Longitudinalwellen vorführt ist ein Modell für die Übertragung von Schallwellen in Luft. Ferner zählen als Modelle Nachbildungen von Experimenten, die für den Unterricht didaktisch aufgearbeitet wurden.

Theoretische Modelle sind spezielle Theorien oder Teile einer Theorie. Sie sind „eine idealisierte Beschreibung eines Gegenstandes und seines Verhaltens.“ (Stöckler, 1995). Sie sind demnach Denkgebäude, die ebenso wie ihre realen Pendanten bestimmte Eigenschaften der Realität berücksichtigen und andere vernachlässigen.

Kircher (1995) schlägt vor, den Modellbegriff ausschließlich auf theoretische Konstrukte anzuwenden. Gegenständliche Nachbildungen von etwas sollten, seiner Ansicht nach als „Realisierung von Modellvorstellungen“ bezeichnet werden. Er räumt aber ein, dass dies weit von der umgangssprachlichen Bedeutung entfernt ist. Stattdessen unterscheidet er fortan „theoretische Modelle“ und „gegenständliche Modelle“ bzw. „Sachmodelle“.

Im vorliegenden Forschungsvorhaben werden unter dem Begriff Modelle sowohl materielle Nachbildungen als auch theoretische Modelle verstanden. Dies wird auch unter dem Bewusstsein vorgenommen, dass theoretische Modelle ein sehr viel höheres Abstraktionsvermögen erfordern als materielle Nachbildungen. Dies begründet sich darin, dass bei materiellen Nachbildungen die äußeren Gegebenheiten ohnehin der Realität ähneln (z.B. Flugzeug im Windkanal) oder auf Grund einer Didaktisierung des Experiments viel Wert auf Anschaulichkeit gelegt wird. Bei theoretischen Modellen haben die verwendeten Zeichen und Symbole keinerlei anschaulichen Charakter. Zur Interpretation bedarf es mathematischer Kenntnisse und Vorstellungskraft. Es wird jedoch angenommen, dass die drei Teilkompetenzen, „Zweck von Modellen“, „Testen von Modellen“ und „Ändern von Modellen“, die in diesem Projekt untersucht werden sollen, unabhängig von der Art des Modells sind. Möglicherweise hat die Art des Modells aber einen Einfluss auf die Schwierigkeit der Aufgaben.

a) Kompetenzfacette: Zweck von Modellen

Naturwissenschaftliche Modelle werden zum Beschreiben, Erklären und Voraussagen von naturwissenschaftlichen Phänomenen genutzt (Oh & Oh, 2011, S. 1116). Einerseits können bekannte Korrelationen und Zusammenhänge im Modell beschrieben werden (Modell von etwas), andererseits können mit Modellen Voraussagen über bislang unbekannte Korrelationen und Zusammenhänge in der Natur getätigt werden (Upmeyer zu Belzen & Krüger, 2010, S. 53).

b) Kompetenzfacette: Testen von Modellen

Modelle können getestet werden, indem Sie mit der Natur abgeglichen werden (Oh & Oh, 2011, S. 1119). Dies kann sowohl durch einen nachträglichen Abgleich mit der Natur (Modell von etwas), als auch durch Ableitung von Hypothesen aus dem Modell, die dann in der Natur getestet werden (Modell für etwas), geschehen. Letztere Perspektive ist die Expertensicht auf Modelle (Upmeier zu Belzen & Krüger, 2010, S. 52f.).

c) Kompetenzfacette: Ändern von Modellen

Im Falle einer mangelhaften Passung der aus der Natur gewonnenen Daten mit dem Modell oder eines Widerspruchs zwischen den aus der Natur gewonnenen Daten und den durch das Modell vorausgesagten Daten, muss das Modell geändert werden (Oh & Oh, 2011, S. 1119). Auch hier besteht die Expertenperspektive darin, Modelle zu verwerfen, weil aus ihnen abgeleitete Hypothesen falsifiziert wurden (Upmeier zu Belzen & Krüger, 2010, S. 53).

4. Forschungsfragen und Hypothesen

Das in Kapitel 3 beschriebene Kompetenzstrukturmodell zur naturwissenschaftlichen Erkenntnisgewinnung zeichnet sich durch eine heterogene Struktur mit den beiden Dimensionen naturwissenschaftlicher Arbeitsweisen *Untersuchungen* und *Modelle nutzen* aus, mit jeweils drei- bzw. vier Kompetenzfacetten (s. Abschnitt 3.3).

Die zentrale Idee bei der Entwicklung des Modells war der Ansatz einer domänenübergreifenden naturwissenschaftlichen Erkenntnisgewinnungskompetenz (Fokussierung auf die schwachen, fächerübergreifenden und fachwissensunabhängigen Arbeitsweisen der naturwissenschaftlichen Erkenntnisgewinnung; s. Kap. 3). Bei der empirischen Überprüfung der Gültigkeit dieser Annahme kommt als potentiell und ggf. weiteres Unterscheidungsmerkmal dennoch eine Differenzierung in die drei naturwissenschaftlichen Domänen bzw. Fächer Physik, Biologie und Chemie hinzu.

Es gibt also aus theoretischer Sicht zumindest zwei Merkmale, die eine mehrdimensionale Struktur des Modells wahrscheinlich machen. Zur Erfassung der hier modellierten naturwissenschaftlichen Erkenntnisgewinnungskompetenz wurde im Rahmen der Arbeit auch ein Testinstrument bzw. Kompetenztest mit entwickelt, und die Analyse der damit erhobenen Daten können später Aufschluss über die Dimensionalität des Modells liefern.

Ein weiteres und zentrales Ziel dieser Arbeit besteht darin zu untersuchen, inwiefern sich Lehramts- und Fachstudierende im Kompetenzstand unterscheiden und wie der Kompetenzstand zu verschiedenen Phasen des Studiums differiert.

Im Weiteren wird auch der Kompetenzstand von Grundschullehramtsstudierenden mit dem Studienfach „Integrierte Naturwissenschaften“ näher untersucht.

Für diese Arbeit sind daher insgesamt die folgenden Forschungsfragen leitend.

4.1 Frage zur Dimensionierung des Kompetenzstrukturmodells

FF1: Welches Modell beschreibt die Daten am besten?

Erste Analysen der Gesamtstichprobe (Hartmann, Mathesius et al., 2015) oder von Teilstichproben (Stiller, Straube, Hartmann, Nordmeier & Tiemann, 2015), kamen zu dem Ergebnis, dass ein eindimensionales Modell die bei der Kompetenzmessung erhobenen Daten am besten beschreibt. Für die Passung von Modellen, die nach Arbeitsweisen oder Fächern differenzieren, gab es geringere Evidenz. Es stellt sich daher die Frage, ob dieses Ergebnis auch für die Teilstichprobe der Physikstudierenden gilt.

H1a: Ein eindimensionales Modell beschreibt die Daten besser als ein Modell, das nach Fächern differenziert.

Der entwickelte Kompetenztest soll auf die schwachen, fächerübergreifenden Methoden der naturwissenschaftlichen Erkenntnisgewinnung fokussieren. Sofern der Fachinhalt bekannt ist, sollte es daher für die Proband/inn/en keinen Unterschied machen, ob sie Aufgaben einer anderen Naturwissenschaft als der Physik bearbeiten (vgl. Giere et al., 2006; Klahr, 2000). Es werden daher drei Modelle getestet:

1A: eindimensionales Modell

1B: zweidimensionales Modell *Physik* vs. *Biologie & Chemie*

1C: dreidimensionales Modell *Physik, Biologie, Chemie*

H1b: Ein in Bezug auf die Arbeitsweisen *Untersuchungen* und *Modelle Nutzen* zweidimensionales Modell beschreibt die Daten besser als ein eindimensionales Modell.

Auch wenn die oben genannten vorläufigen Ergebnisse auf eine eindimensionale Struktur des Modells hindeuten, die sich auch nach den Arbeitsweisen nicht unterscheiden lässt, so gibt es doch theoretische Gründe, die für eine mehrdimensionale Struktur sprechen. So setzt sich das genutzte Kompetenzstrukturmodell aus zwei ursprünglich getrennten Kompetenzstrukturmodellen zusammen (siehe Abschnitt 3.3). Außerdem ist gerade in der Physik die Unterscheidung nach experimenteller und theoretischer Physik lange etabliert und spiegelt sich auch in der universitären Lehre wider. Eine Unterscheidung nach den beiden Arbeitsweisen könnte daher auch in der Datenstruktur zu finden sein. Von den Kompetenzmodellen her wären auch weitere Unterscheidungen nach den sieben Kompe-

tenzfacetten möglich. Da aufgrund des Multimatrix-Designs aber nur nach Arbeitsweisen und nicht nach den einzelnen Facetten unterschieden wurde, kann es sein, dass diese Unterscheidung, auch wenn sie vorhanden wäre, mit den vorliegenden Daten nicht nachweisbar ist.

Es werden die drei folgenden Modelle getestet:

2A: eindimensionales Modell

2B: zweidimensionales Modell nach Arbeitsweisen (*Untersuchungen & Modelle Nutzen*)

2C: siebendimensionales Modell nach Kompetenzfacetten

H1c: Ein eindimensionales Modell beschreibt die Daten besser als ein mehrdimensionales Modell, welches nach Fächern und Arbeitsweisen unterscheidet.

Aus theoretischer Sicht ist aus den oben dargestellten Gründen auch eine Unterscheidung nach Fächern und Arbeitsweisen denkbar. Um diese Hypothese zu testen, werden folgende Modelle miteinander verglichen:

3A: eindimensionales Modell

3B: vierdimensionales Modell nach Fächern (*Physik vs. Biologie & Chemie*) und Arbeitsweisen (*Untersuchungen und Modelle nutzen*)

3C: sechsdimensionales Modell nach *Fächern* und *Arbeitsweisen*

4.2 Fragen zum Kompetenzstand von Physikstudierenden

FF2: Existieren Unterschiede im Kompetenzstand zwischen Lehramtsstudierenden der Physik und Fachstudierenden der Physik?

Zunächst stellt sich die Frage, ob es generelle Unterschiede im Kompetenzstand von Lehramts- und Fachstudierenden der Physik gibt. Lehramtsstudierende beginnen das Studium mit geringeren Eingangsvoraussetzungen im Vergleich zu Fachstudierenden der Physik (Albrecht, 2011). Daher ist ein geringerer Kompetenzstand zu Beginn des Studiums zu erwarten. Sollte sich die Lehramtsausbildung aber als wirksamer erweisen als die der Fachstudierenden (vgl. Fragestellung 4), so könnte dieser Effekt aufgehoben werden.

H2: Lehramtsstudierende weisen einen geringeren Kompetenzstand als Fachstudierende der Physik auf (sig. (Lehramt) < 0,05).

FF3: Existieren Unterschiede im Kompetenzstand Studierender im Bereich der Erkenntnisgewinnung zu verschiedenen Phasen im Studienverlauf?

Im Laufe des Studiums besuchen die Studierenden Lehrveranstaltungen, in denen das zu messende Konstrukt explizit, z. B. in Seminaren zur Wissenschaftstheorie, oder implizit, z.B. im Rahmen von Experimentalpraktika, gefördert wird. Mit steigender Semesterzahl wächst auch die Anzahl der Lerngelegenheiten, welche als Prädiktor für den Erwerb von Fachwissen¹⁷ aber auch Kompetenz gilt (z. B. Blömeke et al., 2010) Es wird demnach erwartet, dass mit fortschreitendem Studium, auch die Erkenntnisgewinnungskompetenz der Studierenden zunimmt.

H3: Der mittlere Kompetenzstand der Studierenden der Physik ist höher, je weiter fortgeschritten die Studierenden im Studium sind (p < 0,05).

¹⁷ Im Sinne des Content Knowledge nach Shulman (siehe Abschnitt 2.3)

FF4: Existieren Unterschiede in den Differenzen im Kompetenzstand zwischen Lehramtsstudierenden der Physik und Fachstudierenden zu den unterschiedlichen Phasen des Studiums?

Grundsätzlich unterscheidet sich die Ausbildung von Lehramtsstudierenden und Fachstudierenden der Physik im Umfang des Studiums. Lehramtsstudierende beschäftigen sich ungefähr nur die Hälfte der Studienzeit mit Physik, die restliche Zeit wird für das zweite Fach, sowie die lehramtsbezogene Berufswissenschaft aufgewendet. Bezogen auf den Aspekt der naturwissenschaftlichen Erkenntnisgewinnung gibt es Unterschiede bei den experimentellen Praktika. Der Anteil von Experimentalpraktika ist bei Fachstudierenden höher, bei Lehramtsstudierenden sind dafür sogenannte Demonstrationspraktika vorgesehen. Lehramtsstudierende besuchen außerdem Didaktikveranstaltungen, in denen das Thema Erkenntnisgewinnung auch explizit thematisiert werden kann (Deutsche physikalische Gesellschaft, 2014).

Mögliche Unterschiede im Kompetenzstand könnten von einer unterschiedlichen Studienzeit herrühren. Gleichzeitig ist aber im Lehramtsstudium eine explizite Förderung dieser Kompetenzen vorgesehen, welche als besonders nützlich angesehen wird (Giere et al., 2006, S. 6). Da sich aus der Theorie nicht ableiten lässt, inwiefern eine explizite Vermittlung eine kürzere Beschäftigungszeit mit dem Thema aufwiegen kann, wird die Nullhypothese geprüft:

H4: Es existieren keine Unterschiede in den Differenzen im Kompetenzstand zwischen Lehramtsstudierenden der Physik und Fachstudierenden zu den unterschiedlichen Phasen des Studiums (sig. (Lehramt*Studienphase) > 0,05; $1 - \beta \geq 0,90$).¹⁸

FF5: Existieren Unterschiede im Kompetenzstand von weiblichen und männlichen Studierenden?

Mögliche Unterschiede zwischen den Geschlechtern im Kompetenzstand wurden in verschiedenen Schulleistungstests beforscht. Im Rahmen des IQB-Ländervergleichs (Pant et al., 2013) wurden explizit der Bereich Erkenntnisgewinnung auf derartige Disparitäten

¹⁸ Grundsätzlich wird hier davon ausgegangen, dass diese querschnittlich erhobenen Daten auch ein Hinweis auf eine Kompetenzentwicklung im Rahmen des Studiums darstellen.

untersucht. Insgesamt zeigt sich ein Vorteil für die Schülerinnen. Besonders stark ist dieser im Kompetenzbereich der Erkenntnisgewinnung im Fach Biologie, etwas schwächer aber trotzdem signifikant auch in den beiden anderen Naturwissenschaften ausgeprägt. Betrachtet man nur die Gymnasien, so zeigt sich ein leichter Vorteil für die männlichen Schüler in den Fächern Chemie und Physik, der Vorteil im Bereich Biologie bleibt für die Schülerinnen bestehen, auch wenn er etwas schwächer ausfällt (Schroeders, Penk, Jansen & Pant, 2013). Woitkowski (2015, S. 222) fand in seiner Analyse der Erkenntnisgewinnungskompetenz von Physikstudierenden signifikante Vorteile für die männlichen Studierenden. Allerdings berücksichtigt er in seiner Studie nur physikalische Aspekte. Aus diesen Ergebnissen kann nicht unbedingt auf die fächerübergreifende Erkenntnisgewinnungskompetenz geschlossen werden. Da das Bild im IQB-Ländervergleich für die Gymnasien über die drei Fächer ebenfalls nicht konsistent ist, wird zunächst die Nullhypothese getestet:

H5: Weibliche Studierende zeigen den gleichen Kompetenzstand wie männliche Studierende ($p > 0,05$, $1 - \beta \geq 0,90$).

FF6: Gibt es Unterschiede im Kompetenzstand von Lehramtsstudierenden, die neben Physik ein weiteres naturwissenschaftliches Fach studieren im Vergleich zu denen, die kein weiteres naturwissenschaftliches Fach studieren?

Ausgehend davon, dass eine größere Anzahl an Lerngelegenheiten zu einem erhöhten Kompetenzstand beiträgt (z.B. Blömeke et al., 2010) und dass sich dieser erhöhte Kompetenzstand von Studierenden mit zwei naturwissenschaftlichen Fächern in der Gesamtkohorte tatsächlich auch zeigte (Hartmann & Mathesius et al., 2015), stellt sich die Frage, ob dieser Vorteil auch für Physikstudierende gilt. In vorläufigen Analysen des Datensatzes zeigte sich widererwarten, dass für Physikstudierende kein Vorteil zu erkennen war (Straube & Nordmeier, 2014). Trotzdem wird die Hypothese untersucht:

H6: Studierende mit einem weiteren naturwissenschaftlichen Fach weisen einen höheren Kompetenzstand auf, als Studierende ohne weiteres naturwissenschaftliches Fach ($p < 0,05$).

4.3 Fragen zum Kompetenzstand von Studierenden im Fach „Integrierte Naturwissenschaften“

FF7: Haben Grundschullehramtsstudierende des Fachs „Integrierte Naturwissenschaften“ den gleichen Kompetenzstand zu Beginn des Studiums wie Lehramtsstudierende der Physik?

An der Freien Universität Berlin wird seit dem Wintersemester 2011/2012 ein das Studienfach „Integrierte Naturwissenschaften“ für das Lehramt an Grundschulen angeboten. Es existieren keine theoretischen oder empirischen Gründe, die für einen geringeren Kompetenzstand dieser Studierenden zu Beginn des Studiums sprechen. Daher wird die Hypothese getestet:

H7: Studierende des Fachs „Integrierte Naturwissenschaften“ haben den gleichen Kompetenzstand wie Physiklehramtsstudierende zu Beginn des Studiums ($p > 0,05$, $1 - \beta \geq 0,90$).

FF8: Haben Studierenden des Faches „Integrierte Naturwissenschaften“ in höheren Fachsemestern einen höheren Kompetenzstand als zu Beginn des Studiums?

Ähnlich wie für den Kompetenzstand im Physikstudium zu verschiedenen Phasen des Studiums, wird auch für die Studierenden des Faches „Integrierte Naturwissenschaften“ angenommen, dass mit steigender Semesterzahl aufgrund der steigenden Anzahl der Lerngelegenheiten ein höherer Kompetenzstand zu erwarten ist. (z.B. Blömeke et al., 2010). Es wird demnach die Hypothese aufgestellt:

H8: Studierende des Fachs „Integrierte Naturwissenschaften“ weisen im höheren Fachsemester einen höheren Kompetenzstand auf als zu Beginn des Studiums ($p < 0,05$).

4.4 Frage zum Kompetenzstand von Studierenden der Sozialkunde

FF9: Welchen Kompetenzstand haben Studierende eines sozialwissenschaftlichen Faches im Vergleich zu Studierenden der Naturwissenschaften?

Grundsätzlich folgen die Sozialwissenschaften den gleichen Forschungsparadigmen, wie die Naturwissenschaften. Es werden die selben schwachen Methoden (Klahr, 2000) verwendet. Der vorliegende Kompetenztest soll ausschließlich auf die methodischen Aspekte fokussieren und Fachinhalte außen vor lassen. Daher müsste es für Studierende eines sozialwissenschaftlichen Faches ebenso möglich sein, die Aufgaben zu beantworten. Gleichzeitig wird aber insgesamt eine naturwissenschaftliche Erkenntnisgewinnungskompetenz geprüft. Demnach ist grundsätzlich davon auszugehen, dass der Kompetenzstand der Studierenden eines sozialwissenschaftlichen Faches geringer ist, als der Studierenden eines naturwissenschaftlichen Faches. Es wird daher die Hypothese geprüft

H9: Studierende der Sozialwissenschaften weisen einen geringeren Kompetenzstand, als Studierende der Naturwissenschaften auf ($p > 0,05$).

5. Methoden

Zur Beantwortung dieser Fragestellungen werden im Folgenden die genutzten Methoden erläutert. Da zu Beginn des Projekts kein geeignetes deutschsprachiges Testinstrument zur Verfügung stand, das das Kompetenzmodell entsprechend abbildet und zusätzlich für die betreffende Altersgruppe geeignet ist, musste ein neuer Test konstruiert werden. Im Weiteren werden zunächst verschiedene Verfahren zur Testkonstruktion diskutiert und eines für die Erstellung des Kompetenztests begründet ausgewählt. Im Anschluss daran werden die statistischen Methoden, die für die Pilotierung und die Auswertung genutzt werden, ausgewählt und dargestellt.

5.1 Verfahren zur Testkonstruktion

Bei der Entscheidung für ein Testverfahren müssen neben den Testgütekriterien Objektivität, Reliabilität und Validität¹⁹ (Moosbrugger & Kelava, 2012a) auch pragmatische Argumente, wie die mit der Erhebung verbundenen Kosten, die unter anderem durch das Erhebungsverfahren und die Stichprobengröße beeinflusst werden, berücksichtigt werden (Stecher & Klein, 1997). Bei der Auswahl muss dementsprechend ein Instrument gefunden werden, was im größtmöglichen Umfang die Testgüte garantiert, gleichzeitig aber auch in großen Proband/inn/en Gruppen ($N > 1000$ Personen, Large-Scale) ökonomisch einzusetzen ist.

5.1.1 Formate zur Erfassung von Kompetenzen

Zur Erfassung von Kompetenzen stehen verschiedene Verfahren zur Verfügung. Gängige Methoden sind u.a. die Selbsteinschätzung von Kompetenzen, die direkte Beobachtung der Proband/inn/en durch Expert/inn/en oder geschulte Beobachter/innen, sowie Paper-and-Pencil-Tests (Hartig & Jude, 2007). Für die Erfassung von Unterrichtshandlungen im Lehramt schlagen Oser, Curcio und Düggeli (2007) außerdem die advokatorische Methode vor, bei der die Handlungen der Lehrperson durch eine weitere, beobachtende Lehrperson gerechtfertigt werden. Speziell für den Bereich der Erkenntnisgewinnung, allerdings für die Schule entwickelt, nutzen Schreiber, Theyßen und Schecker (2009) Simulationsbaukästen.

Die Selbsteinschätzung von Kompetenzen ist eine einfach und schnell einzusetzende Methode zur Erfassung des Kompetenzstandes, weshalb sie auch in großen Pro-

¹⁹ Zu einer ausführlichen Diskussion der Testgütekriterien für dieses Instrument siehe Abschnitt 6.3.1.

band/inn/en Gruppen ökonomisch eingesetzt werden kann. Ihre Testgüte wird allerdings angezweifelt. So wird argumentiert, dass dieses Verfahren nicht objektiv sei, weil vielmehr das eigene Selbstkonzept erhoben würde, was zudem von der sozialen Bezugsnorm abhängig sei (Hartig & Jude, 2007). Aus diesen Gründen wird ein derartiges Verfahren für die vorliegende Studie ausgeschlossen.

Aus Gründen der Testökonomie müssen Verfahren ausgeschlossen werden, die eine weitere Person pro Proband/-in (wie z. B. Rater/innen) erfordern. Bei den angestrebten Proband/inn/en Zahlen von $N > 1000$ würde dies einen erheblichen personellen und damit finanziellen Aufwand bedeuten, weshalb dies nicht zu realisieren ist (Stecher & Klein, 1997). Dementsprechend müssen auch mutmaßlich valide Verfahren wie die direkte Beobachtung und die advokatorische Methode als Testinstrument ausgeschlossen werden. Bei letzterer kommt zudem hinzu, dass diese nicht für den Kompetenzbereich der Erkenntnisgewinnung konstruiert wurde, weshalb ein möglicher Einsatz ohnehin erst evaluiert werden müsste.

Simulationsbaukästen stellen eine ökonomisch einsetzbare Methode dar, um den Kompetenzstand von Proband/inn/en beim Experimentieren zu erheben (Schreiber et al., 2009). Hierbei wird die Validität einer direkten Beobachtung mit der Ökonomie eines Paper-and-Pencil-Tests verbunden. Allerdings ist hierbei auch von hohen Entwicklungskosten auszugehen. Dieses Verfahren wäre für die Erfassung der Kompetenzen in der Kompetenzfacette „Planung und Durchführung“ (von naturwissenschaftlichen Untersuchungen) ein sehr geeignetes Verfahren. Bei den anderen sechs Kompetenzfacetten weist dieses Verfahren aber keine Vorteile gegenüber einem Paper-and-Pencil-Test auf, was bedeutet, dass diese Aspekte des Kompetenzmodells weiterhin durch andere Verfahren (z. B. Paper-and-Pencil Items) erfasst werden müssten. Aufgrund der nur teilweisen Passung auf das Kompetenzmodell und der erwarteten hohen Entwicklungskosten wurde auch von diesem Verfahren Abstand genommen.

5.1.2 Entscheidung für ein Erhebungsverfahren

Die Entscheidung fiel schlussendlich auf einen Paper-and-Pencil-Test, da dieses Verfahren auch in Large-Scale-Studien ökonomisch einzusetzen ist (Stecher & Klein, 1997) und aufgrund der hohen Standardisierung eine gute Auswertungsobjektivität bietet (Moosbrugger & Kelava, 2012a). Außerdem handelt es sich hierbei um ein etabliertes Verfahren zur Kompetenzmessung (Hartig & Jude, 2007). Dabei ist aber zu berücksichtigen, dass

angezweifelt wird, ob hierdurch tatsächlich valide Messungen der Kompetenz möglich sind (Schreiber et al., 2009). Daher muss bei der Testkonstruktion die Validität durch verschiedene Verfahren abgesichert werden. Dieser Aspekt wird in Abschnitt 6.3.1 ausführlich diskutiert.

Nach der Entscheidung für einen Paper-and-Pencil-Test stellte sich weiterhin die Frage, welches Antwortformat für den Test geeignet ist. Häufig werden offene oder geschlossene Formate eingesetzt. Offene Formate erfordern mehr Testzeit, da die Proband/inn/en selber einen Text schreiben müssen. Außerdem ist der Auswertungsprozess aufwändiger, und die Objektivität ist fraglich (Hartig & Jude, 2007). Geschlossene Formate (Multiple-Choice-Test (MC)) sind schnell und objektiv auszuwerten, allerdings erfordert die Konstruktion der richtigen (Attraktor) und der falschen (Distraktor) Antwortmöglichkeiten erhebliche Sorgfalt, um zu verhindern, dass bestimmte Optionen alleine durch Raten ausgeschlossen werden können (Riese & Reinhold, 2014). Es konnte aber gezeigt werden, dass bei einer Konstruktion der MC-Optionen aus echten offenen Antworten durchaus valide Optionen entstehen können (Sadler, 1998). Die Entscheidung fiel daher auf ein geschlossenes Antwortformat. Wesentliche Gründe für diese Entscheidung waren der recht große Umfang des Kompetenzstrukturmodells mit sieben Zellen sowie die durch den Einbezug von drei Fächern bedingte inhaltliche Breite. Beide Dimensionen des Modells müssen ausreichend innerhalb eines Testheftes abgedeckt werden, insbesondere zur Abbildung der innerhalb der Kompetenzdefinition geforderten *Generalisierbarkeit über ähnliche Situationen* (Hartig & Klieme, 2006). Es bedurfte demnach eines Antwortformats, welches durch die Proband/inn/en schnell zu bearbeiten ist. Dem mutmaßlichen Nachteil von MC-Aufgaben in Bezug auf die Validität sollte durch eine sorgfältige Konstruktion der Antwortoptionen begegnet werden. Insbesondere wurde der Aufwand einer solchen Konstruktion geringer eingeschätzt als die Auswertung von offenen Aufgaben. Ein zusätzlicher Punkt, der für die MC-Optionen sprach, war die hohe Objektivität (vgl. Hartig & Jude, 2007; siehe auch Stecher & Klein, 1997).

5.1.3 Itemkonstruktion

Nach der Entscheidung für ein Itemformat müssen die Items für den Kompetenztest selber konstruiert werden. Ein Kompetenztest „[...] is viewed as a sample of tasks from this large domain“ (Shavelson, 2013, S. 40). Die Items, aus denen der Test besteht, können dementsprechend immer nur einen Teil von Situationen aus dem untersuchten Kompetenzbereich berücksichtigen.

Die Items werden auf Grundlage des jeweiligen Kompetenzstrukturmodells entwickelt. Sie müssen in einen entsprechenden Kontext eingebettet werden, um dem in der Kompetenzdefinition festgelegten situativen Charakter gerecht zu werden (Mayer & Wellnitz, 2014). Ein Item muss auf eine bestimmte Kompetenzfacette des Modells abzielen und die dafür beschriebenen Kompetenzen umsetzen. Schecker und Parchmann (2006) stellen dazu fest, dass „ein psychometrisch optimales Item [...] sich einer ganz bestimmten Zelle der Kompetenzmatrix zuordnen“ (S. 62) lässt. Entsprechend abgegrenzt müssen Itemstamm und -impuls formuliert sein. Darauf ist ohnehin zu achten, um auszuschließen, dass andere Lösungswege als die intendierten möglich sind (Schecker & Parchmann, 2006). Zusätzlich müssen die Aufgaben so formuliert sein, dass eine Konfundierung mit anderen Kompetenzausprägungen ausgeschlossen ist (Schecker & Parchmann, 2006). Wie oben dargestellt, ist bei der Nutzung von MC-Items besonderes Augenmerk auf die sorgfältige Konstruktion von Attraktor und den Distraktoren zu legen. Letztere müssen Personen, die nicht über die zum Lösen erforderliche Kompetenz verfügen, als plausible Antwortoption(en) erscheinen (Moosbrugger & Kelava, 2012b), wodurch die Möglichkeit des Ratens reduziert werden soll. Die drei Hauptaspekte, die bei der Auswahl der Distraktoren beachtet werden müssen, sind ihre Auswahlwahrscheinlichkeit (Attraktivität), die Ähnlichkeit mit dem Attraktor und ihre eigene Plausibilität (Moosbrugger & Kelava, 2012b). Um diese Punkte zu erfüllen, sind als Quelle der falschen Antwortmöglichkeiten echte Antworten der Proband/inn/en der ‚Fantasie‘ des Aufgabenstellers vorzuziehen. Echte, aber falsche Antworten sind meistens authentischer und verhindern dadurch die Möglichkeit, die Aufgaben durch Ausschluss von Antworten zu lösen (Bortz & Döring, 2006). Sadler (1998) nutzte z.B. Interviews zur Generierung von Distraktoren und konnte so einen validen und reliablen MC-Test konstruieren. Neben Interviews wird von anderen Autoren (Bortz & Döring, 2006; Moosbrugger & Kelava, 2012b) vorgeschlagen, die Aufgaben zunächst im offenen Format zu entwickeln und einzusetzen. Die Antworten werden anschließend analysiert und ausgewertet. Die so gewonnenen offenen Antworten der Proband/inn/en können dann als Grundlage für die Konstruktion der MC-Optionen genutzt werden. Die offenen Antworten der Proband/inn/en müssen vor der Nutzung sprachlich vereinheitlicht werden (Burton, Sudweeks, Merrill & Wood, 1991).

Während des Konstruktionsprozesses der Items müssen diese laufend durch Expert/inn/en bewertet werden, „*die sich sowohl mit dem Modell als auch mit der unterrichtlichen Praxis auskennen.*“ (Schecker & Parchmann, 2006, S. 62).

5.2 Statistische Verfahren zur Auswertung

Zur Auswertung von Fragebogen-Studien stehen im Wesentlichen zwei Verfahren zur Verfügung. Zunächst die Klassische Testtheorie (KTT), in der, grob gesagt davon ausgegangen wird, dass sämtliche Aufgaben eines Leistungstests dasselbe messen. Zum anderen die probabilistische Testtheorie (PTT) (bzw. auch Item-Response-Theory (IRT)²⁰).

„Der Hauptunterschied zwischen der KTT und der PTT [bzw. IRT] liegt in der messtheoretischen Fundierung. Die KTT nimmt an, dass der Summenwert einer Person in einem Test einen Messwert darstellt und dieser Summenwert Intervallskalenniveau besitzt. Im Rahmen der PTT [bzw. IRT] wird dagegen geprüft, ob es sich bei der Summation der Items überhaupt um eine gültige Verrechnungsvorschrift handelt (Gütekriterium der Skalierbarkeit).“

(Bühner, 2011, S. 300)

Mit Modellen der IRT wird auf das dahinter liegende latente Konstrukt geschlossen. Die daraus gewonnenen Personenparameter werden als Angabe über die Ausprägung des latenten Konstruktes (z. B. einer Kompetenz) interpretiert (Wu & Adams, 2007, S. 18ff.). Diese Modelle können die unterschiedlichen Schwierigkeiten von einzelnen Aufgaben berücksichtigen (das Rasch-Modell bzw. das 1-Parameter-Logistische-Modell (1-PL)) (Neumann, 2014), aber auch unterschiedliche Trennschärfen (das 2-PL-Modell) sowie Rate-Parameter (3-PL) (Moosbrugger, 2012). Bei der Messung von latenten Konstrukten (wie einer Kompetenz), ist die IRT der KTT vorzuziehen (Klieme & Hartig, 2008; Neumann, 2014). Der Vorteil der IRT speziell für die Kompetenzdiagnostik liegt in dem hergestellten Zusammenhang zwischen dem Lösungsverhalten bei der Bearbeitung der entsprechenden Items und dem zugrunde liegenden Kompetenzkonstrukt (Klieme & Hartig, 2008).

„Anders als in der KTT kann mit Hilfe der PTT [bzw. IRT] eine konkrete Verhaltensvorhersage getroffen werden: Im Rahmen der PTT [bzw. IRT] kann ermittelt werden, mit welcher Wahrscheinlichkeit eine Person bei Kenntnis der Itemschwierigkeiten und der Personenfähigkeit ein Item löst. Dies ist möglich, da im Rahmen der PTT [bzw. IRT] ein psychologisch plausibler Zusammenhang zwischen Personenfähigkeit und Itemlösungswahrscheinlichkeit angenommen wird.“

(Bühner, 2011, S. 300)

Durch die Darstellung von Personenfähigkeiten und Aufgabenschwierigkeiten auf einer Skala kann ein Zusammenhang zwischen der Bearbeitung einer einzelnen Aufgabe und dem Kompetenzstand der Proband/inn/en hergestellt werden (vgl. auch Boone, Staver & Yale, 2014). Dies kann zur Beschreibung von Kompetenzniveaus genutzt werden (Klieme

²⁰ Im Folgenden wird die International gebräuchliche Abkürzung IRT verwendet.

& Hartig, 2008). Ein weiterer Vorteil ist, dass es im Rahmen des RASCH-Modells (als ein Modell der IRT, siehe Abschnitt 5.2.1) möglich ist, trotz fehlender Werte – systematisch oder zufällig – alle Proband/inn/en auf einer Skala darzustellen. Diese Technik ermöglicht den Einsatz von systematisch unvollständigen Testheftdesigns (Multi-Matrix-Design), sodass den Proband/inn/en nur ein Teil der Aufgaben zur Bearbeitung vorgelegt werden muss (Boone et al., 2014). Aufgrund dieser Vorteile der IRT gegenüber der KTT im Bereich der Kompetenzdiagnostik, wird im Folgenden auf die IRT zurückgegriffen. Als Messmodell wird das RASCH-Modell ausgewählt, das als einziges IRT-Modell die Anforderungen an eine objektive Messung erfüllt und auch höhere Anforderungen an die Daten stellt (siehe Boone et al., 2014, S. 453 ff.). Gleichzeitig ist es ein etabliertes Modell, das auch z. B. in PISA erfolgreich eingesetzt wurde (OECD, 2009).

5.2.1 Das RASCH-Modell

Das einfachste Modell der IRT ist das RASCH-Modell. Dieses ist wie folgt definiert:

$$P(x_{vi}) = \frac{e^{(x_{vi}(\xi_v - \sigma_i))}}{1 + e^{(\xi_v - \sigma_i)}}$$

Dabei gibt $P(x_{vi})$ die Wahrscheinlichkeit an, dass ein/e Proband/in v das Item i löst, mit ξ_v als individuellem Fähigkeitsparameter eines/-r Probanden/-in v und σ_i den Schwierigkeitsparameter des Items i . Die Steigung dieser logistischen Funktion ist für alle Aufgaben gleich, lediglich ihre Position kann sich auf der σ -Achse verschieben. Damit haben alle Items die gleiche Trennschärfe, d. h. die Items trennen alle gleich gut zwischen verschiedenen Fähigkeitsausprägungen der Personen (Rost, 2004, S. 98). Die Items können aber unterschiedliche Schwierigkeiten annehmen (Moosbrugger, 2012). Das RASCH-Modell, wie oben dargestellt, geht also davon aus, dass die Wahrscheinlichkeit, ein Item zu lösen, nur von der Personenfähigkeit und der Aufgabenschwierigkeit abhängt (Itemhomogenität) (Moosbrugger, 2012, S. 229). Verschiedene Punkte können diese Annahme verletzen, weshalb es gilt, diese bei der Testkonstruktion zu berücksichtigen. Dazu zählen die Vermeidung von Raten bei der Aufgabebearbeitung sowie die Beachtung lokaler Unabhängigkeit der Items. Die Beantwortung eines Items darf also nicht von der Bearbeitung eines anderen abhängig sein (Wu & Adams, 2007, S. 74).

RASCH-Analysen sind auch mit kleinen Stichproben ab ca. 16 Proband/inn/en möglich. Durch eine Erhöhung der Stichprobe wird die Itemkalibrierung stabiler. Ab 150 Personen

geht man davon aus, dass die Stabilität zu 99% in einem Intervall von $\pm 1/2$ -Logit liegt (Linacre, 1994). Da das RASCH-Modell Personen und Items nicht unterscheidet, gelten die selben Bedingungen auch für die Anzahl der Items (Linacre, 1994).

Im Rahmen der RASCH-Analyse stehen verschiedene Statistiken zur Verfügung, um zu prüfen, ob einzelne Items nicht den Bedingungen des RASCH-Modells entsprechen. Auf Grundlage der Infit-Parameter weighted Mean Square (wMNSQ), T-Wert, sowie der klassischen Trennschärfe eines Items und einer optischen Kontrolle der Übereinstimmung des durch das RASCH-Modell angenommenen Verlaufs der Lösungswahrscheinlichkeit (ICC = Item characteristic curve) und dem empirischen Verlauf der Lösungswahrscheinlichkeit des Items, kann eine Abweichung detektiert werden. Der Fit (Passung) der Items auf das RASCH-Modell wird in unstandardisierter Form (als wMNSQ) und als standardisierte T-Statistik ausgegeben (Bond & Fox, 2007, S. 193). Der wMNSQ ist ein Maß für die Abweichung des theoretischen, durch das RASCH-Modell angenommenen, Verlaufs der Lösungswahrscheinlichkeit von dem empirisch beobachteten Verlauf (Bond & Fox, 2007, S. 57). Der Erwartungswert des wMNSQ liegt bei 1. Werte zwischen 0,8 und 1,2 werden als akzeptabel angenommen (Wright & Linacre, 1994). Größere Abweichungen deuten auf ein Problem mit der Aufgabe hin, welches dann in einer inhaltlichen Analyse identifiziert werden muss. Der wMNSQ ist aber stichprobenabhängig, weshalb bei ausreichend großer Proband/inn/enzahl für jedes Item der wMNSQ gegen 1 geht. Durch eine Transformation in eine T-Verteilung kann die Stichprobengröße mit berücksichtigt werden. Die T-Verteilung hat einen Mittelwert von 0 und eine Standardabweichung von 1. Abweichungen $T > 2$ oder $T < -2$ zeigen eine signifikante ($p < 0,05$) Abweichung von dem Modell an (Bond & Fox, 2007, S. 239 ff.).

Die klassische Trennschärfe gibt an, inwiefern das Abschneiden in dem spezifischen Item mit dem Abschneiden in dem gesamten Test zusammenhängt. Hierbei sollten die Werte mindesten 0,2 betragen (Wu & Adams, 2007, S. 64). Generell sollten Items mit höherer Trennschärfe genutzt werden.

Die Infit-Parameter geben nur durchschnittliche Abweichungen der empirischen Daten vom RASCH-Modell an. Eine optische Überprüfung der ICC (durch das RASCH-Modell angenommener Verlauf der Lösungswahrscheinlichkeit, s. o.) mit dem tatsächlichen (empirischen) Verlauf der Lösungswahrscheinlichkeit kann ebenfalls Probleme der Aufgabe aufdecken (Bond & Fox, 2007, S. 243f.; Embretson & Reise, 2000, S. 234) (so zum Beispiel eine nicht modellkonforme, hohe Lösungswahrscheinlichkeit bei geringen Personen-

fähigkeiten). Generell sollte bei einem „guten“ Item die theoretische und die empirische Kurve nahe beieinander liegen (Wu & Adams, 2007, S. 69). (Als Anhaltspunkte für die Beurteilung, inwiefern eine deutliche Abweichung vorliegt, können die Beispiele aus Bond und Fox (2007, S. 254f.) dienen). Grundsätzlich gilt bei der Itemselektion aber: „[...] in real-life, no item fits the Rasch model perfectly!“ (Wu & Adams, 2007, S. 83). Perfekte Passungen sind also weder bei den Fit-Statistiken noch bei der Analyse der ICC zu erwarten.

Je nach Zweck der Erhebung muss die Itemselektion nach unterschiedlichen Kriterien erfolgen. In einer Pilotierung ist es das Ziel, einen Test zu konstruieren, der das zu beobachtende Konstrukt möglichst gut abbildet und daher die besten Items identifizieren sollte. In einer Studie steht das Testinstrument hingegen fest. Das Ziel muss daher sein, aus den vorhandenen Items die meisten Informationen zu erhalten. Entsprechend erfolgt die Itemselektion weniger restriktiv als bei der Pilotierung (Wright & Linacre, 1994). Idealerweise werden alle Items des Tests in die Auswertung mit einbezogen, was bei einer ausführlichen Testkonstruktion auch der Fall sein sollte.

Trotzdem sollten auch in einer Hauptstudie die Itemfit-Indices auf Passung überprüft werden. Items werden dann auf Grundlage der Trennschärfe nur noch aussortiert, wenn diese negativ ist. Auch beim optischen Abgleich von ICC mit der Empirie wird defensiv vorgegangen und Items nur ausgeschlossen, wenn die Verläufe gegensätzlich sind. Bezüglich der wMNSQ-Werte werden nur Items mit Werten $> 2,0$ entfernt, da diese die Messung erheblich stören würden. Items mit wMNSQ-Werten $< 1,0$ werden gar nicht entfernt (Wright & Linacre, 1994). Entsprechend werden Items auch bei T-Werten von $T < -2,0$ nicht entfernt, da „even an item with a very low mean-square tells us a little something that is new and useful“ (Wright & Linacre, 1994).

5.2.2 Normierung

Die im Rahmen dieser Arbeit untersuchten Fragestellungen beziehen sich auf die Teilstichprobe der Studierenden mit einem Fach Physik. Zur Auswertung der erhaltenen Daten werden die Aufgaben zunächst aber auf Basis der gesamten Stichprobe (Studierende der Fächer Biologie, Chemie und Physik) normiert. Dieses Vorgehen bietet den Vorteil, dass dadurch eine sehr genaue Itemparameterschätzung möglich wird (Linacre, 1994). Die Normierung wird so vorgenommen, dass die Summe der Personenparameter gleich Null

gesetzt und die Itemschwierigkeiten daran ausgerichtet werden. Dadurch wird eine normorientierte Interpretation der Personenparameter möglich (Rost, 2004, S. 395). Die Normierung muss unabhängig für ein- und mehrdimensionale Modelle vorgenommen werden.

5.2.3 Personenparameterschätzung

Für auf RASCH-Modellen basierende Auswertungen können zur Schätzung von Personenfähigkeiten verschiedene Parameter genutzt werden. In verschiedenen Simulationsstudien (Monseur & Adams, 2009; Wu, 2005) konnte gezeigt werden, dass die häufig genutzten Personenschätzer Weighted Likelihood Estimates (WLE), Maximum Likelihood Estimates (MLE) und Expected A Posteriori Measures (EAP) verschiedene Probleme aufzeigen, wenn damit Gruppenparameter berechnet werden sollen. Für einen derartigen Fall wird die Nutzung von *Plausible Values* (PVs) empfohlen. PVs sind zufällig gezogene Werte aus der Wahrscheinlichkeitsverteilung der Fähigkeitsparameter eines/-r jeden einzelnen Probanden/-in (Wu, 2005, S. 116). PVs sind aber keine personenbezogenen Testwerte und dürfen dementsprechend nicht zur Individualdiagnose herangezogen werden. Pro Proband/in werden fünf PVs²¹ gezogen. Jede Analyse muss für jeden PV durchgeführt werden. Anschließend werden die Ergebnisse gemittelt (Monseur & Adams, 2009, S. 6). Dies ist ein ideales Vorgehen. Simulationsstudien zeigten aber, dass schon Berechnungen auf Grundlage eines PVs adäquate Gruppenparameter liefern (Wu, 2005, S. 116).

Die Kombination mehrerer PVs und ihrer Varianz berechnet sich folgendermaßen:

Der Mittelwert einer Statistik berechnet sich nach (OECD, 2009)

$$\theta = \frac{1}{M} \sum_{i=1}^M \theta_i$$

wobei M die Anzahl der PVs ist und θ die berechnete Statistik.

Die Gesamtvarianz berechnet sich ebenfalls als Mittelwert aus den Einzelvarianzen:

$$\sigma_{(\theta)}^2 = \frac{1}{M} \sum_{i=1}^M \sigma_{(\theta_i)}^2$$

²¹ Dies ist die Standardeinstellung der Analysesoftware Conquest und wurde auch in PISA so genutzt OECD (2009). Grundsätzlich ist aber auch eine abweichende Anzahl davon möglich.

Die Messvarianz [*measurement variance*] (B) berechnet sich nach

$$B_M = \frac{1}{M-1} \sum_{i=1}^M (\theta_i - \theta)^2$$

Beide Varianzen werden zur Fehlervarianz [*error variance*] (V) wie folgt kombiniert:

$$V = \sigma_{(\theta)}^2 + \left(1 + \frac{1}{M}\right) B_M$$

Die Standardabweichung der Gesamtgruppe entspricht dann der Quadratwurzel aus der Fehlervarianz

$$\sigma_{(Ges\theta)} = \sqrt{V}$$

(Davier, Gonzales & Mislevy, 2009; OECD, 2009)

Der Standardfehler eines Mittelwerts berechnet sich nach

$$SE = \frac{\sigma}{\sqrt{N}},$$

wobei N die Stichprobengröße ist (Bühner & Ziegler, 2009, S. 156).

Sollen Gruppenunterschiede wie Geschlecht, Semester etc. auf Grundlage von PVs eingeschätzt werden, so müssen die jeweiligen Variablen bereits bei der IRT-basierten Schätzung der Personenfähigkeiten als Prädiktoren in ein latentes Regressionsmodell einfließen (s. u.) (Wu, 2005, S. 124).

5.2.4 Latentes Regressionsmodell

Eine Regression ist eine Methode zur Vorhersage einer abhängigen Variable in Abhängigkeit von einer oder mehrerer unabhängiger Variablen. Dieses kann als lineares Modell aus den erhaltenen Messdaten ermittelt werden (Field, 2009, S. 198). Ein latentes Regressionsgrundmodell wird direkt innerhalb des IRT-Modells geschätzt, nicht erst auf Grundlage der Personenparameter. Dies umgeht Probleme, die durch die verschiedenen Personenschätzer resultieren können (Adams & Wu, 2010). Als Ergebnisse erhält man Regressionsgewichte, die angeben, um wieviel sich die abhängige Variable ändert, wenn sich die unabhängige Variable um eine Einheit ändert (Field, 2009, S. 204). Liegen die interessierenden Variablen in nominaler oder ordinaler Form vor (z. B. Geschlecht), müssen diese vorher dummy-codiert werden (Bortz, 2005, S. 483).

5.2.5 Dimensionsanalyse

Ein eindimensionales RASCH-Modell geht davon aus, dass allen Items dieselbe latente Variable zugrunde liegt. Gibt es aber theoretische Gründe, eine mehrdimensionale Struktur der Daten anzunehmen, so lässt sich dies mit mehrdimensionalen RASCH-Analysen modellieren. Mit Hilfe der Deviance-Statistik und Informationskriterien lässt sich die Passung der Modelle vergleichen. Dabei wird davon ausgegangen, dass kein Modell die Daten perfekt beschreibt, stattdessen sucht man das Modell, welches die Daten am besten beschreibt (Bühner, 2011, S. 541). Die Deviance-Statistik dient als Maß dafür, wie gut das Modell auf die Daten passt (Wu, Adams, Wilson & Haldane, 2007). Ein kleinerer Wert steht hier für eine bessere Passung. Durch einen χ^2 -Differenzentest lässt sich prüfen, ob die bessere Passung nach der Deviance-Statistik signifikant ist. Allerdings ist dies an die Bedingung geknüpft, dass die Modelle genestet sind, d. h. ein Modell muss das Obermodell des anderen getesteten Modells sein (Bühner, 2011, S. 348f.).²²

Das Nutzen von mehrdimensionalen Modellen geht mit einer Zunahme der Freiheitsgrade einher, was dem Prinzip der Einfachheit von Modellen widerspricht. Daher bieten sich informationstheoretische Maße an, die jeden zusätzlichen Freiheitsgrad bestrafen (Bühner, 2011, S. 541). Das bedeutet, dass die bessere Passung des Modells ins Verhältnis zur Zunahme der Komplexität des Modells gesetzt wird. Bei einer großen Anzahl von Items wird das *Bayesian Information Criterion* (BIC) als geeignetes Maß genannt (Rost, 2004, S. 344). Es wird außerdem angeraten, dieses in Kombination mit dem *Consistent Akaike Information Criterion* (CAIC) einzusetzen (Bühner, 2011, S. 353).

5.2.6 Differential Item Functioning (DIF)

Ein „Differential Item Functioning“ (im weiteren „DIF“) liegt vor, wenn Personen aus zwei Gruppen bei bestimmten Items trotz gleicher Personenfähigkeit eine unterschiedliche Wahrscheinlichkeit aufweisen, das Item zu lösen (Embretson & Reise, 2000, S. 252). Bestimmte Gruppen werden also durch diese Items systematisch bevorzugt oder benachteiligt. Das Item weist für die beiden Gruppen unterschiedliche Itemfunktionen auf. Zur Detektion von DIF stehen verschiedene Verfahren zur Verfügung, so zum Beispiel eine auf dem RASCH-Modell basierende Technik zur Schätzung von DIF, sowie die Mantel-Haenszel-Methode (MH). Letztere ist eine lang etablierte Methode zur Schätzung von DIF. Während andere Methoden dazu tendieren, Items mit DIF zu identifizieren, auch

²² Der χ^2 -Differenzentest ist u.a. auch daran geknüpft, dass jedes einzelne Antwortmuster mindestens einmal im Datensatz auftaucht. Dieses ist sehr unwahrscheinlich, weshalb die Ergebnisse kritisch betrachtet werden müssen (Bühner (2011, S. 348))

wenn kein DIF vorliegt, so soll die MH die besten Erkennungsraten vorweisen (GÓMEZ-BENITO & Navas-ara, 2000). Üblicherweise werden in der MH *matched groups* verglichen, die auf Basis der Rohwerte zusammengestellt werden. In der hier benutzten Software Conquest kann die MH stattdessen mit PVs durchgeführt werden. Dieser Ansatz soll für unvollständige Testheftdesigns geeignet sein (Sun, 2012). Die einzelnen Items werden gemäß den ETS²³-DIF-Kategorien (Zwick, Thayer & Lewis, 1999) eingeteilt (Sun, 2012). Aus der Analyse ausgeschlossen werden Items, wenn sie in die Kategorie „C-Moderate to large“ fallen.

5.2.7 Mittelwertsvergleiche

Mithilfe eines T-Tests lassen sich Unterschiede zwischen zwei Mittelwerten ermitteln (Field, 2009, S. 325). Die Proband/inn/engruppen, die untereinander verglichen werden, sind unterschiedliche Personen, dementsprechend muss der T-Test für unabhängige Stichproben genutzt werden. Man spricht von einem signifikanten Unterschied der Mittelwerte, wenn die Wahrscheinlichkeit eines α -Fehlers kleiner als 5 % ist. Dies zeigt sich anhand eines p-values von $p < 0,05$. Der Test setzt eine Normalverteilung der Messwerte der Proband/inn/en in beiden Stichproben und eine annäherungsweise Homogenität in den Varianzen voraus (Field, 2009, S. 326). Bei Verletzung der Annahmen für den T-Test können parameterfreie Tests wie der Mann-Whitney-Test eingesetzt werden (Field, 2009, S. 546ff.).

Bei dem Vergleich von mehr als zwei Mittelwerten bietet sich der Einsatz einer ANOVA (Analysis of Variance) an. Diese prüft, ob mehrere Mittelwerte dem gleichen gemeinsamen Mittelwert angehören. Dies geschieht durch eine Analyse der Varianz (Field, 2009, 348f.). Voraussetzungen für die ANOVA sind ebenfalls eine annähernd homogene Varianz und Normalverteilung in den Gruppen (Field, 2009, 359).

Die gesamte Varianz einer Stichprobe unterteilt sich in die systematische (Effekt-) Varianz und der Residualvarianz. Die systematische Varianz ist der Teil der Varianz, der durch die Zugehörigkeit zu einer Gruppe erzeugt und kontrolliert wird. Die Residualvarianz hingegen umfasst unsystematische Einflüsse, die nicht erfasst wurden bzw. nicht intendiert sind (Rasch, Friese, Hofmann & Naumann, 2014). Beide Anteile der Varianz be-

²³ Educational Testing Service

rechnen sich aus den Quadratsummen und den jeweiligen Freiheitsgraden. Für die systematische (Effekt-)Varianz gilt:

$$\sigma_{Eff}^2 = \frac{QS_{Eff}}{df_{Eff}}$$

QS_{Eff} wird aus dem Abstand des Mittelwertes zum Gesamtmittelwert berechnet.

Entsprechend gilt für die Residualvarianz

$$\sigma_{Res}^2 = \frac{QS_{Res}}{df_{Res}}$$

QS_{Res} berechnet sich als Summe aus den Abständen eines jeden einzelnen Wertes zu seinem Gruppenmittelwert.

Der F-Wert errechnet sich aus dem Verhältnis beider Varianzen (Rasch et al., 2014):

$$F = \frac{\sigma_{Eff}^2}{\sigma_{Res}^2}$$

Durch Abgleich mit kritischen F-Werten lässt sich die Signifikanz eines Effektes prüfen. (Rasch et al., 2014)

Bei nicht signifikanten Effekten muss zusätzlich die Teststärke (Power, $1 - \beta$) berechnet werden, um auszuschließen, dass die Alternativhypothese fälschlicherweise abgelehnt wurde. Üblicherweise wird in diesem Fall als Power zumindest $1 - \beta = 0,8$ gefordert (Rasch et al., 2014, S. 57).

5.2.8 Prüfung von Normalverteilung und Varianzhomogenität

Die Normalverteilung kann durch einen Kolmogoroff-Smirnoff-Test (K-S) und einen Shapiro-Wilk-Test (S-W) überprüft werden (Field, 2009, S. 144ff.). Beide vergleichen die Verteilung der Testwerte mit einer Normalverteilung bei gleichem Mittelwert und Standardabweichung. Der S-W-Test ist sensitiver als der K-S-Test und zeigt daher eher eine Abweichung von der Normalverteilung an. Wird der Test signifikant ($p < 0,05$), liegt eine signifikante Abweichung von der Normalverteilung vor. Zusätzlich können durch die Werte von Schiefe ν und Kurtosis ν Abweichungen in der Verteilung identifiziert werden. Beide Werte haben den Normwert 0. Werte niedriger als 0 deuten auf eine linksschiefe bzw. flachgipfelige Verteilung, Werte größer als Null auf eine rechtsschiefe bzw. spitzgip-

felige Verteilung hin. Ob die Abweichung signifikant ist, kann durch Z-Transformation überprüft werden:

$$Z = \frac{X}{SE_x}$$

Wird $Z > 1,96$, so ist die Abweichung signifikant ($p < 0,05$), für $Z > 2,58$ gilt $p < 0,01$. Für Proband/inn/en Gruppen bis 200 Personen sollte nur nach Signifikanzen $p < 0,01$ gesucht werden. Durch die große Stichprobe würden sonst auch geringfügige Abweichungen als Verletzung von Schiefe oder Kurtosis markiert. (Field, 2009, S. 137 f.).

Die Varianzhomogenität kann durch einen Levene-Test überprüft werden (Bühner & Ziegler, 2009, S. 257). Fällt dieser signifikant ($p < 0,05$) aus, so darf die Varianzhomogenität nicht angenommen werden.

5.2.9 Effektstärken

Beobachtete Effekte können durch die Berechnung von Effektstärken klassifiziert werden. Es handelt sich dabei um standardisierte Mittelwertsunterschiede zwischen den zwei untersuchten Gruppen (Bühner & Ziegler, 2009, S. 173). Um Effektstärken für unabhängige Stichproben (wie sie in unserem Fall vorliegen) zu errechnen, eignet sich Cohens d_s (auch Hedges g) (Bühner & Ziegler, 2009, S. 175). Man unterscheidet kleine ($d_s=0,20$), mittlere ($d_s=0,50$) und starke ($d_s=0,80$) Effekte (Cohen, 1988, zitiert nach Bühner & Ziegler, 2009, S. 177)

In mehrfaktoriellen ANOVAs eignet sich als Maß der Effektstärke das partielle η^2 (η_P^2). Diese Effektstärke ist ein Maß für die durch den Effekt produzierte Varianz, ohne die Varianz anderer Faktoren zu berücksichtigen (Pierce, Block & Aguinis, 2004).

Es gilt:

$$\eta_P^2 = \frac{QS_{Eff}}{(QS_{Eff} + QS_{Res})}$$

Angelehnt an die oben genannten Effektstärken unterscheiden sich hier ebenfalls kleine ($\eta_P^2 = 0,01$) mittlere ($\eta_P^2 = 0,09$) und große ($\eta_P^2 = 0,25$) Effekte (Field, 2009, S. 57, 389).

6. Testkonstruktion

In diesem Kapitel wird auf Grundlage des in Abschnitt 3.3 dargestellten Kompetenzmodells und der in Kapitel 5 vorgestellten Verfahren die Konstruktion der Items für den Kompetenztest²⁴ beschrieben. Beispielhaft soll an dieser Stelle die Konstruktion von vier Aufgaben zu vier verschiedenen Facetten des Kompetenzstrukturmodells dargestellt werden. Die hier präsentierten Aufgaben sind aufgrund mangelnder psychometrischer Passung oder anderer Gründe (z. B. ausreichend Items in einer Zelle) nach der Pilotierung aus dem Test ausgeschlossen worden. Da die Studie²⁵ weiterhin läuft, können an dieser Stelle keine eingesetzten Items präsentiert werden. Vom Grundprinzip her gleichen sich die Aufgaben aber.

Die Bewertung der Items folgt dem in Abschnitt 5.2.1 dargestellten Verfahren. In einem ersten Auswertungsschritt wird die Qualität des Testes und der Items evaluiert. Dazu wird eine Modellierung mit einem RASCH-Modell vorgenommen. Ziel der Pilotierung ist, aus den vorhandenen Items eine Auswahl zu treffen, die die Anforderungen des RASCH-Modells erfüllen.

Zur RASCH-Modellierung wurde die Software ACER Conquest 3.0 (Adams, Wu & Wilson, 2012) genutzt.

6.1 Vorgehen bei der Aufgabenkonstruktion

Um ein standardisiertes Aufgabenformat zu erhalten, welches alle Anforderungen an die Testkonstruktion umsetzt, wurde ein Konstruktionsmanual²⁶ auf Grundlage des Kompetenzstrukturmodells erstellt. Darin wird für jede Facette des Kompetenzmodells beschrieben, wie mögliche Aufgaben zu einer Kompetenzfacette aussehen können. Außerdem wird ein einheitlicher Impuls festgelegt. Um dem domänenübergreifenden Charakter des Kompetenztests gerecht zu werden (vgl. Kapitel 3), wird innerhalb des Impulses auf eine Zuordnung zu einer Fachwissenschaft verzichtet, stattdessen wird hier nur von „naturwissenschaftlichen“ Fragestellungen etc. gesprochen. Durch die fachlichen Inhalte wird die Zugehörigkeit zu einer Fachwissenschaft dann festgelegt. Die fachlichen Inhalte wur-

²⁴ Die Darstellung umfasst die Konstruktion der Items mit physikalischem Kontext. Es wurden äquivalente Items für die Chemie und die Biologie entwickelt und für die im Weiteren dargestellten Ergebnisse auch genutzt. Die Konstruktion dieser Items wird bei Stiller (in Vorbereitung) und Mathesius (in Vorbereitung) dargestellt.

²⁵ Es wird angestrebt eine Kohorte in einem echten Längsschnitt zu beforschen.

²⁶ Dieses Manual war Grundlage für alle Items innerhalb des Ko-WADiS-Projekts und wurde nicht innerhalb dieser Arbeit entwickelt.

den so gewählt, dass diese auch für Personen verständlich sind, die das entsprechende Fach nicht studiert haben. Dabei dienten die Standards für den mittleren Schulabschluss (Kultusministerkonferenz, 2005) als Orientierung. Dies geschah, weil die Beantwortung eines Items zur Erkenntnisgewinnung nicht vom Fachwissen der Proband/inn/en abhängen darf (Mayer & Wellnitz, 2014). Das ist gerade in diesem Projekt eine Herausforderung, da eine fächerübergreifende naturwissenschaftliche Erkenntnisgewinnungskompetenz angenommen wird (vgl. Kapitel 3). Die Items mit einem physikalischen Kontext müssen daher auch für Biologie- und Chemiestudierende lösbar sein, genauso wie die Items der anderen Fächer für Physikstudierende. Eine Konfundierung mit universitärem Fachwissen musste daher ausgeschlossen werden, weshalb bei allen Items die zur Lösung notwendigen Informationen im Stamm mitangegeben werden mussten.

Für die Konstruktion der MC-Antwortoptionen wurde ein Verfahren gewählt, das sich an Moosbrugger und Kelava (2012b) anlehnt. Allerdings wurden offene Aufgaben statt halb-offener genutzt. Um trotz der größeren Freiheit in der Gestaltung Antworten zu erhalten, die der intendierten Richtung entsprechen, mussten die Impulse entsprechend eng formuliert sein. Teilweise wurden diese mit vorgeschalteten, dichotomen Ja/Nein-Fragen ergänzt. Dieses Vorgehen sollte sicherstellen, dass die Antworten der Studierenden inhaltlich breit gestreut sind. Hinzu kommt eine schlechte Eignung bestimmter Itemformate (z. B. Beschreibungen von Experimentiersettings) für das halboffene Format. Die offenen Aufgaben wurden im Rahmen einer Prä-Pilotierung an Studierenden eingesetzt. Die erhaltenen Antworten wurden gesichtet und anschließend in richtige und falsche Antworten sortiert. Als MC-Optionen wurden dann Antworten gewählt, die idealerweise mehrfach genannt wurden, auch wenn dies nicht immer möglich war.

Die aus der Prä-Pilotierung erhaltenen Antworten lagen selbstverständlich in sprachlich völlig unterschiedlichen Formaten vor, weshalb diese nicht wortwörtlich in die Multiple-Choice-Items zu übernehmen waren. Vorher mussten Sprache und Format vereinheitlicht werden. Als Grundlage für die sprachliche Anpassung der Items diente im Wesentlichen Burton et al. (1991).

Die aus Sicht des Vorhabens wichtigsten Richtlinien waren dabei

- eine sprachlich parallele Form,
- eine ähnliche Länge,
- eine gleichmäßige Verwendung bzw. Auslassung von Schlüsselwörtern und
- die Vermeidung von Fachwörtern.

Durch die Parallelisierung der Antwortmöglichkeiten waren die Satzanfänge häufig gleich, weshalb diese nur einmal genannt wurden und der zweite Satzteil durch Ankreuzen entsprechend ergänzt werden musste (siehe auch Burton et al., 1991). Die in der Prä-Pilotierung teilweise enthaltenen Ja/Nein-Fragen wurden in die Antwortmöglichkeiten integriert. Jede Aufgabe bot vier Antwortalternativen an, von denen aber jeweils nur eine richtig war (Multiple-Choice-Single-Select).

Die Items wurden während des Konstruktionsprozesses mehrmals durch Expert/inn/en beurteilt. Der Definition von Schecker und Parchmann (2006) folgend, eignen sich für die Begutachtung der Aufgaben projektinterne Hochschullehrer/innen. Ein/e Experte/-in des eigenen Faches sowie mindestens ein/e Experte/-in eines der anderen der beiden beteiligten Fächer beurteilten sowohl die offene Form als auch die geschlossene Form der Items. Insbesondere wurde auf die adäquate Umsetzung des Konstruktionsmanuals und die fachliche Korrektheit geachtet. In der Begutachtung der MC-Items wurden besonders die MC-Optionen fokussiert. Wurden Überarbeitungen angemahnt, wurde der Erfolg dieser durch die Expert/inn/en überprüft. Durch das Hinzuziehen von fachfremden Expert/inn/en sollte sichergestellt werden, dass die Aufgaben tatsächlich unabhängig vom spezifischen Fachwissen lösbar sind. Zusätzlich wurden beide Konstruktionsstadien der Items durch einen psychometrisch ausgewiesenen Experten bewertet, der sich ebenfalls an den oben genannten Kriterien orientierte.

6.1.1 Prä-Pilotierung und Pilotierung

Die Prä-Pilotierung der offenen Aufgaben fand an der Freien Universität Berlin und der Humboldt Universität zu Berlin statt. Dabei wurden 183 offene Aufgaben, die über 20 Testhefte à 15 Aufgaben verteilt waren, bei 259 Studierenden eingesetzt. Im Anschluss daran wurden aus den Ergebnissen dieser Testung die MC-Items konstruiert. Die Pilotierung mit den fertigen MC-Aufgaben wurde zentral für das gesamte Projekt an beiden Berliner Universitäten sowie an den Universitäten in Köln, Aachen und Innsbruck durchgeführt. Insgesamt wurden 578 Testhefte eingesetzt. 59,5 % der Proband/inn/en waren weiblich, das Durchschnittsalter lag bei 23,74 (4,71) Jahren. Auf Grundlage der Pilotierung wurde ein RASCH-Modell berechnet. Die daraus erhaltenen Itemfit-Parameter dienten in Zusammenhang mit einer inhaltlichen Analyse zur Itemsselektion.

6.2 Konstruktionsbeispiele

Im Folgenden werden vier Aufgaben vorgestellt. Zum besseren Verständnis des Konstruktionsprozesses einer Aufgabe, wird jeweils pro Aufgabe chronologisch vorgegangen: Zunächst wird die offene Form der Aufgabe präsentiert, wie sie in der Prä-Pilotierung eingesetzt wurde. Anschließend wird die geschlossene Form der Aufgabe dargestellt. Zu den einzelnen MC-Optionen werden die entsprechenden Antworten der Studierenden genannt, aus denen sie abgeleitet wurden. Nach den jeweiligen geschlossenen Aufgaben werden die Ergebnisse der RASCH-Modellierung der Pilotierung für die entsprechende Aufgabe diskutiert und entsprechend eingeschätzt.

6.2.1 Aufgabenbeispiel 1: „Hypothese“

In der vorliegenden Aufgabe zur Kompetenzfacette „Hypothese“ (Dimension *Naturwissenschaftliche Untersuchungen*, siehe Abschnitt 3.4.1) sind die Proband/inn/en angehalten, eine Hypothese gemäß den oben genannten Kriterien aufzustellen, die das im Itemstamm beschriebene Phänomen erklären kann. Die Auftriebskraft ist den Studierenden aus dem Schulunterricht bekannt (vgl. Kultusministerkonferenz, 2005).

Innerhalb der Aufgabe wird den Proband/inn/en ein Alltagsphänomen (Aufsteigen von Luftblasen in Flüssigkeiten) präsentiert. Es wird dabei auf einen Effekt hingewiesen, der möglicherweise einigen Proband/inn/en auch schon bekannt ist: Luftblasen unterschiedlicher Größen steigen mit unterschiedlichen Geschwindigkeiten an die Oberfläche. Auf Grundlage dieser Beobachtung muss daraufhin eine Hypothese entwickelt werden, deren Beantwortung zur Lösung des Problems beitragen kann. Die Hypothese muss sich demnach auf diese konkrete Beobachtung und das benannte Problem beziehen.

Luftblasen in Öl

Wird eine Flasche mit Öl geschüttelt, so bilden sich Luftblasen verschiedener Größe. Die Luftblasen steigen aufgrund der Auftriebskraft nach oben. Große Luftblasen steigen sehr schnell wieder an die Oberfläche, kleine Blasen wandern nur sehr langsam nach oben.

Stellen Sie eine naturwissenschaftliche Hypothese zu diesem Phänomen auf.

Abbildung 2: Aufgabenbeispiel "Hypothese" (offen)

Luftblasen in Öl

Wird eine Flasche mit einem zähflüssigem Öl, z.B. Sonnenblumenöl, geschüttelt, so bilden sich Luftblasen verschiedener Größe. Die Luftblasen steigen aufgrund der Auftriebskraft nach oben. Große Luftblasen steigen sehr schnell wieder an die Oberfläche, kleine Blasen hingegen nur sehr langsam.

Welche naturwissenschaftliche Hypothese könnten Sie zu diesem Phänomen aufstellen?

Kreuzen Sie an.

- Eine große Menge an Luft in einer Blase bedeutet eine hohe Auftriebskraft.
- Die Zähflüssigkeit eines Stoffes hängt vom Stoff selbst und von der Größe des Fremdkörpers ab.
- Die Auftriebskraft entsteht, weil Öl eine höhere Dichte als Luft hat.
- Je größer die Luftblase ist, desto schneller steigt sie nach oben.

Abbildung 3: Aufgabenbeispiel "Hypothese" (geschlossen)

Entwicklung der Aufgabe:

In den Itemstamm wurde in der geschlossenen Version der Aufgabe die Klarstellung „z. B. Sonnenblumenöl“ zusätzlich eingefügt. Dadurch sollte der Alltagsbezug des Items deutlicher werden.

Tabelle 3: Erläuterungen zu den MC-Optionen und Beispielantworten der Studierenden zum Aufgabenbeispiel „Hypothese“

MC	
1 Attraktor	Die erste MC-Option (Attraktor) ist eine prüfbare und korrekte Hypothese, auch wenn sie fachlich falsch ist. Beispielantworten von Studierenden: <i>„Je größer die Oberfläche der Blasen ist, desto mehr sind sie mit Luft gefüllt [sic] und steigen deswegen schneller in öligen Flüssigkeiten auf.“</i> <i>„Die Auftriebskraft von Luftblasen ist abhängig von der sich darin befindlichen Luftmasse.“</i> <i>„Große Luftblasen (mehr Luft) steigen in Öl schneller auf als kleinere Luftblasen (weniger Luft) aufgrund ihres Luftverhältnisses.“</i> <i>„Eine große Menge an Sauerstoff in einer Luftblase bedeutet eine hohe Auftriebskraft.“</i>
2 Distraktor	Die zweite MC-Option (Distraktor) betrifft lediglich die Zähflüssigkeit und damit nicht die Ursache für das Phänomen. Außerdem ist sie uneindeutig formuliert, da sie zwei Ursachen für die Zähflüssigkeit prüfen will. Beispielantworten von Studierenden: <i>„Viskosität eines Stoffes hängt vom Stoff selbst ab und von der Größe des Fremdkörpers“</i>
3 Distraktor	Die dritte MC-Option (Distraktor) fokussiert ebenfalls falsch, weil nur die Ursache für die Auftriebskraft erklärt werden soll. Beispielantworten von Studierenden: <i>„[...]Luft weist eine geringere Dichte als Öl auf.“</i>
4 Distraktor	Die vierte MC-Option (Distraktor) beschreibt lediglich das Phänomen und kann es dadurch auch nicht erklären. Beispielantworten von Studierenden: <i>„[...]je größer die Luftblase erscheint, desto schneller entweicht die Luftblase nach oben. [...]“</i>

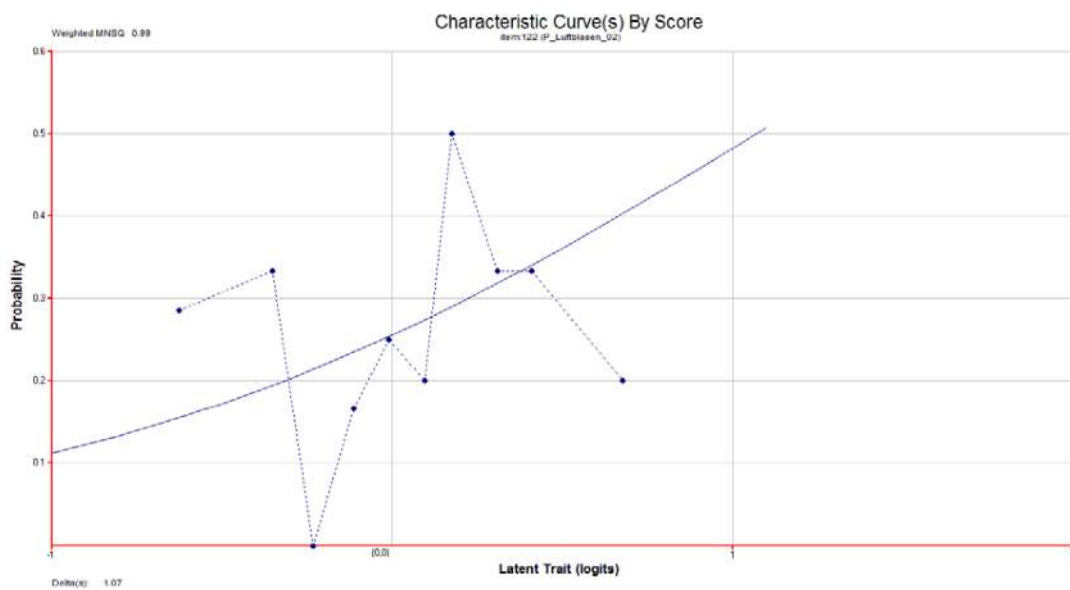


Abbildung 4: ICC zum Aufgabenbeispiel „Hypothese“

Tabelle 4: Item-Fit-Werte zum Aufgabenbeispiel „Hypothese“

Logit	wMNSQ	T	discr.
1,07	0,99	0,0	0,16

Einschätzung

Die Aufgabe weist zufriedenstellende Itemfit-Parameter auf (siehe Tabelle 4), wurde aber aufgrund der geringen Trennschärfe und der ungenügenden Übereinstimmung von theoretischem und empirischem Verlauf der ICC (siehe Abb. 4) im Rahmen der Pilotierung aus dem Test entfernt. Eventuelle inhaltliche Gründe könnten sein, dass sich Studierende mit höherem Kompetenzstand durch die etwas umgangssprachliche Formulierung des Attraktors irritiert fühlten. Außerdem könnte das Fehlen der – möglicherweise einigen Proband/inn/en bekannten – richtigen Erklärung (Anstieg der Auftriebskraft mit r^3 - Anstieg der der Stokesreibung mit r) zusätzlich zu Irritationen geführt haben.

6.2.2 Aufgabenbeispiel 2: „Planung und Durchführung“

Die Kompetenzfacette „Planung und Durchführung“ (Dimension *Naturwissenschaftliche Untersuchungen*, siehe Abschnitt 3.4.1) umfasst auch das Erkennen von Störvariablen bzw. nicht kontrollierten Variablen. Im Folgenden Beispiel wird wieder auf einen Alltagskontext Bezug genommen (Rollen auf einer schiefen Ebene). Der Einfluss unterschiedlicher Oberflächen ist zum einen aus dem Alltag bekannt, allerdings wird dieses Thema auch in der Schule thematisiert (Kultusministerkonferenz, 2005).

Den Proband/inn/en werden eine Hypothese und ein entsprechender experimenteller Aufbau genannt. Dieser soll bewertet werden, inwiefern er zur Prüfung der Hypothese geeignet ist. Um diese Aufgabe lösen zu können, müssen die Proband/inn/en aus der Hypothese zunächst die abhängige und die unabhängige Variable identifizieren. Weiterhin müssen sie diese Erkenntnisse dann auf den experimentellen Aufbau übertragen und darin eventuelle Störvariablen erkennen. Die Proband/inn/en müssen erkennen, dass der Einfluss der Oberflächen der Rampen nicht berücksichtigt wurde und daher diese als Störvariable fungieren kann.

Schiefe Ebene

Man lässt drei Bälle jeweils eine Rampe herab rollen. Die Endgeschwindigkeit wird mit einer Messeinrichtung am Fuße der Rampe bestimmt. Das Gewicht der drei Bälle und die Oberflächenbeschaffenheit sind jeweils gleich. Die erste Rampe ist höher als die zweite und dritte Rampe, davon abgesehen sind sie identisch.

Sie möchten mit einem Versuch die folgende Hypothese prüfen:

„Die Höhe einer Rampe hat Einfluss auf die Endgeschwindigkeit eines herabrollenden Balls.“

Sie nutzen dazu eine Spielplatzrutsche und eine halb so hohe, gleich lange Rampe aus Holz, die Sie selber bauen. Die Endgeschwindigkeit wird am Fuße der Rampen mit einer Messeinrichtung bestimmt.

Beurteilen Sie den beschriebenen Versuchsaufbau im Hinblick auf die zu prüfende Hypothese und begründen Sie ihre Einschätzung.

Abbildung 5: Aufgabenbeispiel "Planung und Durchführung" (offen)

Schiefe Ebene

Sie möchten mit einem Versuch die folgende Hypothese prüfen:

„Die Höhe einer Rampe hat Einfluss auf die Endgeschwindigkeit eines herabrollenden Balls.“

Sie nutzen dazu eine Stahlrutsche auf einem Spielplatz und eine halb so hohe, gleich lange Rampe aus Holz, die Sie selber bauen. Die Endgeschwindigkeit wird am Fuße der Rampen mit einer Messeinrichtung bestimmt.

Wie schätzen Sie den Versuchsaufbau im Hinblick auf die zu prüfende Hypothese ein?

Kreuzen Sie an.

- Der Aufbau ist geeignet. Man muss aber beide Male den gleichen Ball benutzen und darauf achten, dass dieser rollt und nicht rutscht.
- Der Aufbau ist geeignet. Die Rutsche muss aber 50 cm länger sein als die Rampe, damit der Effekt gut sichtbar ist.
- Der Aufbau ist ungeeignet. Der Einfluss der Oberflächen wird nicht berücksichtigt.
- Der Aufbau ist ungeeignet. Fachlich korrekter wäre es den Einfluss des Neigungswinkels auf die Endgeschwindigkeit zu untersuchen.

Abbildung 6: Aufgabenbeispiel "Planung und Durchführung" (geschlossen)

Entwicklung der Aufgabe:

Im Itemstamm wurde der erste Absatz entfernt, der fälschlicherweise in der Prä-Pilotierung in der Aufgabe enthalten war.

Tabelle 5: Erläuterungen zu den MC-Optionen und Beispielantworten der Studierenden zum Aufgabenbeispiel "Planung und Durchführung"

MC	
1 Distraktor	<p>Diese Option zeigt Ansätze von einem Verständnis zur Variablenkontrolle (Ball), allerdings bleiben die unterschiedlichen Oberflächen unberücksichtigt.</p> <p>Beispielantworten von Studierenden: <i>„Die Rotationsenergie beeinflusst den Ausgang des Experiments. Um diesen Einfluss gering zu halten, sollte man ein und denselben Ball benutzen und darauf achten, dass der Ball rollt und nicht rutscht.“</i></p>
2 Distraktor	<p>Die zweite MC-Option (Distraktor) verändert eine zusätzliche Variable im Aufbau und ist daher ebenfalls falsch.</p> <p>Beispielantworten von Studierenden: <i>„wird funktionieren, Rutsche sollte nur etwas länger sein (mind. 50 cm), um Effekt gut sehen zu können“</i></p>
3 Attraktor	<p>Diese Option erkennt den Einfluss der unterschiedlichen Oberflächen auf den Ausgang des Experiments.</p> <p>Beispielantworten von Studierenden: <i>„Der Schwachpunkt des Versuchs liegt bei der unterschiedlichen Oberflächenbeschaffung der Rutsche / Holzrampe“</i> <i>„Oberflächenbeschaffenheit der gewählten Rampen sind unterschiedlich man verändert mehr als eine Sache -> keine genaue Aussage möglich“</i> <i>„Bei dem Versuch wird nicht die Materialeigenschaft miteinbezogen. Spielplatzrutsche besteht aus Metall, selbstgebaute Rampe aus Holz. Die Reibung müsste also auch beachtet werden.“</i></p>
4 Distraktor	<p>Diese Option schätzt den Aufbau als ungeeignet ein, sieht aber das Problem in einer falschen Hypothese.</p> <p>Beispielantworten von Studierenden: <i>„[...] besser: der Neigungswinkel hat Einfluss auf...“</i></p>

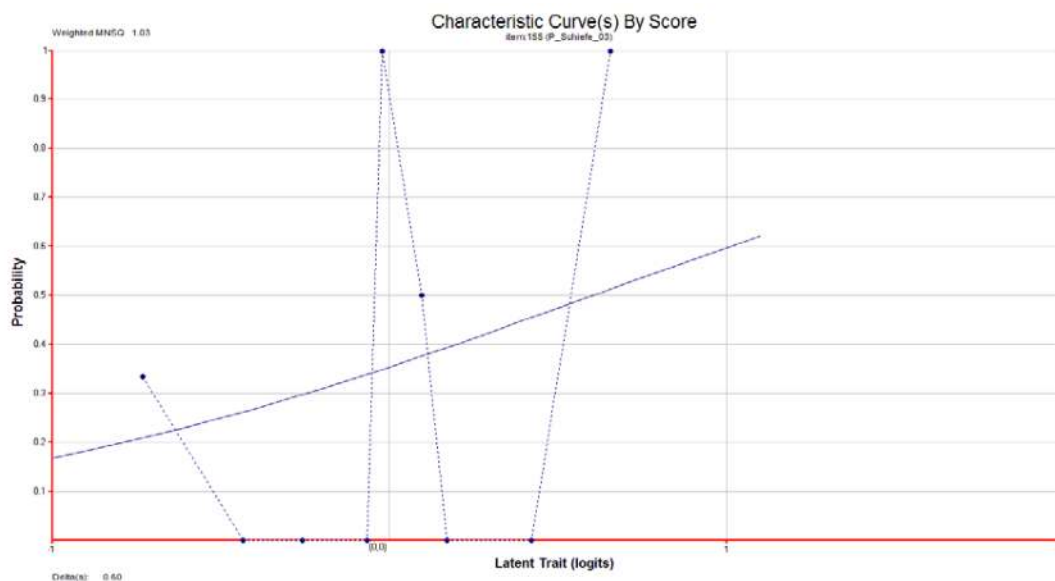


Abbildung 7: ICC des Aufgabenbeispiels "Planung und Durchführung"

Tabelle 6: Item-Fit-Werte des Aufgabenbeispiels "Planung und Durchführung"

Logit	wMNSQ	T	discr.
0,604	1,02	0,2	0,23

Einschätzung

Die Aufgabe weist zufriedenstellende Itemfit-Parameter auf (siehe Tabelle 6), auch wenn die klassische Trennschärfe etwas gering ausfällt. Der Vergleich von theoretischer ICC und dem empirischen Verlauf der Lösungswahrscheinlichkeiten ist nicht zufriedenstellend (siehe Abbildung 7). Aus diesem Grund wurde die Aufgabe aus dem Itempool entfernt. Eine Evaluation dieser Aufgabe ist schwierig, da für sie nur 15 gültige Antworten vorliegen. Möglicherweise wäre bei einer größeren Anzahl von gültigen Antworten die Aufgabe im Test verblieben. Da ausreichend Items zu dieser Facette des Kompetenzmodells vorhanden waren, wurde dies aus pragmatischen Gründen nicht weiter verfolgt.

6.2.3 Aufgabenbeispiel 3: „Auswertung und Interpretation“

Die Dimension *Naturwissenschaftliche Untersuchungen* umfasst auch die Facette „Auswertung und Interpretation“ von Daten. In der hier dargestellten Aufgabe sollen die Ergebnisse einer Messung mit einem Geiger-Müller-Zählrohr interpretiert werden. Das Thema zählt zum Schulwissen und wird bis zum Ende der 10. Jahrgangstufe im Physikunterricht behandelt (Kultusministerkonferenz, 2005). Um dieses Vorwissen zu wiederholen, werden sämtliche relevanten Aspekte innerhalb des Itemstamms wiederholt. Inhaltliche Grundlage dieser Aufgabe ist ein Experiment aus dem Schüler/innen/labor „PhysLab“ (PhysLab, 2014).

Gemäß der oben genannten Operationalisierung der Kompetenzfacette (siehe Abschnitt 3.4.1) umfassen Aufgaben dieser Facette auch die Auswertung von Daten, die im Rahmen von Experimenten gewonnen wurden. Im Fokus dieser Aufgabe stehen der Umgang mit tabellarisch dargestellten Daten und die Interpretation dieser in Bezug auf die Wirkung von Blei zwischen Präparat und Zählrohr. Dabei müssen die Proband/inn/en zunächst die tabellarische Darstellung interpretieren. Außerdem müssen Sie die Bedeutung der einzelnen Zahlenwerte (als ein Impuls) erkennen und verstehen, dass es zu statistischen Schwankungen kommt.

Radioaktivität

Ein Geiger-Müller-Zählrohr kann radioaktive Strahlung detektieren. Es gibt pro detektiertem Strahlungsteilchen einen Impuls (Imp.) an ein Zählgerät weiter. Somit können Impulse nur als ganze Zahlen auftreten. Ein detektiertes Strahlungsteilchen deutet auf einen Zerfall eines Atomkernes in der Umgebung des Geiger-Müller-Zählrohres hin. Durch in der Luft vorkommende natürliche Radionuklide wie Rn-222 besteht immer ein natürlicher Strahlungshintergrund, der mit der sogenannten Nullzählrate erfasst wird. Diese muss für jeden Ort neu bestimmt werden, weil sie stark von den Gegebenheiten der Umgebung abhängt.

Sie befestigen ein Ra-226-Präparat in einem Abstand von $s = 1$ cm vor einem Geiger-Müller-Zählrohr. Zwischen Zählrohr und Präparat werden nacheinander mehrere $b = 2$ mm starke Bleiplatten eingebracht und jeweils für den Zeitraum von $t = 60$ s die Impulse gemessen.

Sie erhalten folgende Messreihen, wobei die Nullzählrate von $N_0 = 17 \frac{\text{Imp.}}{60\text{s}}$ schon berücksichtigt ist:

Stärke der Bleiplatten	N-N ₀	N-N ₀	N-N ₀
ohne Bleiplatten	39341	40084	40034
2 mm	169	166	161
4 mm	151	142	147
6 mm	144	131	126
8 mm	123	130	133
10 mm	131	124	121

Interpretieren Sie das Ergebnis bezüglich der Wirkung von Blei zwischen Präparat und Zählrohr.

Abbildung 8: Aufgabenbeispiel „Auswertung und Interpretation“ (offen)

Radioaktivität

Ein Geiger-Müller-Zählrohr kann radioaktive Strahlung detektieren. Es gibt pro detektiertem Strahlungsteilchen einen Impuls (Imp.) an ein Zählgerät weiter. Somit können Impulse nur als ganze Zahlen auftreten. Ein detektiertes Strahlungsteilchen deutet auf den Zerfall eines Atomkerns in der Umgebung des Geiger-Müller-Zählrohres hin. Durch in der Luft vorkommende natürliche Radionuklide wie Radon (Rn-222) besteht immer ein natürlicher Strahlungshintergrund, der mit der sogenannten Nullzählrate erfasst wird. Diese muss für jeden Ort neu bestimmt werden, weil sie stark von den Gegebenheiten der Umgebung abhängt.

Sie befestigen ein Radium-Präparat (Ra-226) in einem Abstand von $s = 1$ cm vor einem Geiger-Müller-Zählrohr. Zwischen Zählrohr und Präparat werden nacheinander mehrere, jeweils $b = 2$ mm starke Bleiplatten eingebracht und jeweils für den Zeitraum von $t = 60$ s die Impulse gemessen.

Sie erhalten folgende Messreihen, wobei die Nullzählrate von $N_0 = 17 \frac{\text{Imp.}}{60\text{s}}$ schon berücksichtigt ist:

<i>Stärke der Bleiplatten</i>	N-N ₀	N-N ₀	N-N ₀
ohne Bleiplatten	39341	40084	40034
2 mm	169	166	161
4 mm	151	142	147
6 mm	144	131	126
8 mm	123	130	133
10 mm	131	124	121

Welche Interpretation lässt die Messreihe bezüglich der Wirkung von Blei zwischen Präparat und Zählrohr zu?

Kreuzen Sie an.

- Je dicker eine Bleiplatte ist, desto weniger radioaktive Strahlung wird detektiert.
- Je stärker die Bleiplatten, desto geringer ist die Stärke der gemessenen Impulse.
- Durch die Bleiplatten wird radioaktive Strahlung abgeschirmt.
- Der Schutz vor radioaktiver Strahlung steigt mit jeder zusätzlichen Bleiplatte deutlich an.

Abbildung 9: Aufgabenbeispiel „Auswertung und Interpretation“ (geschlossen)

Entwicklung der Aufgabe:

Im Itemstamm wurde die Klarstellung „Radon“ vor Rn-22 eingefügt, da dies Nicht-Physiker/innen bekannter sein dürfte.

Tabelle 7: Erläuterungen zu den MC-Optionen und Beispielantworten der Studierenden zum Aufgabenbeispiel „Auswertung und Interpretation“

MC	
1	<p>Diese Option sieht einen kontinuierlichen Rückgang der detektierten Strahlung, der aber nicht zu erkennen ist. Stattdessen kommt es im Bereich von 6mm – 10 mm zu Schwankungen.</p> <p>Beispielantworten von Studierenden: <i>„je dicker die Bleischicht, desto geringer ist die gezählte Pulsrate [...]“</i> <i>„je dicker es wird, desto weniger Strahlungen werden gemessen, da sie nicht durch das Präparat gelangen“</i> <i>„Je dicker / stärker die Bleiplatten zwischen Präparat und Zählrohr, desto weniger Impulse / Strahlenteilchen werden von dem Geiger-Müller-Zählrohr detektiert.“</i></p>
2	<p>Diese Option sieht ebenfalls den kontinuierlichen Rückgang. Zusätzlich interpretiert sie die Messdaten falsch, da von „Stärke von Impulsen“ die Rede ist.</p> <p>Beispielantworten von Studierenden: <i>„Je stärker die Bleiplatte, desto geringer ist der Impuls (detektierten [sic] Strahlungsteilchen), welche erfasst werden können.“</i></p>
3	<p>Diese Option erkennt die abschirmende Wirkung von Bleiplatten. Sie stellt keine weiteren quantifizierenden Ergebnisse dar, weshalb diese Antwort die richtige ist.</p> <p>Beispielantworten von Studierenden: <i>„Bleiplatten mildern Strahlung abschirmend ab (aber nicht vollständig).“</i> <i>„Blei lässt wenig Strahlung passieren.“</i> <i>„Blei schirmt Strahlung ab.“</i> <i>„Blei hat eine abschirmende Wirkung bezüglich radioaktiver Strahlung.“</i></p>
4	<p>Diese Option sieht einen Zusammenhang zwischen der Stärke der Bleiplatten und Schutz vor radioaktiver Strahlung. Zusätzlich nutzt diese eine mutmaßliche Alltagsvorstellung von „Strahlenschutz“</p> <p>Beispielantworten von Studierenden: <i>„mit steigender Dicke → mehr Strahlenschutz“</i></p>

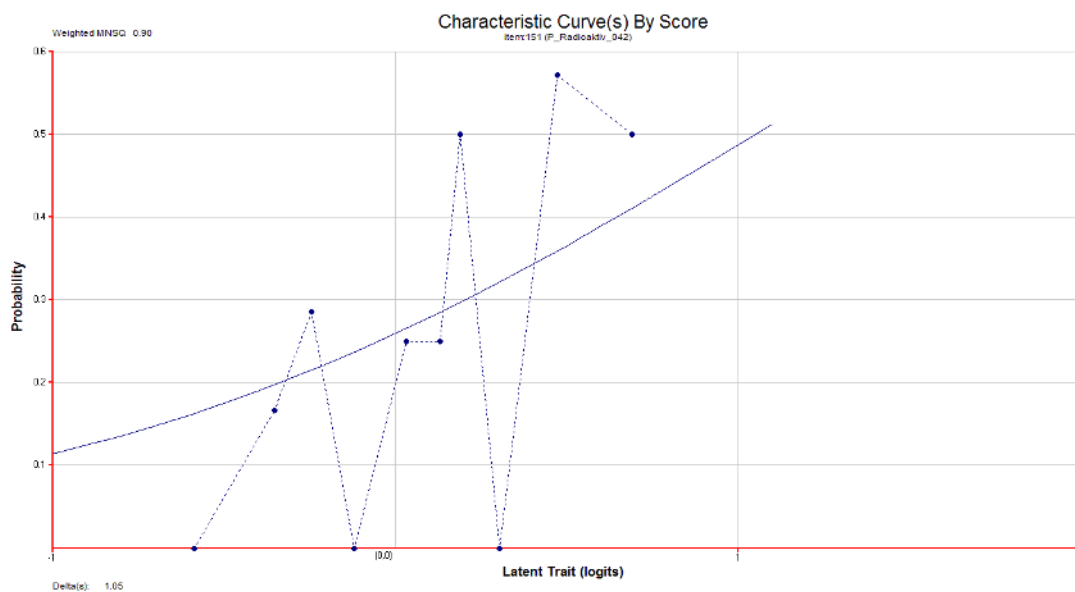


Abbildung 10: ICC des Aufgabenbeispiels „Auswertung und Interpretation“

Tabelle 8: Item-Fit-Werte des Aufgabenbeispiels „Auswertung und Interpretation“

Logit	wMNSQ	T	discr.
1,052	0,90	-0,7	0,31

Einschätzung:

Die Itemfit-Parameter sind insgesamt zufriedenstellend (siehe Tabelle 8). Der Blick auf die Verläufe der theoretischen ICC und den empirischen Verlauf der Lösungswahrscheinlichkeiten (siehe Abbildung 10) offenbart aber Probleme mit dieser Aufgabe. Sie wurde daher aus dem Itempool entfernt. Das Problem dieser Aufgabe könnte darin liegen, dass die Antwortoption eins durch ihre je-desto-Formulierung eine hohe Attraktivität besitzt, die korrekte Antwort drei möglicherweise einigen Studierenden zu banal erschien (siehe Tabelle 7). Ein weiteres Problem besteht darin, dass hier keine Hypothese vorgegeben und stattdessen eine explorative Interpretation der Daten erwartet wird. Eine Weiterentwicklung dieser Aufgabe könnte demnach eine prüfbare Hypothese enthalten über deren Stützung oder Ablehnung dann die Proband/inn/en entscheiden müssen.

Aufgabenbeispiel 4: „Zweck von Modellen“

In Bezug auf die Kompetenzfacette „Zweck von Modellen“ (Dimension *Modelle nutzen*, siehe Abschnitt 3.4.2) soll der Zweck eines naturwissenschaftlichen Modells erkannt werden. In diesem Fall handelt es sich um ein theoretisches Modell. Die Thematik der Massepunkte wird teilweise in der Schule behandelt (Kultusministerkonferenz, 2005). Das Modell wird daher detailliert beschrieben, um den Zweck zu erkennen.

Die Proband/inn/en sollen in diesem Fall aus der genannten Beschreibung dieses Modells seine Einsatzmöglichkeiten erkennen. Mögliche Antworten können z. B. in die Richtung zielen, dass dieses Modell dazu eingesetzt werden kann, Voraussagen über das Verhalten von Körpern abzuleiten. Andererseits wären auch Antworten möglich, die auf den Abgleich mit der Natur zielen, z. B. die Abschätzung von Kräften.

Massepunkt										
In der Mechanik werden räumlich ausgedehnte Körper modellhaft häufig als Massepunkte betrachtet. Dabei wird angenommen, dass sich die gesamte Masse eines Körpers im Schwerpunkt des Körpers konzentriert. Dies ermöglicht mathematische Rechnungen bezüglich des Verhaltens von komplexen Körpern. Beispielsweise werden bei der Beschreibung der Bewegung des Mondes um die Erde beide Himmelskörper als punktförmig angesehen.										
Erläutern Sie, zu welchem Zweck dieses Modell in den Naturwissenschaften eingesetzt werden kann.										

Abbildung 11: Aufgabenbeispiel „Zweck von Modellen“ (offen)

Massepunkt

In der Mechanik wird ein räumlich ausgedehnter Körper häufig modellhaft als Massepunkt betrachtet. Dabei wird angenommen, dass sich die gesamte Masse des Körpers in seinem Schwerpunkt konzentriert. Dies ermöglicht mathematische Berechnungen bezüglich des Verhaltens von komplexen Körpern. Beispielsweise werden bei der Beschreibung der Bewegung des Mondes um die Erde beide Himmelskörper als punktförmig angesehen.

Zu welchem Zweck kann dieses Modell in den Naturwissenschaften eingesetzt werden?

Kreuzen Sie an.

- Das Modell vereinfacht die Abschätzung der Gravitationskräfte zwischen Körpern.
- Mit dem Modell können die Gravitationskräfte zwischen Mond und Erde exakt berechnet werden.
- Das Modell kann die Lage und die Größe zweier Körper zueinander erklären.
- Mit dem Modell kann man die Gravitationskräfte zwischen dem Mond und einzelnen Punkten der Erde abschätzen.

Abbildung 12: Aufgabenbeispiel "Zweck von Modellen" (geschlossen)

Entwicklung der Aufgabe:

Im ersten Satz des Itemstamms wurden die Wörter „häufig“ und „modellhaft“ vertauscht, um zu einem besseren Verständnis des Satzes beizutragen.

Tabelle 9: Erläuterungen zu den MC-Optionen und Beispielantworten der Studierenden zum Aufgabenbeispiel „Zweck von Modellen“

MC	
1 Attraktor	Diese Option erklärt den Zweck des Modells korrekt. Unter anderem wird durch die vereinfachende Betrachtung möglich die Gravitationskräfte abzuschätzen. Beispielantworten von Studierenden: <i>„Bestimmung von Anziehungs-/ [.]-Kräften[...]“</i>
2 Distraktor	Diese Option ist teilweise richtig. Eine Berechnung der Gravitationskräfte zwischen Mond und Erde ist möglich, allerdings ist keine exakte Berechnung möglich. Beispielantworten von Studierenden: <i>„dies ermöglicht eine genauere Berechnung der Reaktion der unterschiedlichen Massen auf äußere Einflüsse“</i> <i>„Berechnung der Erde-Mond-Konstellation“</i>
3 Distraktor	Diese Option ist falsch, weil durch das Modell keine Möglichkeit besteht die Lage und Größe der Körper abzuschätzen Beispielantworten von Studierenden: <i>„Z.B. Bei der Beschreibung eines Atoms. 1. Verständnis der Lage 2. Raumbeziehungen 3. Größe zueinander“</i>
4 Distraktor	Diese Option ist falsch, weil sie ein Abschätzen der Gravitationskräfte zwischen Mond und einzelnen Punkten der Erde betrifft. Beispielantworten von Studierenden: <i>„Berechnung der Erde-Mond-Konstellation“</i> (Dieser Distraktor wurde aus Ermangelung anderer passender Antworten ebenfalls aus dieser Antwort abgeleitet)

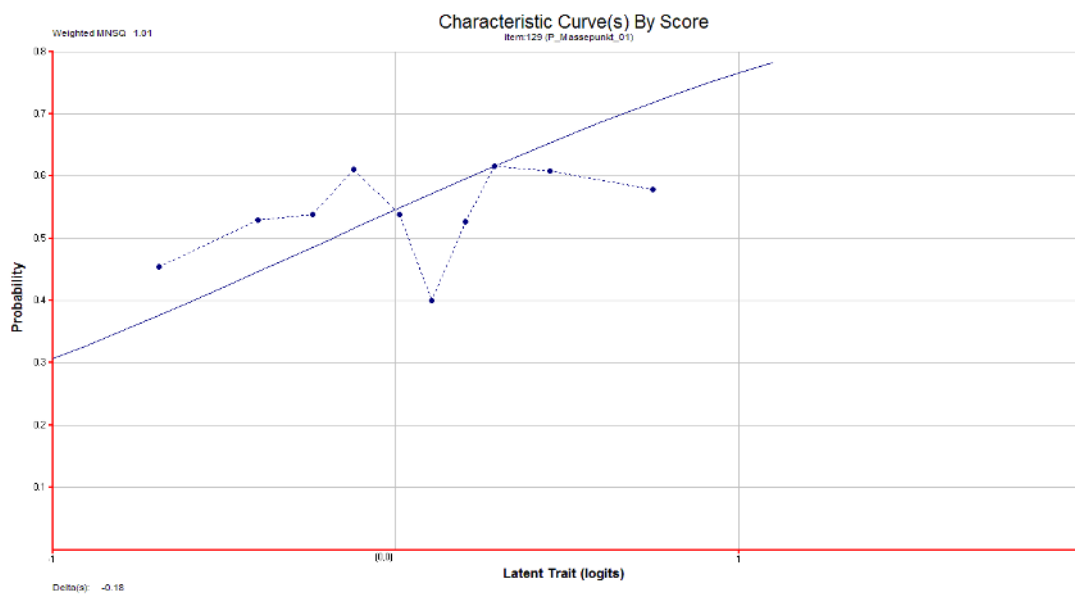


Abbildung 13: ICC des Aufgabenbeispiels "Zweck von Modellen"

Tabelle 10: Item-Fit-Werte des Aufgabenbeispiels "Zweck von Modellen"

Logit	wMNSQ	T	discr.
-0,184	1,01	0,3	0,22

Einschätzung:

Die Itemfit-Parameter dieses Items (siehe Tabelle 10) sind, abgesehen von der etwas zu niedrigen Trennschärfe, zufriedenstellend. Der empirische Verlauf der Lösungswahrscheinlichkeiten stimmt aber nur ungenügend mit dem der theoretischen ICC (siehe Abbildung 13) überein. Es zeigt sich ein Underfit. Da für diese Facette Aufgaben mit besserer Modellpassung bereitstanden, wurde diese Aufgabe aus dem Test entfernt. Die Problematik dieser Aufgabe liegt möglicherweise darin, dass im Itemstamm das Modell nur im Zusammenhang mit Bewegungen beschrieben wird, Gravitationskräfte aber erst in den Antworten benannt werden. Außerdem werden im Stamm Erde und Mond explizit benannt, was in Antwortoption vier (siehe Abbildung 12) wieder aufgegriffen wird. Diese beiden Wörter hatten eventuell Signalcharakter für einige Proband/inn/en.

6.3 Beschreibung des fertigen Tests

Der fertige Kompetenztest besteht aus 123 Items, die sich wie folgt über die einzelnen Fächer, Dimensionen und Kompetenzfacetten verteilen:

Tabelle 11: Übersicht über die Items (B=Biologie, C=Chemie, P=Physik)

Naturwissenschaftliche Erkenntnisgewinnung				
Naturwissenschaftliche Untersuchungen	Fragestellungen	Hypothesen	Planung und Durchführung	Auswertung und Interpretation
<i>Items</i>	B:6 C:5 P:8	B:6 C:4 P:7	B:8 C:5 P:9	B:7 C:4 P:6
Modelle Nutzen	Zweck von Modellen	Testen von Modellen	Ändern von Modellen	
<i>Items</i>	B:3 C:7 P:6	B:6 C:5 P:7	B:6 C:3 P:5	

Die Länge eines Testheftes wird durch verschiedene Faktoren begrenzt. So muss die zeitliche Belastung, der die Proband/inn/en ausgesetzt werden, zumutbar sein (Moosbrugger & Kelava, 2012a), wodurch auch Ermüdungserscheinungen vorgebeugt werden sollen. Außerdem wird durch institutionelle Vorgaben (Lehrveranstaltungen, in denen der Test eingesetzt werden soll) die Testzeit durch externe Faktoren ebenfalls limitiert (Mayer & Wellnitz, 2014).

Um diese Einschränkungen zu berücksichtigen, gleichzeitig aber der Breite der Kompetenz mit seinen 123 konstruierten Aufgaben gerecht zu werden, wurden die Aufgaben im Multi-Matrix-Design²⁷ über mehrere Testhefte verteilt. Dieses Vorgehen ist mit der Einschränkung verbunden, dass sich die erhobenen Daten weniger zur Feststellung individueller Personenparameter eignen. Zur Schätzung von Gruppenfähigkeiten können die Daten aber weiterhin genutzt werden (Gonzales & Rutkowski, 2010, S. 126).

Aus den Erfahrungen der Pilotierungsstudie wurde die Anzahl der Items auf 18 pro Testheft festgelegt. Die Domäne Physik ist daher mit sechs Items in jedem Testheft vertreten, die anderen beiden Fächer jeweils auch mit sechs. Damit sollte eine Testzeit von ca. 30

²⁷ Das hier dargestellte Design wurde innerhalb des Gesamtprojekts Ko-WADiS entworfen und nicht speziell für diese Dissertation.

Minuten erreicht werden. Die Items wurden im Rotationsdesign auf die Testhefte verteilt. Zunächst wurden für die beiden Dimensionen *Naturwissenschaftliche Untersuchungen* und *Modelle Nutzen* Blöcke von je drei Items gebildet. Dadurch wird erreicht, dass in jedem Testheft jede Dimension und jedes Fach mit mindestens drei Items vertreten ist (Gonzales & Rutkowski, 2010, S. 130). Diese werden entsprechend Tabelle 12 über 20 Testhefte verteilt. Um alle Proband/inn/en unabhängig vom bearbeiteten Testheft auf einer Skala abbilden zu können, müssen die Testhefte untereinander verlinkt werden. Dies wird in diesem Fall durch Ankerblöcke erreicht, die jeweils in mindestens zwei Testheften vorkommen (Frey, Hartig & Rupp, 2009, S. 43). Sowohl die Fächer als auch die Blöcke wechseln die Position innerhalb der Testhefte, um entsprechenden Positionseffekten entgegenzuwirken (Gonzales & Rutkowski, 2010, S. 135f.). Durch die unterschiedliche Anzahl von Items pro Fach und Dimension ergibt sich ein unbalanced²⁸ incomplete²⁹ Testdesign (Gonzales & Rutkowski, 2010).

Tabelle 12: Übersicht über das Rotationsdesign (P=Physik, C=Chemie, B=Biologie, U=Untersuchungen, M=Modelle)

TH30	TH31	TH32	TH33	TH34	TH35	TH36	TH37	TH38	TH39
PU1	BU2	CU1	PU3	BU5	CU2	PU4	BU8	CU3	PU6
PM1	BM2	CM3	PM2	BM5	CM4	PM4	BM3	CM5	PM5
CU1	PU2	BU3	CU2	PU3	BU6	CU3	PU5	BU9	CU4
CM1	PM1	BM3	CM2	PM3	BM1	CM3	PM4	BM4	CM4
BU1	CU1	PU2	BU4	CU2	PU4	BU7	CU3	PU5	BU1
BM1	CM4	PM2	BM4	CM5	PM3	BM2	CM1	PM5	BM5
TH40	TH41	TH42	TH43	TH44	TH45	TH46	TH47	TH48	TH49
BU2	CU4	PU7	BU5	CU5	PU9	BU8	CU6	PU10	BU2
BM1	CM1	PM1	BM4	CM2	PM2	BM2	CM3	PM4	BM5
PU6	BU3	CU5	PU8	BU6	CU6	PU9	BU9	CU1	PU1
PM6	BM2	CM5	PM1	BM5	CM1	PM3	BM3	CM2	PM5
CU4	PU7	BU4	CU5	PU8	BU7	CU6	PU10	BU1	CU4
CM2	PM6	BM3	CM3	PM2	BM1	CM4	PM3	BM4	CM5

²⁸ Die Blöcke tauchen unterschiedlich oft im Testheft auf.

²⁹ Nicht alle Proband/inn/en erhalten alle Aufgaben.

In der Anfangsphase der Erhebungen wurde ein leicht anderes Design eingesetzt, was vom Grundprinzip aber dem oben angegebenen entspricht. Da später noch sechs weitere Aufgaben hinzugenommen wurden, wurde das Design entsprechend angepasst. Für die im Anschluss durchgeführten Auswertungen werden alle Fälle unabhängig vom Testheftdesign verwendet. Dies kann geschehen, weil beide Designs untereinander durch die 117 Aufgaben, die in beiden Designs vorhanden sind, geankert sind.

6.3.1 Testgüte

Die Güte eines Tests wird häufig durch die drei Kriterien Objektivität, Reliabilität und Validität beschrieben (Moosbrugger & Kelava, 2012a). Objektiv ist ein Test, wenn das gemessene Merkmal unabhängig von Testleiter und -auswerter gemessen wird. Außerdem müssen für die Interpretation ebenso klare Regeln existieren. Es werden die drei Aspekte Durchführungs-, Auswertungs- und Interpretationsobjektivität unterschieden (Moosbrugger & Kelava, 2012a, S. 8). Die Durchführungsobjektivität betrifft die Bedingungen, unter denen der Test durchgeführt wird. Im Rahmen unseres Tests wurde diese durch ein einheitliches Instruktionsblatt gewährleistet. Inhaltliche Fragen wurden während des Testes nicht beantwortet. Die durch äußere Rahmenbedingungen häufig abweichende Testzeit wurde dadurch berücksichtigt, dass nicht bearbeitete Aufgaben am Ende des Testheftes als *missing not reached* kodiert wurden. Dadurch wurde in der Auswertung die Aufgabe so gewertet wie alle anderen, bedingt durch das Multi-Matrix-Design, nicht vorliegenden Aufgaben. Die Auswertungsobjektivität wird durch den ausschließlichen Einsatz von Multiple-Choice-Aufgaben gewährleistet (Hartig & Jude, 2007, S. 30). Falscheingaben beim Abtippen der Testhefte wurden durch eine Kontrolle von 10% der Eingaben und einer augenscheinlichen Kontrolle der Muster der Testhefte in den Dateien vorgebeugt. Die Interpretationsobjektivität wird durch den geringen Spielraum bei der Interpretation von MC-Aufgaben unterstützt.

Die Reliabilität eines Testes gibt an, wie exakt der Test das fokussierte Konstrukt misst, wie groß also der Messfehler ist. Im Rahmen der RASCH-Modellierung wird ein Reliabilitätsmaß (EAP/PV-Reliabilität) geschätzt, wobei einige Einschränkungen zu berücksichtigen sind (Adams, 2005). Die Diskussion der Reliabilität erfolgt nach der Auswertung im Ergebnisteil dieser Arbeit.

„Ein Test gilt dann als valide [...], wenn er das Merkmal, das er messen soll, auch wirklich misst und nicht irgendein anderes.“ (Moosbrugger & Kelava, 2012a, S. 13). Dieser Definition folgend, beschäftigt sich die Validität also vor allem mit den inhaltlichen Aspekten eines Tests. Daher ist dieser Punkt als notwendiges Kriterium der Testgüte anzusehen. Wenn durch

den Test etwas anderes als das intendierte Konstrukt gemessen wird, ist der Test nicht mehr einzusetzen (Hartig, Frey & Jude, 2012, S. 144). Als wesentliche Aspekte der Validität werden die Punkte Inhaltsvalidität, Kriteriumsvalidität und Konstruktvalidität angesehen (Schmiemann & Lücken, 2014; Schnell, Hill & Esser, 2011). Die Diskussion der Validität dieses Test soll sich auf diese Punkte beschränken (vgl. Hartig & Jude, 2007), auch wenn in der Forschung noch weitere Aspekte der Validität diskutiert werden (z. B. Messick, 1995).

Die Inhaltsvalidität beschäftigt sich mit der Frage, ob „möglichst alle Aspekte der Dimension, die gemessen werden sollte, berücksichtigt wurden.“ (Schnell et al., 2011, S. 141). Schon bei der Konstruktion des Testes muss dieser Punkt beachtet werden (Schmiemann & Lücken, 2014). In unserem Fall wird dieser Aspekt durch eine theoriegeleitete Konstruktion des Kompetenzmodells abgesichert. Das genutzte Modell ist bereits etabliert und wurde in verschiedenen Tests erfolgreich eingesetzt (Mayer, 2007). Die einzelnen Aufgaben werden ergänzend durch Expertenratings beurteilt, die unter anderem die Übereinstimmung von Kompetenzmodell und Items absichern sollten (siehe auch Schmiemann & Lücken, 2014). Die Validität innerhalb der Testitems wird zusätzlich durch die Konstruktion auf Grundlage von offenen Aufgaben unterstützt (Sadler, 1998). Die Inhaltsvalidität wird auch post-hoc durch den Einsatz von Think-aloud-Studien überprüft (Schmiemann & Lücken, 2014). Bei Kompetenztests ist hierbei vor allem auf die kognitiven Aspekte im Sinne der Kompetenz zu achten. Es bedarf also *“evidence that the tasks evoke the kinds of thinking and reasoning that are part of the inference on which a judgment of competence is made“* (Shavelson, 2013, S. 36). Die Physikaufgaben aus dem Bereich der Untersuchungen werden in einer Begleitstudie durch Think-aloud-Studien evaluiert (Klemm, in Vorbereitung).

Die Kriteriumsvalidität „bezieht sich auf den Zusammenhang zwischen den empirisch gemessenen Ergebnissen des Messinstruments und einem anders gemessenen empirischen („externen“) Kriterium.“ (Schnell et al., 2011, S. 147). Unterschieden wird eine prädiktive und eine konkurrierende Validität (Schmiemann & Lücken, 2014). Hartmann, Upmeier zu Belzen, Krüger und Pant (2015) untersuchen die konkurrierende Validität dieses Testinstruments im Sinne einer *Known-Group-Validity*. Dabei stellten sich die erwarteten Gruppenunterschiede (z.B. Bachelor-Master) ein. Sie folgern daraus, dass die Ergebnisse die konkurrierende Validität des Instruments stützen (Hartmann & Upmeier zu Belzen et al., 2015). Die prädiktive Validität wurde bislang nicht weiter untersucht. Denkbar wäre, mit Hilfe des Instruments Voraussagen über die Ergebnisse in Experimentalpraktika aufzustellen.

Unter Konstruktvalidität versteht man die Feststellung empirisch abgesicherter Zusammenhänge mit anderen Konstrukten, die vorher theoretisch vorhergesagt wurden (Schnell et al., 2011, S. 148). Denkbar ist hier eine konvergente Validierung mit anderen Instrumenten, die ebenfalls Aspekte naturwissenschaftlicher Erkenntnisgewinnung messen sollen. Dazu wird ein Teilsample der Items aus der Physik mit einem Fragebogen zur Erhebung der Qualität in Experimentalpraktika (Rehfeldt, Mühlenbruch & Nordmeier, 2015) abgeglichen. Erwartet wird hier eine positive Korrelation zwischen dem Kompetenzzuwachs, welcher durch den hier beschriebenen Kompetenztest gemessen wird und dem selbsteingeschätzten Kompetenzzuwachs, welcher durch den Fragebogen erhoben wird (Rehfeldt, Straube, Stiller, Kratschmar & Nordmeier, in Vorbereitung). Daneben wurde ebenfalls eine diskriminante Validierung (Schmiemann & Lücken, 2014) durchgeführt. Dazu wurden die kognitiven Grundfähigkeiten einer Teilstichprobe mit einem Intelligenz-Struktur-Test (Liepmann, Beauducel, Brocke & Nettelstroth, 2012) erhoben. Dadurch soll sichergestellt werden, dass der Test etwas anderes als kognitive Grundfähigkeiten misst (vgl. Schmiemann & Lücken, 2014).

7. Auswertung

Im Rahmen dieser Arbeit wurde auf Grundlage von theoretischen Überlegungen ein Kompetenzstrukturmodell für den Bereich der Naturwissenschaftlichen Erkenntnisgewinnung entwickelt (siehe Kapitel 2 und 3). Die zentralen Forschungsfragen wurden bezüglich der Dimensionierung dieser Kompetenz und zum Kompetenzstand Studierender aufgestellt (siehe Kapitel 4). Da zum Zeitpunkt der Studie kein deutschsprachiges Testinstrument zur Erfassung der Erkenntnisgewinnungskompetenz zur Verfügung stand, wurde ein Kompetenztest entwickelt (siehe Kapitel 6). Die Testgüte wurde durch verschiedene Schritte im Verlauf der Testkonstruktion und durch Post-Hoc-Analysen abgesichert und überprüft (siehe Abschnitt 6.3.1). In diesem Kapitel folgt nun mit Hilfe der oben beschriebenen Methoden (siehe Abschnitt 5.2) die Bearbeitung der oben beschriebenen Fragestellungen und Hypothesen (siehe Kapitel 4).

Zur RASCH-Modellierung wurde die Software ACER Conquest 3.0 (Adams et al., 2012) genutzt. Weitere statistische Berechnungen wurden mit IBM® SPSS 22® (IBM Corp., 2013) durchgeführt. Für die Berechnung von Effektstärken und Test-Stärken wurde G*Power 3.1 (Faul, Erdfelder, Lang & Buchner, 2007) genutzt. Die Erstellung von Diagrammen und sonstige Berechnungen erfolgten mit Microsoft® Excel 2010® (Microsoft, 2010).

7.1 Beschreibung der Stichprobe

Die Normierungsstichprobe bestand aus 2651 Studierenden der Biologie, Chemie und Physik (sowohl Lehramtsstudierende als auch Fachstudierende). Daraus wurden fünf Proband/inn/en entfernt, weil für sie bei weniger als drei Items gültige Antworten vorlagen. Damit blieben 2646 Studierende. Sie stammen zu 43,0% von der Freien Universität Berlin und zu 27,2% von der Humboldt Universität zu Berlin. Weitere 19,2% stammen aus Österreich von den Universitäten in Wien, Salzburg und Innsbruck. Die restlichen Proband/inn/en stammen von verschiedenen Universitäten aus Deutschland (z. B. Hannover, Potsdam, Aachen). Weiblich sind 53,8 % der Proband/inn/en, das Durchschnittsalter beträgt 22,7 Jahre (SD=4,5). In einem Lehramtsstudium waren 52,5% der Studierenden immatrikuliert. Die Verteilung über die Fächer sieht folgendermaßen aus:

Tabelle 13: Verteilung der Proband/inn/en der Normierungsstichprobe nach Fächern

Fach	Physik	Biologie	Chemie
N	695 (26,3 %)	1663 (62,8 %)	552 (20,9 %)

Durch die Möglichkeit, zwei Fächer kombiniert zu studieren, sind in Tabelle 13 Doppelnennungen enthalten. Daher ergibt sich dort eine Gesamtsumme, die höher ist als die Zahl der Proband/inn/en.

Die in dieser Studie fokussierte Gruppe der Physikstudierenden ist im Durchschnitt 23,1 Jahre (SD = 5,0) alt (Fach: 22,1 (4,9) Jahre; Lehramt: 24,1 (5,0) Jahre). In der Gruppe befinden sich 179 weibliche Studierende (25,8 %). Ein Lehramtsstudium absolvieren 342 Studierende (49,2 %). Ein Lehramtsstudium Physik zusammen mit Biologie oder Chemie studieren 64 Studierende (9,2 %). Die Verteilung über die Universitäten sieht folgendermaßen aus:

Tabelle 14: Verteilung der Proband/inn/en der Normierungsstichprobe nach Universitäten

Universität	N	%
Freie Universität Berlin	296	42,6
Humboldt Universität zu Berlin	203	29,2
Universität Wien/Innsbruck	77	11,0
Sonstige (Deutschland)*	81	11,6
Fehlend**	38	5,5

* andere Universitäten in Deutschland

** keine Angaben zur Universität im Testheft

7.2 Fragestellung 1: Dimensionierung des Kompetenzstrukturmodells

Zunächst wird der Fragestellung nachgegangen, inwiefern man die theoretisch plausible Dimensionierung des Kompetenzstrukturmodells nach Arbeitsweisen und die nicht plausible Dimensionierung nach Fächern in den empirisch erhobenen Daten wiederfinden kann.

FF1: Welches Modell beschreibt die Daten am besten?

Zur Beantwortung dieser Fragestellung werden entsprechend der untenstehenden Hypothesen verschiedene ein- und mehrdimensionale RASCH-Modelle spezifiziert. Für die berechneten Modelle werden die Deviance-Statistik sowie die Freiheitsgrade angegeben. Aus diesen beiden Größen werden die Informationskriterien CAIC und BIC sowie der χ^2 -Test berechnet (vgl. Abschnitt 5.2.5) und pro Hypothese tabellarisch dargestellt.

H1a: Ein eindimensionales Modell beschreibt die Daten besser als ein Modell, das nach Fächern differenziert.

Der vorliegende Kompetenztest soll auf die schwachen, fächerübergreifenden Methoden der naturwissenschaftlichen Erkenntnisgewinnung fokussieren. Sofern der Fachinhalt bekannt ist, sollte es daher für die Proband/inn/en keinen Unterschied machen, ob sie Aufgaben einer anderen Naturwissenschaft als der Physik bearbeiten (vgl. Giere et al., 2006; Klahr, 2000). Es werden daher drei Modelle getestet:

1A: eindimensionales Modell

1B: zweidimensionales Modell *Physik* vs. *Biologie & Chemie*

1C: dreidimensionales Modell *Physik, Biologie, Chemie*

Tabelle 15: Prüfung der Modellpassung Hypothese H1a

Modell	Deviance [$-2 \ln(L)$]	f	CAIC	BIC	χ^2 (gegen 1D)
1A (1D)	14738	125	15680,99	15.555,99	-
1B (2D)	14735	128	15700,62	15572,62	0,39
1C (3D)	14735	132	15730,80	15598,80	0,86

Die Informationskriterien CAIC und BIC sind jeweils für das eindimensionale Modell am niedrigsten und deuten daher auf eine bessere Passung des eindimensionalen Modells hin. Der χ^2 -Test zeigt ebenfalls keinen signifikanten Vorteil für die mehrdimensionalen Modelle an. Die Hypothese H1a wird damit gestützt.

H1b: Ein In Bezug auf die Arbeitsweisen *Untersuchungen* und *Modelle nutzen* zweidimensionales Modell beschreibt die Daten besser als ein eindimensionales Modell.

Auch wenn die oben genannten vorläufigen Ergebnisse auf eine eindimensionale Struktur des Modells hindeuten, die sich auch nach den Arbeitsweisen nicht unterscheiden lässt, so gibt es doch theoretische Gründe (siehe Kapitel 4), die für eine mehrdimensionale Struktur sprechen. Es werden folgende Modelle getestet:

2A: eindimensionales Modell

2B: zweidimensionales Modell nach Arbeitsweisen (*Untersuchungen & Modelle Nutzen*)

2C: siebendimensionales Modell nach Kompetenzfacetten

Tabelle 16: Prüfung der Modellpassung Hypothese H1b

Modell	Deviance [$-2 \ln(L)$]	f	CAIC	BIC	χ^2 (gegen 1D)
2A (1D)	14738	125	15680,99	15.555,99	-
2B (2D)	14741	128	15706,62	15578,62	0,39
2C (7D)	14679	158	15870,94	15712,94	<0,01

Der χ^2 -Test wird für das siebendimensionale Modell signifikant, allerdings zeigen die Informationskriterien CAIC und BIC jeweils eine bessere Passung des eindimensionalen Modells an. Die Hypothese H1b muss daher verworfen werden.

H1c: Ein eindimensionales Modell beschreibt die Daten besser als ein mehrdimensionales Modell, welches nach Fächern und Arbeitsweisen unterscheidet.

Aus theoretischer Sicht ist aus den oben dargestellten Gründen (siehe Kapitel 4) auch eine Unterscheidung nach Fächern und Arbeitsweisen denkbar. Allerdings sprechen die beiden zuvor getesteten Hypothesen gegen diese Vermutung. Um diese Hypothese trotzdem getestet zu haben, werden folgende Modelle miteinander verglichen:

3A: eindimensionales Modell

3B: vierdimensionales Modell nach Fächern (*Physik* vs. *Biologie & Chemie*) und Arbeitsweisen (*Untersuchungen* und *Modelle nutzen*)

3C: sechsdimensionales Modell nach *Fächern* und *Arbeitsweisen*

Tabelle 17: Prüfung der Modellpassung Hypothese H1c

Modell	Deviance [$-2 \ln(L)$]	f	CAIC	BIC	χ^2 (gegen 1D)
3A (1D)	14738	125	15680,99	15.555,99	-
3B (4D)	14734	137	15767,91	15630,91	0,98
3C (6D)	14708	150	15839,59	15689,59	0,22

Die Informationskriterien CAIC und BIC deuten auf eine bessere Passung des eindimensionalen Modells hin. Der χ^2 -Test zeigt ebenfalls keinen signifikanten Vorteil für die mehrdimensionalen Modelle an. Die Hypothese H1c wird damit gestützt.

7.3 Fragestellungen zum Kompetenzstand von Physikstudierenden

Im Folgenden sollen die Fragestellungen zum Kompetenzstand verschiedener Studiengruppen bearbeitet werden. Da die Ergebnisse aus der Fragestellung 1 zur Dimensionierung des Kompetenzmodells (vgl. Abschnitt 7.2) auf eine eindimensionale Struktur der Daten hindeuten, beziehen sich die im Folgenden dargestellten Analysen auch auf eine eindimensionale Struktur. Bevor mit der Analyse zu den einzelnen Fragestellungen begonnen werden kann, muss der Datensatz auf Items untersucht werden, die die Ergebnisse verzerren könnten. Dazu gehören Items, die sich nicht gemäß dem RASCH-Modell (siehe Abschnitt 5.2.1) verhalten und Items, die sich für die untersuchte Gruppe anders verhalten (DIF, siehe Abschnitt 5.2.6). Diese Analyse betrachtet die einzelnen Items in Bezug zu den anderen Items derselben Dimension und kann daher erst nach der Prüfung einer eventuellen Dimensionierung durchgeführt werden. Nachdem diese Items entfernt wurden, müssen die Items normiert werden (siehe Abschnitt 5.2.2).

7.3.1 Vorbereitung der Daten

Nach der oben beschriebenen Pilotierung sollten im Test nur noch solche Items verblieben sein, die den Anforderungen des RASCH-Modells entsprechen. Bevor weitere Analysen gerechnet werden, muss diese Annahme aber nochmals überprüft werden. Dazu müssen auf Grundlage der RASCH-Modellierung Items identifiziert werden, die sich nicht entsprechend des RASCH-Modells verhalten und daher aus der Analyse ausgeschlossen werden müssen (vgl. Abschnitt 5.2.1). Dazu wird zunächst ein eindimensionales RASCH-Modell spezifiziert. Die Auswahl geschieht, wie oben beschrieben anhand der Infit-Parameter, der klassischen Trennschärfe, sowie einer Beurteilung der ICC.

Die klassische Trennschärfe bewegt sich im Rahmen von **0,07 – 0,53**. Ein Item weist damit eine besonders niedrige klassische Trennschärfe auf. Insgesamt haben acht Items eine Trennschärfe von $< 0,2$. Sie werden aber aus den oben genannten Gründen (siehe Abschnitt 5.2.1) nicht aus dem Test entfernt. Der $wMNSQ$ ist mit Werten zwischen 0,91 und 1,07 auch nicht auffällig, was aber bei einer großen Stichprobe auch so zu erwarten war. Deshalb werden die T-Werte zur Identifizierung eines möglichen Underfits von Aufgaben genutzt.

Bei einer Aufgabe aus der Facette „Testen von Modellen“ ($T = 3,0$) deutet sich anhand des T-Werts ein Problem mit dieser Aufgabe an. Daher wird bei dieser Aufgabe die ICC genauer betrachtet.

Aus Abbildung 14 wird deutlich, dass der durch den T-Wert von 3,0 angezeigte Underfit auch im Abgleich von ICC (durchgezogene Linie) und empirisch ermittelter Lösungswahrscheinlichkeit (gestrichelte Linie) zu finden ist. (Die Kurve ist flacher als angenommen.) Die Aufgabe wird daher aus der Auswertung ausgeschlossen.

Da sich die Fit-Statistik nach Entfernen von Items mit Underfit neu justiert (Wright & Linacre, 1994), wird zunächst dieses Item entfernt.

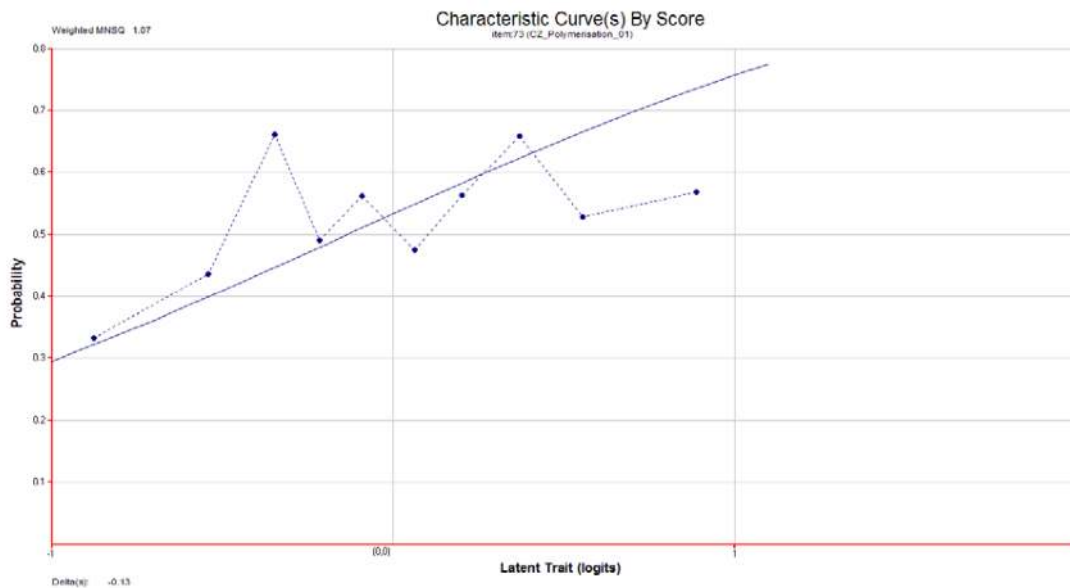


Abbildung 14: ICC Item 73 (Underfit)

Mit den verbleibenden 122 Items wird erneut ein eindimensionales RASCH-Modell gerechnet. Die wMNSQ-Werte liegen abermals zwischen 0,91 und 1,07 und sind damit wieder unauffällig. Die T-Werte liegen zwischen -2,7 und 1,7. Insgesamt zeigen zwei Items mit T-Werten von $T \leq -2,0$ einen Overfit. Diese werden aber aus den oben beschriebenen Gründen nicht entfernt. Bei einer Neugestaltung der Studie wäre zu überlegen, diese aus dem Test zu entfernen und durch produktivere Items zu ersetzen. Kein Item zeigt anhand des T-Werts einen Underfit. Alle weiteren ICCs wurden in dieser Modellierung einer optischen Kontrolle unterzogen. Dabei zeigen sich weitestgehend modellkonforme Verläufe der empirischen Lösungswahrscheinlichkeit, auch wenn sich die Items in ihrer Qualität unterscheiden. Bei einem Item aus der Facette „Auswertung und Interpretation“ und einem aus der Facette „Ändern von Modellen“ wurde ein Underfit durch grafischen Abgleich diagnostiziert. Beispielhaft ist hier die ICC des Items aus Facette „Ändern von Modellen“ abgedruckt (Abbildung 15). Insbesondere zwischen -0,5 Logits und 1,0 Logits

nimmt die Wahrscheinlichkeit, das Item zu lösen, trotz steigender Fähigkeit, ab. Eine ähnliche Charakteristik ist auch bei dem anderen entfernten Item zu finden.

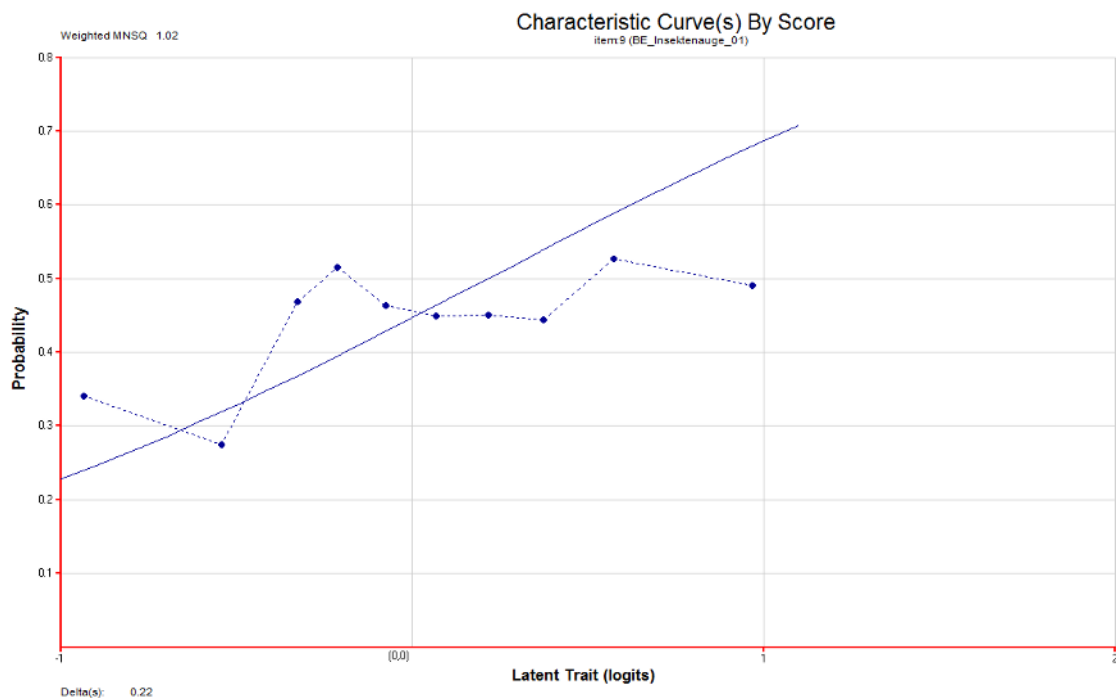


Abbildung 15: ICC Item 9 (Underfit)

In der dritten Modellierung wurden das Item, welches im ersten Durchgang, sowie die zwei Items, welche im zweiten Durchgang als nicht modellkonform identifiziert wurden, entfernt. Die wMNSQ-Werte liegen zwischen 0,93 und 1,08, die T-Werte zwischen -2,2 und 1,9. Eine Abweichung vom Modell wird durch diese Werte demnach nicht angezeigt. Im nächsten Schritt wird erneut eine grafische Überprüfung der ICC vorgenommen. Dabei zeigen sich keine größeren Residuen, weshalb keine weiteren Items entfernt werden

Insgesamt wurden drei Items aus der weiteren Analyse ausgeschlossen: zwei Items aus den Facetten „Auswertung und Interpretation“ und „Ändern von Modellen“ sowie ein Item aus der Facette „Zweck von Modellen“.

7.3.2 Differential Item Functioning (DIF)

Innerhalb der Gesamtstichprobe existieren verschiedene Gruppen, für die es Möglichkeiten für DIF gibt. Hierbei ist insbesondere die Frage zu klären, ob für die Physikstudierenden im Vergleich zur Gesamtgruppe ein DIF existiert. Diese Items müssen vor der Auswertung aus dem Datensatz entfernt werden, weil sie das Ergebnis verzerren würden.

Zur Untersuchung auf DIF wurde eine Dummycodierung vorgenommen. Die Physikstudierenden (sowohl Lehramt, als auch Fach-Physikstudierende) wurden mit 1 (Focalgroup), alle anderen Studierenden mit 0 (Referencegroup) codiert. Es wurde eine MH-DIF-Analyse mit fünf matched groups auf Basis von PVs durchgeführt (siehe Abschnitt 5.2.6). Die Ergebnisse sind in der folgenden Tabelle dargestellt.

Tabelle 18: Verteilung der Items in die DIF-Kategorien

ETS-DIF-Category	Anzahl der Items
A (Negligible)	115
B (Slight to Moderate)	3
C (Moderate to Large)	2

Auf Grund der MH-DIF-Analyse mussten den Ergebnissen zufolge zwei Items aus der Analyse ausgeschlossen werden. Zusammen mit den drei Items, die aufgrund einer mangelhaften Passung auf das RASCH-Modell entfernt werden mussten, verblieben 118 Items in der Auswertung. Da die jeweiligen Facetten des Kompetenzmodells durch fünf bzw. sechs Items weiterhin ausreichend besetzt blieben, wurde die Inhaltsvalidität durch das Entfernen der Items nicht wesentlich beeinträchtigt.

7.3.3 Normierung

Die 118 in der Analyse verbleibenden Items wurden im nächsten Schritt normiert (siehe Abschnitt 5.2.2). Dazu wurde ein eindimensionales RASCH-Modell mit den Proband/inn/en aller drei Fächer gerechnet. Die daraus erhaltenen Item-Schwierigkeiten wurden exportiert und als Grundlage für alle weiteren Berechnungen genutzt.

Die Logit-Werte der Normierung befinden sich in einem Bereich von -2,685 bis 1,697 Logits. Die Wright-Map dieser RASCH-Modellierung offenbart eine gute Abdeckung aller Fähigkeitsbereiche.

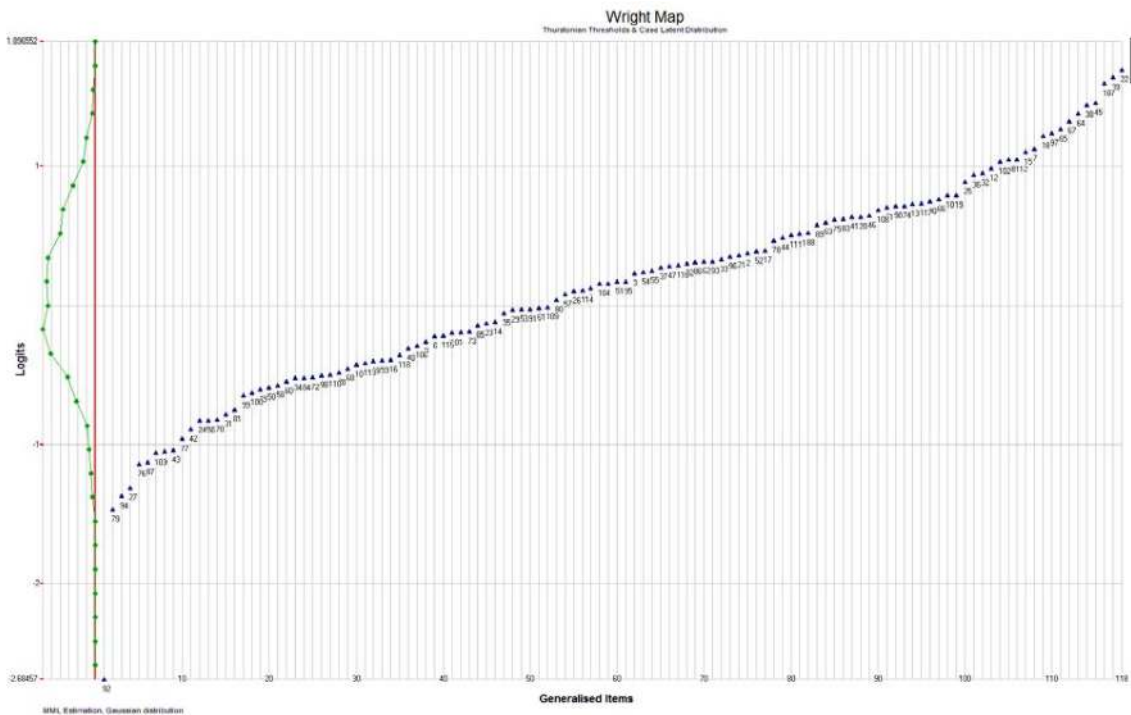


Abbildung 16: Wright-Map der Normierung

Die EAP/PV-Reliabilität dieser Normierung liegt bei $Rel_{EAP/PV} = 0,500$.

7.3.4 Bearbeitung der Fragestellungen 2-5

Zur Prüfung der folgenden Hypothesen werden RASCH-Modellierungen (siehe Abschnitt 5.2.1) vorgenommen, bei denen die Itemschwierigkeiten aus der Normierung als fixierte Werte in das Modell importiert werden (vgl. Abschnitt 7.3.3). Da die Fragestellungen verschiedene Personenmerkmale betreffen, müssen diese vor der Analyse in ein latentes Hintergrundmodell (siehe Abschnitt 5.2.4) modelliert werden. Aus dieser Modellierung erhält man Regressionsgewichte, die unten tabellarisch dargestellt werden. Weiterhin werden pro Proband/in fünf PVs exportiert. Zur Prüfung der Hypothesen werden die Mittelwerte der einzelnen Proband/inn/engruppen mit Hilfe einer ANOVA untersucht (siehe Abschnitt 5.2.7). Die Annahmen für diese Verfahren (Varianzhomogenität und Normalverteilung) werden ebenfalls geprüft (vgl. Abschnitt 5.2.8). Bedingt durch die fünf PVs erhält man auch fünf Statistiken, die entsprechend Abschnitt 5.2.3 kombiniert werden müssen.

Da für eine Analyse des Kompetenzstandes der Studierenden in jedem einzelnen Semester (vgl. Forschungsfrage 2) nicht ausreichend Daten vorliegen, wird eine Einteilung in

drei Studienphasen vorgenommen. Diese orientierte sich am Bachelor-Master-System. Studierende, die nicht in diesem System studierten, wurde entsprechend der Tabelle den Studienphasen zugeordnet. Der Wechsel der Zuordnung nach dem 6. Semester erfolgte auf Grundlage der Regelstudienzeit des Bachelors³⁰.

Tabelle 19: Aufteilung der Studienphasen

Studienphase	Bachelor-Master	Sonstige (z.B. Diplom)
BA1	1.-2. Semester Bachelor	1.-2. Semester
BA2	≥ 3. Semester Bachelor	3.-6. Semester
MA	Master (alle Semester)	≥ 7. Semester

Die Zuordnung erfolgte durch eine Dummymodierung (*BA2* und *MA*). Weiterhin wurden in das Hintergrundmodell die Zugehörigkeit zu einem Lehramtsstudium, die Interaktionen *BA2* × *Lehramt* und *MA* × *Lehramt* sowie das Geschlecht mit aufgenommen.

Für die Berechnung können nur Fälle genutzt werden, für die für alle Regressionsvariablen Antworten vorlagen. Dies trifft für 682 Fälle zu. Die Zellen sind wie folgt durch Personen besetzt:

Tabelle 20: Verteilung der Proband/inn/en Fragestellungen 2 bis 5

Studienphase	Fach	Lehramt	Gesamt
BA1	135 (♀27)	52(♀14)	187(♀41)
BA2	135(♀26)	173(♀66)	308(♀92)
MA	73(♀5)	114(♀40)	187(♀45)

Die EAP/PV-Reliabilität dieser Modellierung liegt bei $Rel_{EAP/PV} = 0,585$. Diese Modellierung klärt 13% ($R^2 = 0,13$)³¹ der Varianz auf. Es ergeben sich folgende latente Regressionsgewichte:

³⁰ Bei unvollständigen Angaben wurde anhand der weiteren demografischen Angaben eine Zuordnung vorgenommen. Z. B. wurde ein 20 Jähriger Proband im 2. Semester ohne Angabe einer Studienart in die Gruppe BA1 einsortiert. Ein 30 Jähriger Proband wurde nicht zugeordnet, weil auf Grund des Alters auch ein Masterstudium möglich wäre.

³¹ Die Varianzaufklärung wurde anhand der in den Raschmodellen geschätzten Varianz errechnet.

Tabelle 21: Unstandardisierte Regressionsgewichte (in Logits) des Hintergrundmodells Fragestellungen 2 bis 5

Prädiktorvariable	B	SE(B)
Constant	0,173*	0,069
Bachelor 2 (BA2)	0,097	0,095
Master	0,502***	0,113
Lehramt (1=Lehramt)	-0,386**	0,130
Lehramt x Bachelor 2	0,233	0,158
Lehramt x Master	0,110	0,174
Geschlecht (1=weiblich)	-0,187**	0,070

*p<,05; **p<,01; ***p<,001

Innerhalb des latenten Regressionsmodells deuten sich schon bestimmte Effekte an. So zeigen die Prädiktorvariablen *Master*, *Lehramt* und *weiblich* signifikante Effekte an. Zur Beantwortung der Fragestellungen 2-5 wird eine faktorielle ANOVA mit den Faktoren *Studienphase*, *Lehramt* und *weiblich* berechnet. Für diese Modellierung werden fünf PVs pro Fall exportiert und in fünf einzelne ANOVAs importiert.

Zunächst werden die Voraussetzungen für die ANOVA geprüft. Für alle sechs Gruppen und für alle fünf PVs wird die Normalverteilung mit dem K-S-Test und dem S-W-Test geprüft, außerdem werden Schiefe und Kurtosis berechnet. Für PV1 zeigt der S-W-Test für zwei Subgruppen signifikante Abweichungen von der Normalverteilung an. Diese treten aber in den Gruppen auf, die hohe Proband/inn/enzahlen aufweisen. Dadurch werden auch schon kleine Abweichungen signifikant (Field, 2009). Die weiteren Subgruppen sind aber unauffällig und zeigen keine signifikanten Abweichungen von der Normalverteilung an. Insbesondere bleibt der K-S-Test durchgängig ohne signifikantes Ergebnis. Da für den übergroßen Anteil der Subgruppen keine Verletzung der Normalverteilung angezeigt wird, die Abweichung auch nur durch den grundsätzlich konservativeren S-W-Test (Field, 2009, S. 148) erkannt wird und die ANOVA generell auch als robust gegenüber einer derartigen Verletzung gilt (Bühner & Ziegler, 2009, S. 368), wird die ANOVA als Testverfahren beibehalten.

Der Levene-Test zur Überprüfung der Varianzgleichheit zeigt keine Verletzung der Varianzhomogenität für alle fünf PVs an ($F_{PV1}(11,670) = 1,189$,n.s. ; $F_{PV2}(11,670) = 1,199$, n.s. ; $F_{PV3}(11,670) = 0,868$, n.s. ; $F_{PV4}(11,670) = 1,118$,n.s. ; $F_{PV5}(11,670) = 0,346$, n.s.)

Die hier dargestellten Ergebnisse sind durch Kombination der Teststatistiken (siehe Abschnitt 5.2.3) aus den Ergebnissen von fünf ANOVAs für jeden einzelnen PV entstanden. Aus den gemittelten Quadratsummen wurden F-Werte errechnet (siehe Abschnitt 5.2.7). Die einzelnen ANOVAs sind im Anhang aufgeführt. Die Signifikanz wurde durch Abgleich mit kritischen F-Werten ermittelt (Backhaus, Erichson, Plinke & Weiber, 2000). Die Effektstärken wurden aus den Quadratsummen ermittelt (siehe Abschnitt 5.2.9).

Tabelle 22: Kombinierte ANOVA Fragestellungen 2 bis 5

Abhängige Variable:	PV1, PV2, PV3, PV4, PV5 (kombiniert)						
Quelle	Typ III Quadratsumme	df	Quadratischer Mittelwert	F	Sig.	η^2	r
Korrigiertes Modell	37,341	6	6,224	21,594	<0,001	0,161	0,4
Konstanter Term	8,733	1	8,733	30,301	<0,001	0,043	0,21
Studienphase	26,878	2	13,439	46,63	<0,001	0,121	0,35
Lehramt	8,106	1	8,106	28,126	<0,001	0,040	0,2
Weiblich	4,825	1	4,825	16,741	<0,001	0,024	0,15
Studienphase * Lehramt	1,551	2	0,776	2,691	>0,05	0,008	0,09
Fehler	194,540	675	0,288				
Gesamtsumme	255,513	682					
Korrigierter Gesamtwert	231,881	681					

Tabelle 23: Kombinierte Deskriptive Statistik Fragestellungen 2 bis 5

	M	SD ²	Mess- varianz	Fehler- varianz	SD gesamt	N	SE
BA1 Gesamt	0,0257	0,3083	0,0001	0,3084	0,5553	187	0,0406
BA2 Gesamt	0,1125	0,3068	0,0001	0,3069	0,554	308	0,0316
Master Gesamt	0,4676	0,3168	0,0009	0,3179	0,5638	187	0,0412
BA 1 Fach	0,1266	0,2811	0,0001	0,2812	0,5303	135	0,0456
BA 2 Fach	0,1984	0,3198	0,0003	0,3202	0,5659	135	0,0487
MA Fach	0,6542	0,3005	0,0017	0,3025	0,55	73	0,0644
BA 1 LA	-0,2361	0,2866	0,0029	0,2901	0,5386	52	0,0747
BA 2 LA	0,0454	0,2879	0,0002	0,2881	0,5367	173	0,0408
MA LA	0,3481	0,2928	0,0007	0,2936	0,5418	114	0,0507
Fach	0,2698	0,3348	0,0001	0,3349	0,5787	343	0,0312
LA	0,1014	0,3234	0,0002	0,3236	0,5689	339	0,0309
nicht weiblich	0,2431	0,3291	0,0288	0,3637	0,6031	504	0,0269
weiblich	0,0631	0,3375	0,0094	0,3488	0,5906	178	0,0443

Ausgehend von dieser ANOVA können nun die Fragestellungen 2-5 bearbeitet werden:

FF2: Existieren Unterschiede im Kompetenzstand zwischen Lehramtsstudierenden der Physik und Fachstudierenden der Physik?

H2: Lehramtsstudierende weisen einen geringeren Kompetenzstand als Fachstudierende der Physik auf. (sig. (Lehramt)<0,05).

Die ANOVA zeigt signifikante Unterschiede zwischen Lehramtsstudierenden und Fachstudierenden der Physik $F(1,675) = 28,126$, $p < 0,001$, $\eta_p^2 = 0,040$, $r = 0,20$. Die Hypothese H2 wird daher gestützt.

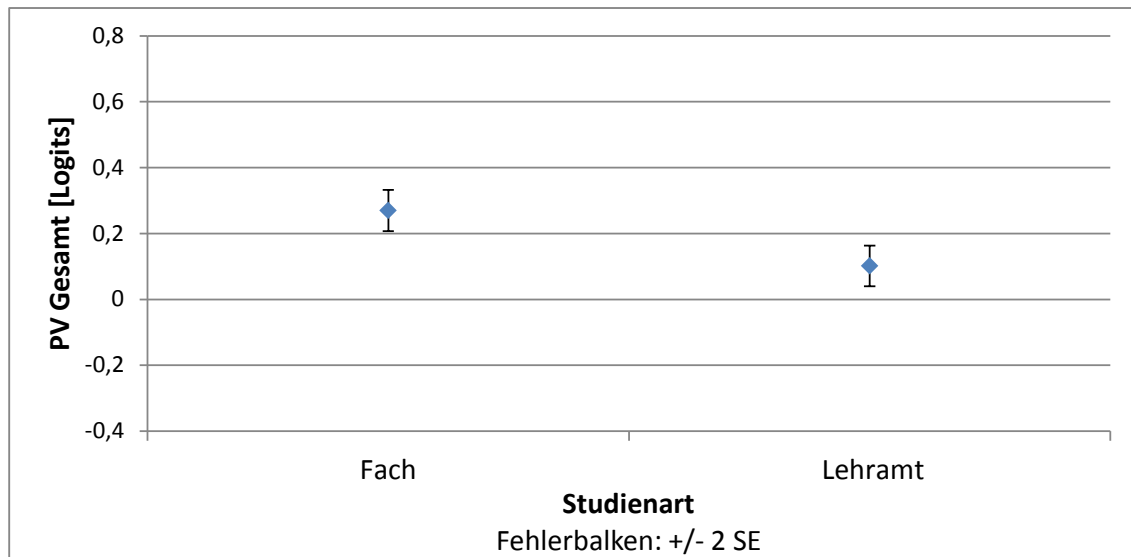


Abbildung 17: Kompetenzstand Fach- und Lehramtsstudierender (Fragestellung 2)

FF3: Existieren Unterschiede im Kompetenzstand Studierender im Bereich der Erkenntnisgewinnung zu verschiedenen Phasen im Studienverlauf?

H3: Der mittlere Kompetenzstand der Studierenden der Physik ist höher, je weiter fortgeschritten die Studierenden im Studium sind ($p < 0,05$).

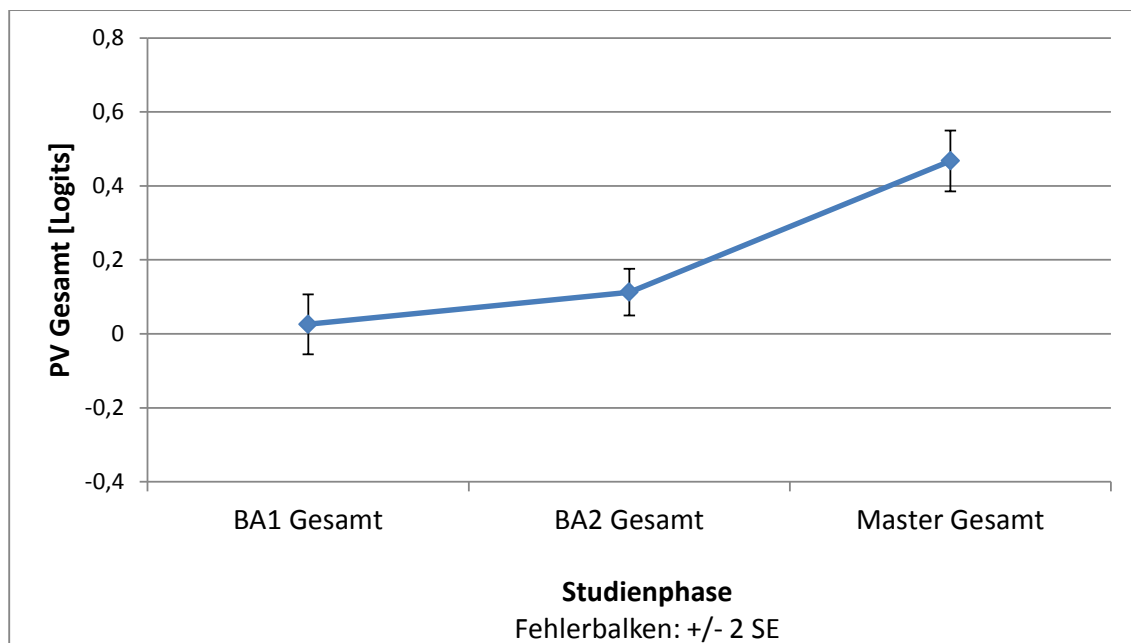


Abbildung 18: Kompetenzstand Studierende zu verschiedenen Phasen des Studiums (Fragestellung 3)

Die ANOVA zeigt für die Variable Studienverlauf eine signifikante Veränderung über den Studienverlauf an $F(2,675) = 46,630$, $p < 0,001$, $\eta^2 = 0,121$, $r = 0,35$.

Unterschiede zwischen den drei Gruppen werden mit zwei T-Tests (siehe Abschnitt 5.2.7) genauer untersucht³².

Tabelle 24: T-Test: Bachelor Phase 1 gegen Bachelor Phase 2

Gruppe	N	M	SD	df	t	P (1-seitig)	d	r
Bachelor 1	187	0,0257	0,5553	493	-1,69	0,045	0,156	0,08
Bachelor 2	308	0,1125	0,554					

Der T-Test zwischen den Gruppen zeigt keinen signifikanten Unterschied zwischen den Gruppen Bachelor 1 und Bachelor 2 (Senkung des Signifikanzniveaus auf $p < 0,025$). Eine Teststärke von $1 - \beta \geq 0,80$ wird für Effektstärken von $d \geq 0,23$ erreicht.

Tabelle 25: T-Test: Bachelor Phase 2 gegen Master

Gruppe	N	M	SD	df	t	p (1-seitig)	d	r
Bachelor 2	308	0,1125	0,554	493	-6,87	<0,01	0,636	0,30
Master	187	0,4676	0,5638					

Der T-Test zwischen den Gruppen Bachelor 2 und Master zeigt einen signifikanten Unterschied zwischen den Gruppen. Die Ergebnisse stützen die Hypothese H3 für die Differenz im Kompetenzstand zwischen den Phasen Bachelor 2 und Master.

³² Berechnet mit SPSS 22, nach der Methode von Field (2009, S. 336). Senkung des Signifikanzniveaus auf $p < 0,025$ nach Bonferoni (Field (2009, S. 373)

FF4: Existieren Unterschiede in den Differenzen im Kompetenzstand zwischen Lehramtsstudierenden der Physik und Fachstudierenden zu den unterschiedlichen Phasen des Studiums?

H4: Es existieren keine Unterschiede in den Differenzen im Kompetenzstand zwischen Lehramtsstudierenden der Physik und Fachstudierenden zu den unterschiedlichen Phasen des Studiums (sig. (Lehramt*Studienphase) > 0,05; $1 - \beta \geq 0,90$)

Die ANOVA zeigt für die Interaktion Lehramt*Studienphase keine signifikanten Unterschiede an $F(2, 675) = 2,691$, $p > 0,05$, $\eta^2 = 0,008$. Die Teststärke (Power) erreicht Werte von $1 - \beta \geq 0,9$ bei Effektstärken von $\eta^2 = 0,017$. Die Hypothese H4 wird demnach für starke und mittlere Effekte gestützt.

Auf Abbildung 19 ist der Kompetenzstand zu den drei erhobenen Phasen des Studiums für Fach- und Lehramtsstudierende geplottet. Es zeigt sich ein Zuwachs über die Semester, der weitestgehend synchron verläuft, auch wenn der Unterschied im Kompetenzstand in der zweiten Phase des Bachelors deutlich abnimmt. In der Masterphase des Studiums ist diese Differenz im Kompetenzstand wieder ablesbar.

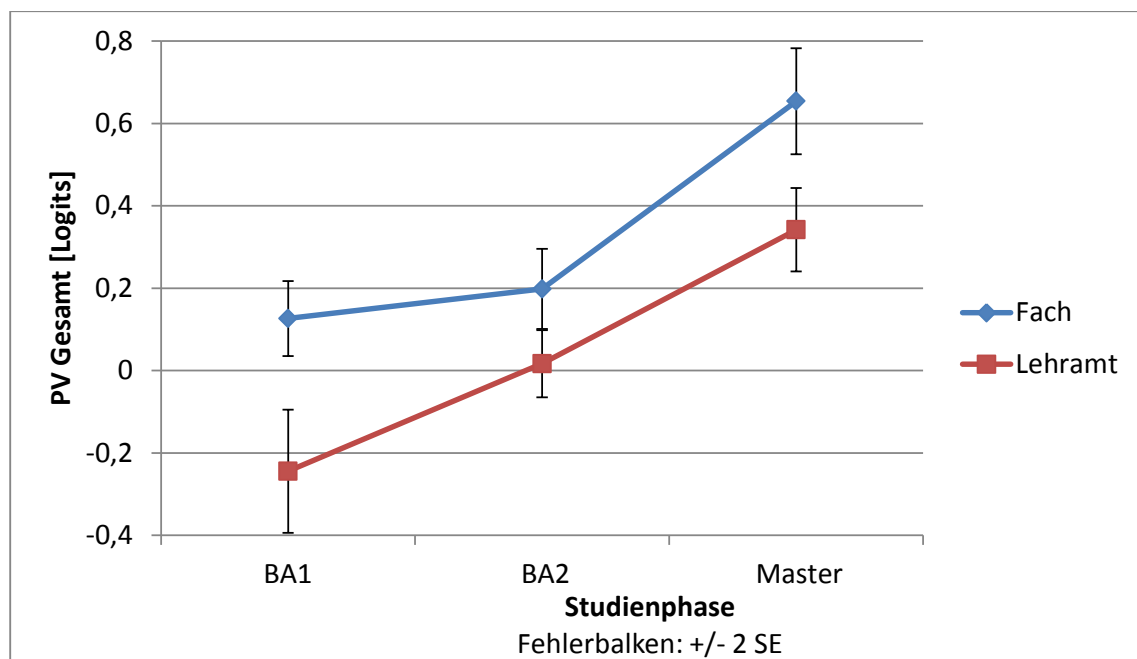


Abbildung 19: Kompetenzstand Lehramts- und Fachstudierender zu verschiedenen Phasen des Studiums (Fragestellung 4)

FF5: Existieren Unterschiede im Kompetenzstand von weiblichen und männlichen Studierenden?

H5: Weibliche Studierende zeigen den gleichen Kompetenzstand wie männliche Studierende ($p > 0,05$; $1 - \beta \geq 0,90$).³³

Die ANOVA zeigt signifikante Unterschiede zwischen weiblichen und nicht weiblichen Studierenden, $F(1,675) = 16,741$, $p < 0,001$, $\eta^2 = 0,024$, $r = 0,15$, zu Gunsten der nicht weiblichen Studierenden an. Auf eine Power-Analyse kann aufgrund des signifikanten Ergebnisses verzichtet werden. Die Hypothese H5 muss daher verworfen werden.

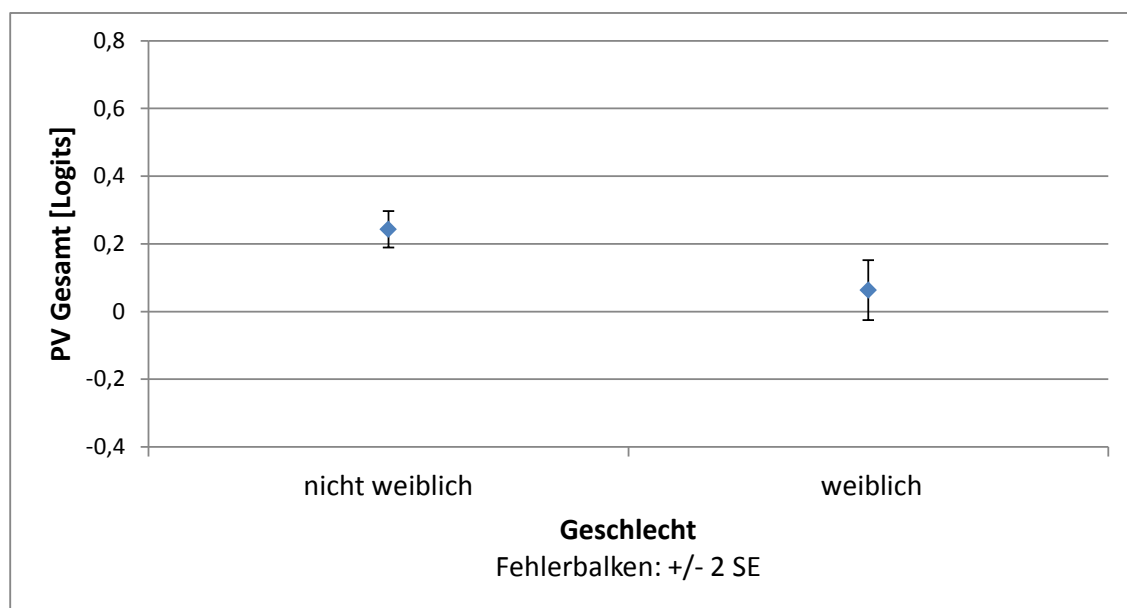


Abbildung 20: Kompetenzstand nicht- weiblicher und weiblicher Studierender (Fragestellung 5)

³³ Bei der Prüfung dieser Hypothese wird nicht zwischen Fach- und Lehramtsstudierenden unterschieden.

FF6: Gibt es Unterschiede im Kompetenzstand von Lehramtsstudierenden, die neben Physik ein weiteres naturwissenschaftliches Fach studieren im Vergleich zu denen, die kein weiteres naturwissenschaftliches Fach studieren?

H6: Studierende mit einem weiteren naturwissenschaftlichen Fach weisen einen höheren Kompetenzstand auf, als Studierende ohne weiteres naturwissenschaftliches Fach ($p < 0,05$).

Zur Beantwortung dieser Fragestellung wird eine erneute RASCH-Modellierung mit latentem Hintergrundmodell vorgenommen. In der Stichprobe werden nur Lehramtsstudierende aufgenommen. Es werden die in der ersten Modellierung signifikanten Prädiktorvariablen *BA2*, *MA* und *weiblich*, sowie als neue Variable *Zwei naturwissenschaftliche Fächer* aufgenommen. Die Variable *Lehramt* kann bei dieser Modellierung entfallen.

Im latenten Hintergrundmodell der RASCH-Modellierung ergeben sich folgende Regressionsgewichte:

Tabelle 26: Unstandardisierte Regressionsgewichte (in Logits) des Hintergrundmodells

Prädiktorvariable	B	SE(B)
Constant	-0,232*	0,115
Bachelor 2 (BA2)	0,340*	0,124
Master (MA)	0,625***	0,129
Geschlecht (1=weiblich)	-0,209*	0,086
Zwei naw. Fächer (1=2 naw. Fächer)	0,093	0,106

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

Die Prädiktorvariable *zwei naw. Fächer* wird im Hintergrundmodell nicht signifikant ($p > 0,05$).

Die Tests auf Abweichung von der Normalverteilung fallen alle nicht signifikant aus (siehe Anhang). Gleiches gilt für die Tests auf Varianzhomogenität ($F_{PV1}(1,337) = 0,122$, n.s.; $F_{PV2}(1,337) = 0,433$, n.s.; $F_{PV3}(1,337) = 1,233$, n.s.; $F_{PV4}(1,337) = 0,440$, n.s.; $F_{PV5}(1,337) = 0,400$, n.s.). Die erhaltenen PVs werden gemittelt und die Standardabweichungen entsprechend korrigiert. Für die beiden betreffenden Gruppen wird ein T-Test gerechnet. Der T-Test fällt nicht signifikant aus. Eine Teststärke von $1 - \beta \geq 0,80$ wird

für Effektstärken von $d \geq 0,35$ erreicht. Die Hypothese H6 muss daher für mittlere und starke Effekte verworfen werden

Tabelle 27: T-Test: Lehramtsstudierende der Physik gegen Lehramtsstudierende der Physik mit einem weiterem naturwissenschaftlichen Fach

Gruppe	N	M	SD	df	t	P (1-seitig)	r
Nur Physik	278	0,076	0,5553	337	-0,04	0,49	0.00
Physik + 2. Nawi Fach	61	0,0793	0,5587				

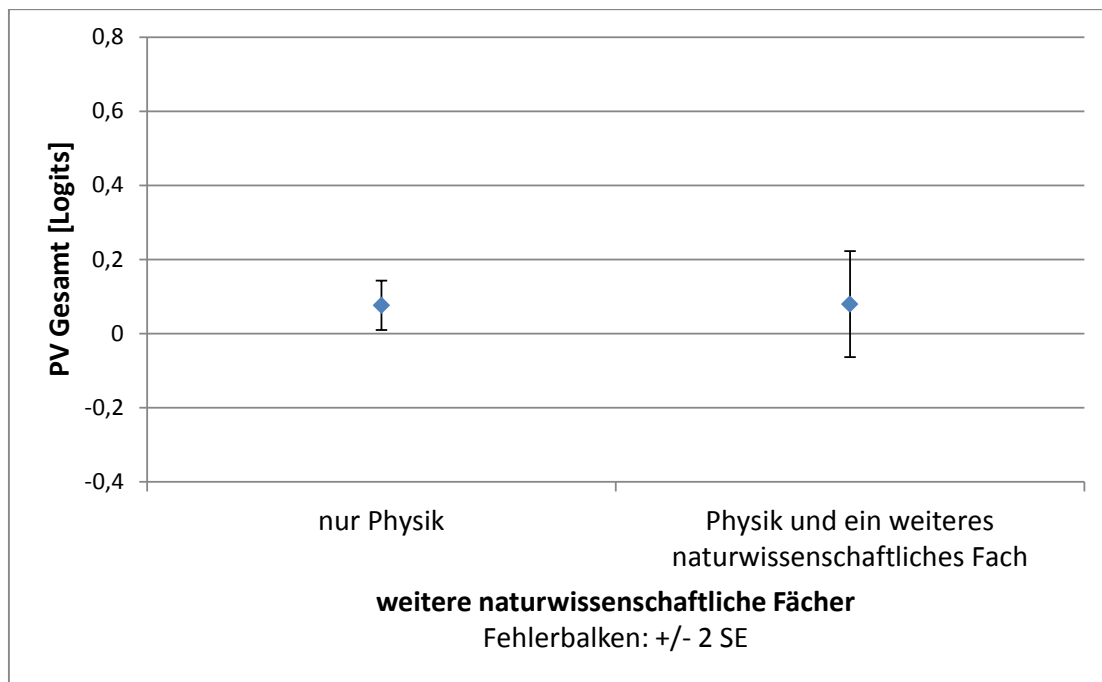


Abbildung 21: Kompetenzstand Lehramtsstudierender der Physik und Lehramtsstudierende der Physik mit einem weiteren naturwissenschaftlichen Fach (Fragestellung 6)

7.4 Fragestellungen 7 und 8: Kompetenzstand von Studierenden im Fach „Integrierte Naturwissenschaften“

Die Fragestellungen 7 und 8 betreffen Grundschullehramtsstudierende im Fach „Integrierte Naturwissenschaften“.

Zur Beantwortung dieser Fragestellungen wird eine RASCH-Modellierung (siehe Abschnitt 5.2.1) mit den Proband/inn/en, die das Fach Integrierte Naturwissenschaften studieren, vorgenommen. Als Vergleichsgruppe werden die Proband/inn/en, die Physik im ersten Semester studieren, herangezogen. In das Hintergrundmodell (siehe Abschnitt 5.2.4) werden die Prädiktorvariablen *Naw1* (Integrierte Naturwissenschaften im 1. Semester) und *Naw2* (Integrierte Naturwissenschaften in einem höheren Fachsemester) einbezogen. Die Proband/inn/en verteilen sich wie folgt auf die einzelnen Gruppen:

Tabelle 28: Verteilung der Proband/inn/en

Studienart	Physik-Lehramt	Int. Nawi. (1.Sem)	Int. Nawi. (> 1.Sem)	Gesamt
N	98 (♀30)	27(♀21)	32(♀26)	157(♀77)

Von den Studierenden im Fach Integrierte Naturwissenschaften, die nicht im ersten Fachsemester studieren, befindet sich der Großteil im dritten (33,3 %) und fünften (57,6 %) Semester. Der Rest befindet sich in anderen Bachelorsemestern.

Im Hintergrundmodell ergeben sich folgende Regressionsgewichte:

Tabelle 29: Unstandardisierte Regressionsgewichte (in Logits) des Hintergrundmodells

Prädiktorvariable	B	SE(B)
Constant	0,095	0,083
weiblich	-0,244	0,130
Naw1	-0,260	0,167
Naw2	-0,038	0,160

Die Tests auf Normalverteilung (siehe Anlage) zeigen für die Gruppe Naw2 bei PV2 und für die Gruppe Naw1 bei PV4 eine Verletzung der Annahme der Normalverteilung an.

Um die Auswertung weiterhin mit einem T-Test durchführen zu können, werden bei den entsprechenden Gruppen die betreffenden PVs aus der Berechnung ausgenommen. Die kombinierten Mittelwerte und Standardabweichungen für die Gruppen Naw1 und Naw2 werden daher nur auf Grundlage von jeweils vier PVs berechnet.

Tabelle 30: Deskriptive Statistik Fragestellungen 7 und 8

	M	SD ²	Mess- varianz	Fehler- varianz	SD gesamt	N	SE
Lehramt PH	0,0162	0,2346	0,0001	0,2358	0,4856	98	0,0491
Naw1 (1. Sem)	-0,3934	0,1987	0,0425	0,2518	0,5018	27	0,0966
Naw2 (> 1. Sem)	-0,1295	0,2052	0,0046	0,2110	0,4593	32	0,0812

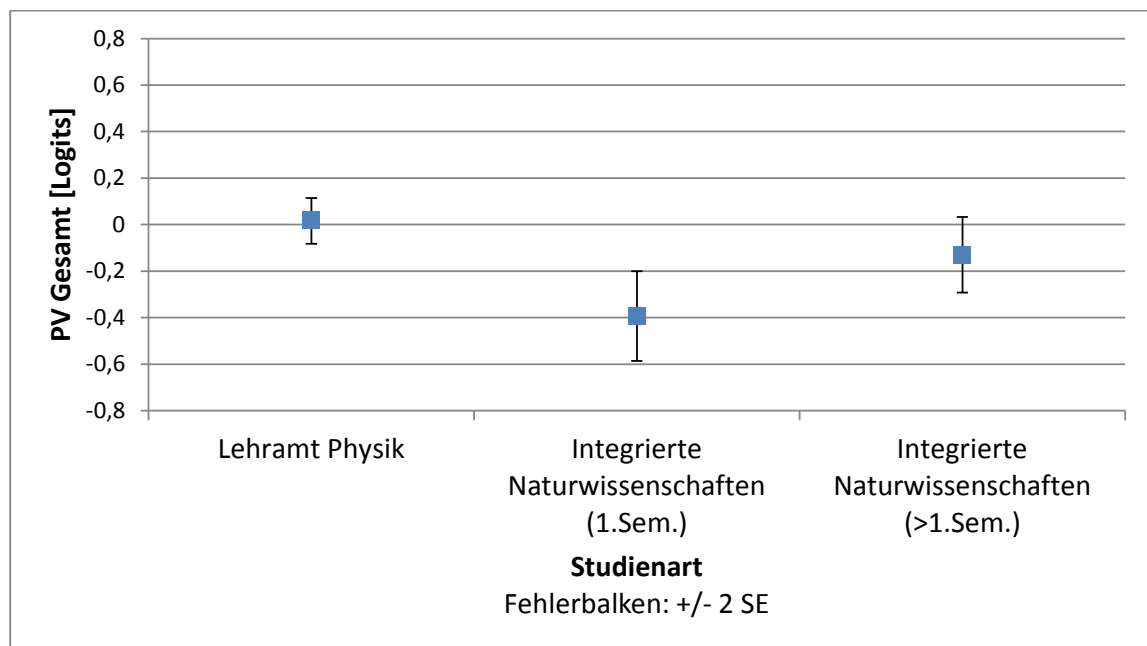


Abbildung 22: Kompetenzstand Grundschullehramtsstudierender (Integrierte Naturwissenschaften) zu verschiedenen Zeiten des Studiums mit Vergleichsgruppe (Fragestellungen 7 und 8)

FF7: Haben Grundschullehramtsstudierende des Fachs „Integrierte Naturwissenschaften“ den gleichen Kompetenzstand zu Beginn des Studiums wie Lehramtsstudierende der Physik?

H7: Studierende des Fachs „Integrierte Naturwissenschaften“ haben den gleichen Kompetenzstand wie Physiklehramtsstudierende zu Beginn des Studiums ($p > 0,05$, $1 - \beta \geq 0,90$).

Der Levene-Test zur Überprüfung der Varianzgleichheit zeigt keine Verletzung der Varianzhomogenität für alle fünf PVs an ($F_{PV1}(1,123) = 2,525$, n.s. ; $F_{PV2}(1,123) = 0,113$, n.s. ; $F_{PV3}(1,123) = 0,382$, n.s. ; $F_{PV4}(1,123) = 0,050$, n.s. ; $F_{PV5}(1,123) = 0,838$, n.s.).

Tabelle 31: T-Test Physik-Lehramtsstudierende gegen Studierende des Faches Integrierte Naturwissenschaften

Gruppe	N	M	SD	df	t	P (2-seitig)	d	r
Physik LA	98	0,0162	0,4856	122	3,85	<0,01	0,83	0,33
Integrierte Naturwissenschaften (1.Sem)	27	-0,3934	0,5018					

Der T-Test zeigt einen signifikanten Unterschied zwischen den beiden Gruppen an. Die Effektstärke $d = -0,83$ zeigt einen starken Effekt zu Ungunsten der Studierenden des Faches Integrierte Naturwissenschaften an. Die Hypothese H7 muss daher verworfen werden.

FF8: Haben Studierenden des Faches „Integrierte Naturwissenschaften“ in höheren Fachsemestern einen höheren Kompetenzstand als zu Beginn des Studiums?

H8: Studierende des Faches „Integrierte Naturwissenschaften“ weisen im höheren Fachsemester einen höheren Kompetenzstand auf als zu Beginn des Studiums ($p < 0,05$).

Der Levene-Test zur Überprüfung der Varianzgleichheit zeigt keine Verletzung der Varianzhomogenität für alle fünf PVs an ($F_{PV1}(1,57) = 2,523$, n.s. ; $F_{PV2}(1,57) = 0,465$, n.s. ; $F_{PV3}(1,57) = 0,178$, n.s. ; $F_{PV4}(1,57) = 0,003$, n.s. ; $F_{PV5}(1,57) = 0,002$, n.s.).

Tabelle 32: T-Test: Studierende des Faches Integrierte Naturwissenschaften 1. Semester gegen Semester <1.

Gruppe	N	M	SD	df	t	P (1-seitig)	d	r
Integrierte Naturwissenschaften (1.Sem)	27	-0,3931	0,4634	58	-2,19	0,015	0,57	0,28
Integrierte Naturwissenschaften (< 1. Sem)	32	-0,1295	0,4593					

Das Signifikanzniveau wird auf $p = 0,025$ gesenkt. Da eine gerichtete Hypothese vorliegt, kann mit einer 1-seitigen Signifikanzprüfung gerechnet werden. Der T-Test wird signifikant. Die Hypothese H8 wird daher gestützt.

7.5 Fragestellung 9: Kompetenzstand von Studierenden der Sozialkunde

FF9: Welchen Kompetenzstand haben Studierende eines sozialwissenschaftlichen Faches im Vergleich zu Lehramtsstudierenden der Physik?

H9: Studierende der Sozialwissenschaften weisen einen geringeren Kompetenzstand, als Studierende der Naturwissenschaften auf ($p > 0,05$).

Um diese Hypothese zu prüfen, wurde eine Gruppe von Studierenden eines sozialwissenschaftlichen Fachs getestet. Da durch die Unterschiede in der fachlichen Ausbildung bestimmte Voraussetzungen des RASCH-Modells verletzt sein könnten (z. B. ausschließliche Abhängigkeit von dem latenten Konstrukt) und es zudem nicht um eine Erhebung zur Leistungsdiagnostik sondern vielmehr um eine Abschätzung geht, wird zur Beantwortung dieser Fragestellung auf die KTT zurückgegriffen. Dabei wurde nur ein Testheft berücksichtigt, um Probleme mit unvollständigen Datensätzen zu vermeiden. Die Auswahl des Testheftes erfolgte zunächst durch eine Analyse der eingesetzten Aufgaben. Es wurden vier Testhefte identifiziert, bei denen alle Aufgaben eine (nach dem damaligen Zeitpunkt) zufriedenstellende Passung nach dem RASCH-Modell³⁴ zeigten. Die Schwierigkeiten der Aufgaben wurden pro Heft auf eine Linie geplottet (siehe Abbildung 23).

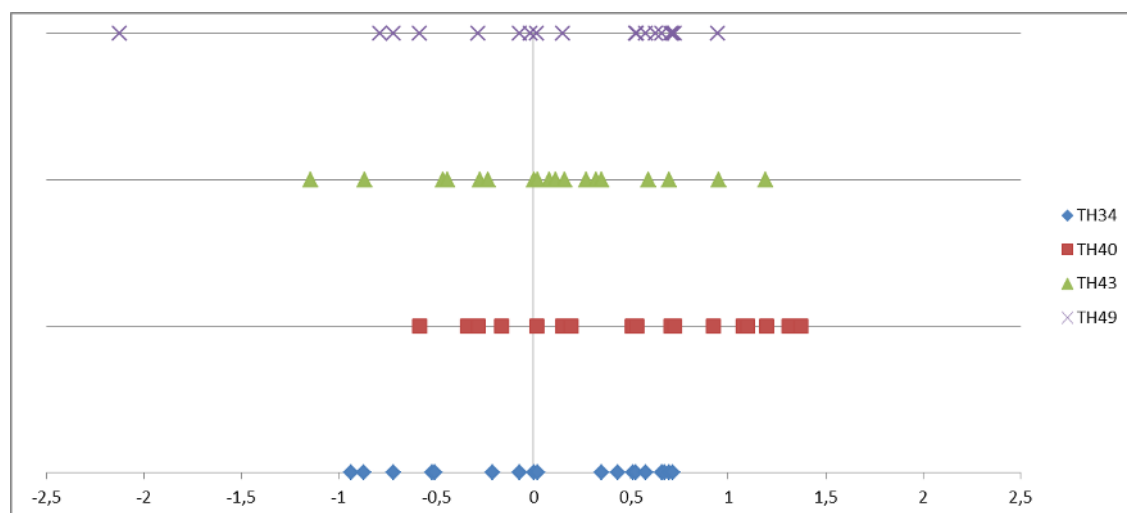


Abbildung 23: Logit-Werte der Aufgaben von vier Testheften

³⁴ Die Passung nach dem Rasch-Modell wurde trotz des Einsatzes der KTT berücksichtigt, da dadurch davon auszugehen ist, dass die Items tatsächlich das selbe Konstrukt messen.

Das Testheft sollte einen möglichst breiten Schwierigkeitsbereich abdecken. Daher wurde Testheft 43 ausgewählt. Die drei Teile Biologie, Chemie und Physik wurden in den Testheften an unterschiedliche Positionen rotiert. Es wurde eine Gruppe von 28 Studierenden³⁵ der Sozialkunde (Lehramtsstudierende mit weiterem Fach) getestet. 15 Personen sind weiblich (53,6 %), der Altersdurchschnitt liegt bei 28,5 Jahren (SD = 5,1). Keiner der Studierenden hat eines der drei Naturwissenschaften als Nebenfach. Das Testheft eines Proband/inn/en musste aus der Analyse ausgeschlossen werden, weil die Bearbeitung vorzeitig abgebrochen wurde und zehn Aufgaben unbearbeitet blieben. In anderen Heften unbearbeitete Aufgaben wurden mit 0 = Falsch kodiert, weil ausreichend Zeit zur Bearbeitung der Aufgaben zur Verfügung stand.

Die Vergleichsgruppe besteht aus Proband/inn/en der drei Naturwissenschaften, die das entsprechende Testheft 43 bearbeitet haben. Hieraus werden nur Lehramtsstudierende berücksichtigt. Ebenfalls werden alle Fälle ausgeschlossen, bei denen fehlenden Werte bei den Items vorlagen (z. B. wenn die Bearbeitung eines Testheftes aus Zeitgründen vorher abgebrochen wurde). Die Vergleichsgruppe umfasst nach Bereinigung 48 Proband/inn/en, von denen 21 weiblich und 27 männlich sind. Der Altersdurchschnitt liegt bei 24,3 (SD = 4,5). 29 Proband/inn/en sind im Bachelorstudium, 14 Proband/inn/en im Masterstudium³⁶. 35 Proband/inn/en studieren ein Fach Biologie, 7 Chemie und 13 Physik³⁷.

Aufgrund der großen inhaltlichen Distanz der naturwissenschaftlichen Studiengänge zu dem Studium der Sozialkunde und dadurch möglicher Unterschiede in der Funktion der Items werden zwei getrennte Reliabilitätsanalysen durchgeführt.

Zunächst wird für die Vergleichsgruppe der Lehramtsstudierenden mit naturwissenschaftlichem Fach die Analyse durchgeführt. Die Skalierung erreicht ein $\alpha_{\text{Cronbach}} = 0,575$. Zwei Items weisen mit $r = -0,008$ und $r = -0,048$ negative Trennschärfen auf und werden daher aus der Analyse entfernt. Mit den verbleibenden 16 Items wird eine erneute Analyse durchgeführt. Diese Skalierung erreicht $\alpha_{\text{Cronbach}} = 0,622$. Auch in dieser Skalierung bleiben zwei Items mit Trennschärfen $r < 0,05$ auffällig. Sie werden jedoch zunächst nicht entfernt.

³⁵ Die Größe der Stichprobe war durch Anzahl möglicher Proband/inn/en begrenzt. Abschätzungen mit G*Power (Faul, Erdfelder, Lang und Buchner (2007) zeigen, dass sich mittlere bis starke Effekte dadurch nachweisen lassen sollten.

³⁶ Bzw. einem semestermäßigem Äquivalent (s.o.), zu 48 fehlende Werte sind unbekannt

³⁷ Mehrfachnennungen möglich

Tabelle 33: Reliabilitätsanalyse Physik-Lehramtsstudierende

Analyse Nr.	1	2
α_{Cronbach}	0,575	0,622
Itemanzahl	18	16
Trennschärfe	-0,048 – 0,487	0,028 – 0,498
davon neg. Trennschärfe	2	0

Die zweite Reliabilitätsanalyse wird für die Gruppe der Studierenden der Sozialkunde durchgeführt. In der ersten Skalierung wird lediglich $\alpha_{\text{Cronbach}} = 0,163$ erreicht. Neun Items weisen negative Trennschärfen auf. Für die nächste Skalierung werden zunächst Items mit negativen Trennschärfen von $r < -0,05$ entfernt. Dies trifft auf fünf Items zu. Die Skalierung mit den verbleibenden 13 Items erreicht $\alpha_{\text{Cronbach}} = 0,403$. Ein Item bleibt mit einer negativen Trennschärfe weiterhin auffällig und wird ebenfalls aus der Analyse entfernt. Die Reliabilität erreicht $\alpha_{\text{Cronbach}} = 0,444$. Allerdings weist in dieser Skalierung ein weiteres Item eine negative Trennschärfe auf, was demnach auch entfernt wird. Die vierte Analyse erreicht $\alpha_{\text{Cronbach}} = 0,478$. Die Items weisen allesamt positive Trennschärfen auf. Insgesamt müssen bei der Gruppe der Sozialkunde-Studierenden sieben Items aus dem Itempool entfernt werden.

Tabelle 34: Reliabilitätsanalyse Studierende der Sozialkunde

Analyse Nr.	1	2	3	4
α_{Cronbach}	0,163	0,403	0,444	0,478
Itemanzahl	18	13	12	11
Trennschärfe	-0,145 – 0,334	-0,054 – 0,384	-0,025 – 0,433	0,028 – 0,427
davon neg. Trennschärfe	9	1	1	0

Da es sich um unterschiedliche Items handelt, die für die beiden Vergleichsgruppen entfernt wurden, verbleiben neun Items, für die jeweils positive Trennschärfen erreicht wurden. In einer dritten Reliabilitätsanalyse werden nun die Antworten aller Proband/inn/en gemeinsam analysiert.

Die Analyse erreicht $\alpha_{\text{Cronbach}} = 0,572$. Die Trennschärfe deckt einen Bereich von $r = 0,106$ bis $r = 0,509$ ab. Sechs Items weisen eine Trennschärfe von $r < 0,3$ auf. Auf eine weitere Prüfung der Hypothese wird aufgrund der geringen Trennschärfe des größten Teils der Items verzichtet. Gründe für das Scheitern dieser Untersuchung werden im Diskussions-
teil (siehe Kapitel 9) besprochen.

8. Interpretation der Ergebnisse

Die Untersuchung der Dimensionalität unterteilte sich in drei Hypothesen, die einzeln untersucht wurden. Hypothese H1a fokussierte eine mögliche Dimensionierung nach Fächern. Hierbei zeigte sich, dass sowohl die Informationskriterien, als auch der χ^2 -Test für eine eindimensionale Struktur sprechen, die nicht nach Fächern differenziert. Hypothese H1b betraf eine mögliche Differenzierung nach Arbeitsweisen. Hier wurde angenommen, dass sich eine Unterscheidung nach *Untersuchungen* und *Modelle Nutzen* auch in der Datenstruktur wieder finden lässt. Die Ergebnisse deuten auch hier insgesamt auf eine bessere Passung eines eindimensionalen Modells hin, auch wenn der χ^2 -Test zu Gunsten des siebendimensionalen Modells ausfällt. Diese bessere Passung geht auch mit einer starken Zunahme der Freiheitsgrade einher. Diese Zunahme steht im Missverhältnis zur Zunahme der Modellpassung, was durch die Informationskriterien angezeigt wird. Die Hypothese H1b muss insgesamt abgelehnt werden. Auch hier zeigt sich demnach die beste Passung eines eindimensionalen Modells. Die Hypothese H1c sagte eine bessere Passung eines eindimensionalen Modells gegenüber Modellen voraus, die sowohl nach Fächern als auch nach Arbeitsweisen unterscheiden. Die Informationskriterien, wie auch der χ^2 -Test stützen diese Hypothese. Insgesamt deuten alle drei getesteten Hypothesen auf die beste Passung eines eindimensionalen Modells hin.

Die Hypothese H2 betraf generelle Unterschiede im Kompetenzstand zwischen Lehramtsstudierenden und Fachstudierenden der Physik. Hier wurde aufgrund der geringeren Eingangsvoraussetzungen von Lehramtsstudierenden angenommen, dass ein Unterschied zwischen beiden Gruppen existiert zu Gunsten der Fachstudierenden. Die ANOVA weist hier auf einen hochsignifikanten Effekt hin. Lehramtsstudierende weisen demnach einen geringeren Kompetenzstand auf als Fachstudierende der Physik. Dabei handelt es sich um einen schwachen bis mittleren Effekt ($\eta^2 = 0,040$).

Die Hypothese H3 betraf Unterschiede im Kompetenzstand der Erkenntnisgewinnung zu verschiedenen Phasen des Physikstudiums. Hierbei wurde zunächst noch nicht zwischen Lehramtsstudierenden und Fachstudierenden unterschieden. Die ANOVA zeigt einen signifikanten Einfluss der Studienphase auf den Kompetenzstand an. Zwei T-Tests als Paarvergleiche deuten auf einen nicht signifikanten Unterschied der Kompetenz von Studienphase *Bachelor 1* zu Studienphase *Bachelor 2* und einen hochsignifikanten Unterschied von *Bachelor 2* zum *Master* hin. Entsprechend wird auch im latenten Hintergrundmodell

die Variable *Ba2* nicht signifikant. Die Effektstärke zwischen *Bachelor 2* und *Master* weist mit $d = 0,636$ auf einen mittleren Effekt hin.

Die Hypothese H4 betraf Unterschiede in den Differenzen im Kompetenzstand zu verschiedenen Phasen des Studiums von Lehramtsstudierenden und Fachstudierende der Physik. Es wurde kein Unterschied angenommen, was durch die nicht signifikanten Interaktionseffekte im latenten Hintergrundmodell *Lehramt* \times *BA2* und *Lehramt* \times *Master* sowie durch den nicht signifikanten Interaktionseffekt in der ANOVA *Studienphase***Lehramt* ($p > 0,05$) gestützt wird. Die Power-Analyse zeigt, dass für mittlere bis große Effekte die Teststärke ausreicht um Fehler zweiter Art auszuschließen. Dieses Ergebnis unterstreicht ebenfalls die Hypothese. Allerdings reicht die Teststärke nicht aus, um auch kleinere bis mittlere Effekt auszuschließen. Die Hypothese muss daher differenziert beantwortet werden. So sind zumindest keine mittleren und großen Effekte beim Unterschied in den Differenzen im Kompetenzstand zu verschiedenen Phasen des Studiums zu erwarten, jedoch bleibt eine Unsicherheit, dass doch kleine Effekte existieren.

Die Hypothese H5 fokussierte auf mögliche Unterschiede im Kompetenzstand von weiblichen und männlichen Studierenden der Physik. Aufgrund der uneindeutigen Forschungslage wurde erwartet, dass beide Gruppen sich im Kompetenzstand nicht unterscheiden. Die ANOVA fällt im Gegensatz zur Erwartung aber signifikant aus. Die errechnete Effektstärke deutet auf einen kleinen Effekt zu Gunsten der männlichen Studierenden hin ($\eta^2 = 0,024$).

Die Hypothese H6 sagte positive Effekte des Studiums eines weiteren naturwissenschaftlichen Faches voraus. So wurde angenommen, dass Studierenden mit zwei naturwissenschaftlichen Fächern einen signifikant höheren Kompetenzstand aufweisen, als Studierende die nur Physik studieren. Der T-Test fiel nicht signifikant aus, die Hypothese H6 muss daher abgelehnt werden.

Die siebte und achte Fragestellung betrafen den Kompetenzstand von Studierenden der Grundschulpädagogik mit dem Nebenfach Integrierte Naturwissenschaften. Zunächst wurde die Hypothese H7 aufgestellt, dass die Studierenden mit einem zu Physiklehramtsstudierenden vergleichbaren Kompetenzstand das Studium beginnen. Der direkte Vergleich der beiden Gruppen zeigt starke Effekte zu Ungunsten der Studierenden der Integrierten Naturwissenschaften. Eine genauere Analyse der Regressionsgewichte zeigt jedoch keinen signifikanten Effekt. Die Unterschiede zwischen beiden Gruppen sind dem-

nach nicht alleine durch die Zugehörigkeit zum Studienfach integrierte Naturwissenschaften zu erklären.

Die weitere Hypothese H8 betraf die Unterschiede im Kompetenzstand von Studierenden in höheren Fachsemestern im Vergleich zu Studierenden im ersten Semester im Rahmen des Studiums der Integrierten Naturwissenschaften. Hier wurde angenommen, dass aufgrund des Studiums Studierende höherer Fachsemester einen höheren Kompetenzstand aufweisen. Der T-Test zwischen beiden untersuchten Gruppen fällt signifikant aus. Die Hypothese H8 kann demnach mit den hier vorliegenden Daten gestützt werden.

Die Hypothese H9 sagte einen höheren Kompetenzstand von Naturwissenschaftsstudierenden gegenüber Studierenden der Sozialkunde voraus. Die Prüfung dieser Hypothese wurde aufgrund des stark dezimierten Aufgabenpools und der geringen Trennschärfe der verbliebenen Items nicht durchgeführt.

9. Diskussion

Innerhalb dieser Arbeit wurden Items mit physikalischem Kontext für die Erfassung der Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung bei Studierenden entwickelt. Es wurden bestehende Kompetenzmodelle adaptiert, auf deren Grundlage dann Items entwickelt und pilotiert wurden.

Der Kompetenztest fokussiert vor allem auf die fächerübergreifenden Aspekte der naturwissenschaftlichen Denk- und Arbeitsweisen. Einschränkend könnte argumentiert werden, dass er damit auf den kleinsten gemeinsamen Nenner der drei Fächer Biologie, Chemie und Physik fokussiert. Kritisch könnte zudem angemerkt werden, dass Forschung generell nicht so abläuft, wie sie durch das Kompetenzmodell paradigmatisch beschrieben wird. Beiden Kritikpunkten kann durchaus zugestimmt werden. Es ist aber gerade die Stärke dieses Tests, die Gemeinsamkeiten der drei Naturwissenschaften zu betonen. Insbesondere unter besonderer Berücksichtigung integrativer naturwissenschaftlicher Fächer ist dies ein Vorteil. Für die einzelnen Fächer können nun aufbauend auf diesem Test weitere Instrumente entwickelt werden, die die Besonderheiten der Fächer berücksichtigen. Der zweite Punkt ist durchaus kritischer zu sehen. Jedoch ist hier zu entgegnen, dass selbst wenn Wissenschaftlerinnen und Wissenschaftler nicht konkret so arbeiten, der dargestellte Forschungsablauf dennoch der grobe theoretische Unterbau ist, der implizit ihrer Forschung zu Grunde liegt. Schon Klahr (2000, S. 12f.), auf dessen Arbeiten das hier beschriebene Kompetenzmodell aufbaut, betont, dass ein erfolgreicher Forschender viel mehr wissen muss, als durch das Modell dargestellt wird. Das Modell berücksichtigt lediglich die mentalen Prozesse, die dem wissenschaftlichen Denken zu Grunde liegen. Diese sind domänenübergreifend. Ein/e Physiker/in ist nicht automatisch auch ein/e Chemiker/in oder Biologe/-in, aber „[...]they all use the same weak methods to help solve their respective problems. When their activity is described as search in a problem space, each can understand the rationale of the other's activity[...]“ (Klahr & Simon, 2001, S. 78). Wendet man den weiter oben definierten Begriff des Modells auf das Kompetenzmodell an, so wird deutlich, dass eben auch in diesem Modell nur bestimmte Dinge berücksichtigt und andere weggelassen wurden. Trotzdem bleibt das Kompetenzmodell ein Modell für die reale Forschungstätigkeit von Wissenschaftlerinnen und Wissenschaftlern. Wie jedes Modell ist es eine *Annäherung an die Wirklichkeit*, es bleibt aber unvollständig. So bleiben einige Eigenheiten der Physik innerhalb des Modells unberührt. Dazu zählen unter anderem das Problem der Nebenhypothese (Duhem-Quine-These), der Umgang mit Wahrscheinlichkeitshypothesen, der

praktische Umgang mit Experimenten und Geräten sowie Aspekte der *Nature of Science*. Letztere fallen beide aber ohnehin unter andere Aspekte der naturwissenschaftlichen Erkenntnisgewinnung (Mayer, 2007). Dies ist demnach nicht unbedingt als Mangel des vorliegenden Tests anzusehen sondern vielmehr als Forschungsdesiderat.

Im Rahmen der Aufgabenentwicklung wurden bereits bewährte Methoden und Verfahrensweisen wie eine Entwicklung entlang eines Manuals und regelmäßige Expertenratings eingesetzt. Als sehr erfolgreich kann außerdem die Konstruktion der MC-Antwortoption auf Grundlage von echten Studierendenantworten gesehen werden. Zum einen ersparen diese Zeit, weil diese schon vorliegen, zum anderen kommen die oben erwähnten Vorteile wie eine höhere Authentizität zum Tragen. Nach Abschluss verschiedener Pilotierungsrunden konnten daher 20 Testhefte mit insgesamt 123 Items erstellt und eingesetzt werden.

Die erste Fragestellung betraf die Dimensionierung des Kompetenzstrukturmodells. Hier wurden verschiedene, aus der Theorie begründete, mehrdimensionale Modelle getestet. Es zeigte sich aber jeweils, dass die Evidenz für ein eindimensionales Modell am größten ist. So sprechen die Ergebnisse für die bessere Passung eines eindimensionalen Modells im Vergleich zu einem, das nach Fächern unterscheidet. Dies entspricht den Ergebnissen, die auch schon für die Gesamt- (Hartmann & Mathesius et al., 2015) und Teilstichprobe (Physik- und Chemiestudierende) (Stiller et al., 2015) gefunden wurden. Dass das gleiche Ergebnis auch für eine (homogenere) Stichprobe von Physikstudierenden gefunden wurde, unterstreicht diese vorhergenannten Ergebnisse nochmals deutlich. Es entspricht auch den theoretischen Ergebnissen, dass die schwachen wissenschaftlichen Methoden in allen (Natur-)Wissenschaftsdisziplinen (Klahr, 2000) zu finden sind und dass sich Wissenschaftlerinnen und Wissenschaftler unabhängig vom Fachwissen darüber verständigen können (s. o.).

Auch für eine Strukturierung nach Kompetenzdimensionen ergibt sich keine Evidenz. Dies deckt sich mit den Erkenntnissen von anderen Studien (z.B. Gut-Glanzmann, 2012; Wellnitz, 2012; Woitkowski, 2015). Allerdings gab es auch Studien, die mit einem ähnlichen Kompetenzmodell zu gegenteiligen Ergebnissen kamen: So konnte Grube (2010, S. 57) für den Bereich Experimentieren eine bessere Passung eines vierdimensionalen Modells berichten. Die zitierten Studien fokussieren allerdings immer nur auf eine Fachwissenschaft. Insofern können Vergleiche lediglich Anhaltspunkte bieten. Studien, die eine fächerübergreifende Kompetenzmodellierung in diesem Bereich zum Ziel hatten, sind

nicht bekannt. Es bleibt bei dem hier vorliegenden Test allerdings eine gewisse Unsicherheit, dass eine Dimensionierung möglicherweise besteht, diese aber auf Grund der, durch das Multi-Matrix-Design bedingten, geringen Anzahl an Items pro Dimension, nicht nachweisbar ist (vgl. Gonzales & Rutkowski, 2010).

Entsprechend der Ergebnisse der ersten beiden Dimensionsprüfungen zeigte auch eine Unterscheidung nach Fächern und Kompetenzdimensionen keine bessere Passung als das eindimensionale Modell.

Bei der anschließenden Itemselektion, die auf Grundlage des eindimensionalen Modells geschah, wurden Items entfernt, die eine mangelnde Passung auf das RASCH-Modell oder ein DIF zeigten. Insgesamt wurden dabei aber nur fünf Items entfernt. Der Einfluss auf die inhaltlichen Aspekte des Modells sollte demnach gering sein, weil alle Zellen des Kompetenzmodells weiterhin mit ausreichend Items abgedeckt sind. Für den aktuell im Projekt *ValidiS*³⁸ weiterverfolgten Längsschnitt wurden aufgrund dieser Ergebnisse Testhefte konstruiert, die nur noch RASCH-konforme Items enthielten. Aufgrund der zeitlichen Einschränkungen des Projekts war dies für diese Studie aber nicht mehr möglich, weshalb die oben genannten fünf Items für die Auswertung entfernt werden mussten.

Die in der Auswertung verbleibenden Items zeigen eine zufriedenstellende Passung auf das verwendete RASCH-Modell. Die Reliabilität ist mit $Rel_{EAP/PV}=0,500$ zwar niedrig, und in der Nomenklatur des Cronbachs α (welches als äquivalent angesehen werden kann (Rost, 2004)) als „schlecht“ einzuordnen. Allerdings weisen andere Kompetenztests zur Erkenntnisgewinnung ähnliche Ergebnisse auf (z. B. Woitkowski (2015, S. 191): $Rel_{EAP/PV} = 0,518$; Wellnitz (2012, S. 110): $Rel_{EAP/PV} = 0,59$). Zudem hängt die Stabilität der Messung (und damit die Reliabilität) mit der Anzahl der Items zusammen (vgl. Linacre, 1994), die bedingt durch die Testlänge mit 18 Items kurz ist. Es wird außerdem argumentiert, dass der Nutzen dieses Reliabilitätsmaßes bei den hier genutzten Matrix-Samplings ohnehin wenig Aussagekraft hat (Adams, 2005).

Bei der Betrachtung des Kompetenzstandes wurden Unterschiede bei Lehramtsstudierenden und Fachstudierenden der Physik nachgewiesen. Im Mittel weisen Fachstudierende einen höheren Kompetenzstand auf. Bei der Betrachtung des Kompetenzstandes zu verschiedenen Phasen des Studiums zeigt sich erwartungsgemäß ein Unterschied im Kompetenzstand zwischen den Phasen Bachelor 2 und Master. Zwischen den Phasen Bachelor 1

³⁸ Validierungsstudie zum wissenschaftlichen Denken im naturwissenschaftlichen Studium

und Bachelor 2 wird hingegen kein signifikanter Zuwachs angezeigt. Weiterhin konnten keine Interaktionseffekte zwischen den Variablen Studienphase und Lehramt nachgewiesen werden. Interpretiert man diese Ergebnisse im Sinne eines Quasi-Längsschnitts, sprechen diese dafür, dass sich die Entwicklung dieser Kompetenz zwischen Lehramtsstudierenden und Fachstudierenden nicht unterscheidet. Bringt man dieses Ergebnis mit dem Ergebnis aus Fragestellung zwei zusammen, so zeigt sich, dass Lehramtsstudierende mit einem niedrigeren Kompetenzstand das Studium beginnen und es auch mit einem niedrigen Kompetenzstand als Fachstudierende verlassen. Dabei entwickeln sich beide Gruppen aber annähernd parallel. Entgegen früherer Argumentationen scheint also kein Unterschied zwischen der impliziten Vermittlung z. B. in Laborpraktika für Fachstudierende und der expliziten Vermittlung dieser Kompetenz in Didaktikveranstaltungen zu bestehen (vgl. Straube & Nordmeier, 2014). Die Ergebnisse deuten vielmehr darauf hin, dass der Kompetenzerwerb gleich gut gelingt, die Differenzen aber aus unterschiedlichen Eingangsvoraussetzungen resultieren. Dies deckt sich mit Ergebnissen von Albrecht (2011, S. 80, 86), der bei Lehramtsstudierenden im Vergleich zu Fachstudierenden feststellte, dass diese weniger häufig Mathematik- oder Physikleistungskurse in der Schule besuchten und eine geringere Note in der Hochschulzugangsberechtigung aufwiesen. Es scheint nicht zu gelingen, diese Differenzen innerhalb des Studiums auszugleichen. Aus dieser vorliegenden Studie geht aber nicht hervor, ob ein derartiger Ausgleich überhaupt notwendig ist. Möglicherweise reicht das erreichte Niveau für das Lehramt entsprechend aus.

Diese Ergebnisse stehen teilweise im Widerspruch zu den Ergebnissen der Gesamtstichprobe (Hartmann & Mathesius et al., 2015). Hier zeigte sich zwar ebenfalls ein genereller Vorteil für die Fachstudierenden. Bei genauerer Untersuchung wird aber deutlich, dass ein Interaktionseffekt zum Tragen kommt: So haben im Bachelorstudium die Fachstudierenden zwar einen Vorteil, dieser kehrt sich aber im Masterstudium um. Um die Differenzen zu den hier dargestellten Ergebnissen interpretieren zu können, sind fächervergleichende Analysen notwendig. Denkbar wäre z. B. dass Fachstudierende der Physik aufgrund ihres Studiums einen stärkeren Kompetenzzuwachs aufweisen als die Fachstudierenden der anderen Fächer. Genauso wäre es aber auch möglich, dass die Ausbildung im Bereich Erkenntnisgewinnung für Lehramtsstudierende der Physik weniger erfolgreich als in den anderen Fächern ist.

Eine weitere Fragestellung betraf mögliche Unterschiede im Kompetenzstand von weiblichen und männlichen Studierenden. Bisherige Forschungsarbeiten zeigten ein heteroge-

nes Bild. Allerdings wurden entweder nur Schülerinnen und Schüler untersucht, oder die Untersuchung fokussierte sich nur auf ein Fach. Entsprechend wurde hierbei zunächst die Nullhypothese angenommen. Die Ergebnisse dieser Studie zeigen, dass die Nullhypothese abgelehnt werden muss, also tatsächlich Unterschiede im Kompetenzstand existieren – zu Gunsten der männlichen Studierenden. Die uneinheitliche Forschungslage macht die Interpretation dieses Ergebnisses schwer. Ausgehend von der Annahme, dass Studierende beider Geschlechter sich bewusst für ein derartiges Studium entschieden haben, sollten Selbstbild- oder Stereotypeffekte, wie sie zum Beispiel von Hannover und Kessels (2002) beschrieben werden, keinen Einfluss haben. Möglicherweise kommen hierbei auch wieder unterschiedliche Eingangsvoraussetzungen zum Tragen. Riese und Reinhold (2012) diskutieren einen möglichen Einfluss von Effekten des physikalischen Selbstkonzepts, können aber die ebenfalls in ihrem Fachwissenstest vorhandenen Unterschiede nicht erklären. An diesem Punkt sollte weitere Forschung anschließen. Unter anderem könnte in einem mehrdimensionalen Modell nach Fächern untersucht werden, ob die in der Sekundarstufe I nach Fächern variierenden Vorteile (Schroeders et al., 2013) sich auch in der Hochschulausbildung fortsetzen.

Generell ist für diese Fragestellungen aber anzumerken, dass das verwendete Modell lediglich 13 % der Varianz aufklären kann. Demnach scheinen weitere Prädiktoren zu existieren, die im Rahmen dieser Studie nicht untersucht wurden.

Nur die Lehramtsstudierenden betreffend wurde außerdem untersucht, ob Unterschiede im Kompetenzstand zwischen solchen, die Physik und ein weiteres naturwissenschaftliches Fach studieren und solchen, die Physik und ein anderes, nicht naturwissenschaftliches Fach studieren, existieren. Es wurde die Hypothese aufgestellt, dass auf Grund der höheren Anzahl an Lerngelegenheiten Lehramtsstudierende mit einem weiteren naturwissenschaftlichen Fach einen höheren Kompetenzstand aufweisen. Die Hypothese wird abgelehnt. Der Datensatz offenbart keine Evidenz für einen derartigen Unterschied. Dies deckt sich mit vorhergehenden, vorläufigen Analysen (Straube & Nordmeier, 2014), widerspricht aber Ergebnissen, die für die Gesamtkohorte errechnet wurden (Hartmann & Mathesius et al., 2015)³⁹. Anscheinend haben Lehramtsstudierende der Physik keinen Vorteil von einem weiteren naturwissenschaftlichen Fach. Dies kann auf zwei Weisen interpretiert werden: Entweder nutzen die Studierenden mit zwei naturwissenschaftlichen

³⁹ Es sei angemerkt, dass diese Analyse in die Gruppe der Studierenden mit einem Fach, auch Fachstudierende einbezogen, also nicht ausschließlich Lehramtsstudierende

Fächern ihren Vorteil nicht aus oder aber die Lehramtsstudierenden mit einem Fach erreichen einen höheren Kompetenzstand als Studierende anderer Fächer (Biologie oder Chemie) mit nur einem Fach. Hier können detailliertere Analysen der Gesamtkohorte Aufschluss darüber geben.

Genauer untersucht wurde auch der Kompetenzstand von Studierenden mit einem Fach Integrierte Naturwissenschaften. Ein T-Test zeigt einen Unterschied im Kompetenzstand zu Beginn des Studiums im Vergleich zu Lehramtsstudierenden der Physik zu Beginn des Studiums an. Im latenten Hintergrundmodell dieser Modellierung wird diese Variable aber nicht signifikant. Demnach lässt sich dieser Effekt nicht alleine durch die Zugehörigkeit zur Gruppe der Studierenden der Integrierten Naturwissenschaften erklären. Möglicherweise kann der Unterschied durch den übergroßen Anteil an weiblichen Studierenden in dieser Gruppe erklärt werden, auch wenn in dieser Modellierung die Prädiktorvariable *weiblich* ebenfalls nicht signifikant wurde. Allerdings zeigte sich in vorhergehenden Untersuchungen mit Physik(lehramts)studierenden diese Variable als hoch signifikant. Der durch den T-Test angezeigte Unterschied kommt demnach möglicherweise durch eine Wechselwirkung beider Variablen zustande. Genauere Analysen können aber aufgrund der zu geringen Größe der Stichprobe nicht durchgeführt werden.

Eine weitere Hypothese zielte auf Differenzen im Kompetenzstand von Studierenden im ersten Semester verglichen mit Studierenden in höheren Fachsemestern im Rahmen des Studiums der Integrierten Naturwissenschaften ab. Diese Hypothese konnte durch die vorliegenden Daten gestützt werden. Es zeigte sich ein mittlerer Effekt. Die in den ersten Semestern stattfindenden Veranstaltungen zur Wissenschaftstheorie und zum wissenschaftlichen Arbeiten zeigen vermutlich hier ihre Wirkung.

Die letzte Fragestellung beschäftigte sich mit der Frage, welchen Kompetenzstand Studierende eines sozialwissenschaftlichen Faches im Vergleich zu Studierenden mit naturwissenschaftlichen Fächern aufweisen. Die Beantwortung dieser Fragestellung führte zu erheblichen Problemen, da in der Gruppe der Sozialkunde-Studierenden ein Großteil der Items negative Trennschärfen aufwies. Dies kann möglicherweise durch motivationale Effekte erklärt werden. So ist es denkbar, dass einige Proband/inn/en den Kompetenztest nicht ernst genommen und ohne langes Nachdenken die Antworten gegeben haben. Weiterhin scheint es aber auch plausibel, dass bestimmte Begriffe nur Studierenden der naturwissenschaftlichen Fächer bekannt sind und diese daher das Antwortverhalten beeinflussen haben. Denkbar und sehr wahrscheinlich ist daher die Existenz von DIF bei den

Items, der aber im Rahmen dieser Studie nicht untersucht wurde. Von den insgesamt eingesetzten 18 Items wiesen nur noch neun Items in beiden Gruppen positive Trennschärfen auf. Von diesen neun Items zeigten aber sechs Items Trennschärfen von $r < 0,3$ auf. Dementsprechend war die Reliabilität dieser verbliebenen Items mit $\alpha_{\text{Cronbach}} = 0,572$ ebenfalls niedrig. Auf eine Prüfung der Hypothese wurde demnach verzichtet. Die bei der Gruppe der Lehramtsstudierenden mit naturwissenschaftlichen Fächern bei zwei Items auftretende negative Trennschärfe hat keine Relevanz für die weiteren Ergebnisse dieser Studie. Ausschlaggebend bleibt die Itemselektion im Rahmen der RASCH-Modellierung.

Ausblick

Nach dieser Bestandsaufnahme des Kompetenzstandes der Lehramtsstudierenden folgt als Forschungsdesiderat zunächst, die hier erhaltenen Ergebnisse anhand eines echten Längsschnittes zu überprüfen. Weiterhin sind konkrete Kompetenzstufen abzuleiten und forschungsbasiert Mindest- und Regelstandards für Lehramtsstudierende festzulegen. Auch die weitere Kompetenzentwicklung nach Abschluss des Studiums, während des Referendariats und des Berufslebens sollte erforscht werden. Hierbei ist auch zu untersuchen, welchen konkreten Einfluss der Kompetenzstand der Lehrenden auf die Kompetenzentwicklung der Lernenden hat. Weiterhin können im Rahmen des Studiums konkrete Interventionen getestet werden, um somit das Lehramtsstudium in dieser Beziehung weiter zu optimieren. Für den regelmäßigen Einsatz zur Erhebung des Kompetenzstandes muss der Kompetenztest anwenderfreundlicher werden. Hier wäre denkbar, diesen Test in ein *Computer adaptiv Test* (CAT) (Bortz & Döring, 2006) zu integrieren. Dadurch wäre es auch möglich, völlig neue kompetenzbezogene Prüfungsformen in das Studium einzubeziehen (siehe z. B. Huber, 2008). Einige dieser Desiderate werden bereits im Nachfolgeprojekt *ValidiS* thematisiert.

10. Literaturverzeichnis

- Abell, S. K. (2007). Research on science teacher knowledge. In S. K. Abell & N. G. Lederman (Hrsg.), *Handbook of research on science education* (S. 1105–1150). Mahwah, NJ: Erlbaum.
- Adams, R. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation*, 31 (2-3), 162–172.
- Adams, R. & Wu, M. (ACER, Hrsg.). (2010). *Unidimensional Latent Regression*. Zugriff am 22.05.2015. Verfügbar unter <http://www.acer.edu.au/files/Conquest-Tutorial-5-UnidimensionalLatentRegression.pdf>
- Adams, R., Wu, M. & Wilson, M. (2012) ConQuest 3.0.1 [Computer software]: Australian Council for Educational Research.
- Aiello-Nicosia, M. L., Sperandio-Mineo, R. M. & Valenza, M. A. (1984). The relationship between science process abilities of teachers and science achievement of students: an experimental study. *Journal of Research in Science Teaching*, 21 (8), 853–858.
- Albrecht, A. (2011). *Längsschnittstudie zur Identifikation von Risikofaktoren für einen erfolgreichen Studieneinstieg in das Fach Physik*. Dissertation, Freie Universität Berlin. Berlin. Verfügbar unter http://www.diss.fu-berlin.de/diss/servlets/MCRFileNodeServlet/FUDISS_derivate_000000010456/Dissertation_Druckversion_Andre_Albrecht_UB.pdf
- Andersson, G. (1988). *Kritik und Wissenschaftsgeschichte. Kuhns, Lakatos' und Feyerabends Kritik des kritischen Rationalismus* (Die Einheit der Gesellschaftswissenschaften, Bd. 54). Tübingen: Mohr, J.C.B., (Paul Siebeck).
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2000). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (Springer-Lehrbuch, 9., überarb. und erw. Aufl.). Berlin: Springer.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 469–520.
- Baumert, J., Stanat, P. & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider et al. (Hrsg.), *PISA 2000* (S. 15–68). Berlin, Heidelberg: Springer.
- Bleichroth, W., Dahncke, H., Jung, W., Kuhn, W., Merzyn, G. & Weltner, K. (1999). *Fachdidaktik Physik* (2. Aufl.). Köln: Aulis-Verlag.
- Blömeke, S., Suhl, U., Kaiser, G., Felbrich, A., Schmotz, C. & Lehmann, R. (2010). Lerngelegenheiten und Kompetenzerwerb angehender Mathematiklehrkräfte im internationalen Vergleich. *Unterrichtswissenschaft*, 38 (1), 29–50.
- Blömeke, S., Zlatkin-Troitschanskaia, O., Kuhn, C. & Fege, J. (2013). Modeling and measuring competencies in higher education: tasks and challenges. In S. Blömeke, O. Zlatkin-Troitschanskaia, C. Kuhn & J. Fege (Hrsg.), *Modeling and measuring competencies in higher education*.

- Tasks and challenges* (Professional and VET learning, Bd. 1, S. 1–10). Rotterdam: Sense Publishers.
- Bond, T. G. & Fox, C. M. (2007). *Applying the Rasch model. Fundamental measurement in the human sciences*. Mahwah, NJ: Erlbaum.
- Boone, W. J., Staver, J. R. & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Dordrecht: Springer.
- Borowski, A., Kirschner, S., Liedtke, S. & Fischer, H. E. (2011). Vergleich des Fachwissens von Studierenden, Referendaren und Lehrenden in der Physik. *Physik und Didaktik in Schule und Hochschule*, 10 (1), 1–9.
- Borowski, A., Neuhaus, B. J., Tepner, O., Wirth, J., Fischer, H. E., Leutner, D. et al. (2010). Professionswissen von Lehrkräften in den Naturwissenschaften (ProwiN). Kurzdarstellung des BMBF-Projekts. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 341–349.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler*. Heidelberg: Springer Medizin.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation*. Berlin: Springer.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (Psychologie, 3., aktualisierte und erw. Aufl). München: Pearson Studium.
- Bühner, M. & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler* (Ps Psychologie). München: Pearson Studium.
- Burton, S., Sudweeks, R., Merrill, P. & Wood, B. (1991). *How to Prepare Better Multiple-Choice Test Items: Guidelines for University Faculty*: Brigham Young University Testing Services and the Department of Instructional Science.
- Bybee, R. W. (2002). Scientific Literacy - Mythos oder Realität. In W. Gräber, P. Nentwig, T. Koballa & R. Evans (Hrsg.), *Scientific Literacy. Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung* (S. 21–43). Opladen: Leske + Budrich.
- Cauet, E., Borowski, A. & Fischer, H. E. (2015). Testen wir relevantes Wissen? Professionswissen von Physiklehrkräften. In S. Bernholt (Hrsg.), *Heterogenität und Diversität - Vielfalt der Voraussetzungen im naturwissenschaftlichen Unterricht* (S. 136–138). Kiel: IPN. Zugriff am 22.04.2015. Verfügbar unter http://www.gdcp.de/images/tb2015/TB2015_136_Cauet.pdf
- Chalmers, A. F. (2007). *Wege der Wissenschaft. Einführung in die Wissenschaftstheorie* (Fünfte, völlig überarbeitete und erweiterte Auflage). Berlin, Heidelberg: Springer.
- Cohen, J. (1988). *Statistical power analysis for the behavioural science*. Hillsdale: Lawrence Erlbaum.
- Davies, M. v., Gonzales, E. & Mislevy, R. J. (2009). What are plausible values and why are they useful? In M. v. Davies & D. Hastedt (Hrsg.), *Issues and Methodologies in Large-Scale Assessments* (IERI Monograph Series, Bd. 2, S. 9–36). Princeton, Hamburg: IERInstitute.
- Deutsche physikalische Gesellschaft. (2006). *Thesen für ein modernes Lehramtsstudium im Fach Physik*. Bad Honnef: Deutsche physikalische Gesellschaft.

- Deutsche physikalische Gesellschaft. (2014). *Zur fachlichen und fachdidaktischen Ausbildung für das Lehramt Physik*. Zugriff am 03.08.2015. Verfügbar unter <https://www.dpg-physik.de/veroeffentlichung/broschueren/studien/lehramtsstudie-2014.pdf>
- Drieschner, E. (2008). *Bildungsstandards praktisch. Perspektiven kompetenzorientierten Lehrens und Lernens* (1. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Duhem, P. (1998). Physical Theory and Experiment. In M. Curd & J. A. Cover (Hrsg.), *Philosophy of science. The central issues* (1. Aufl., S. 257–279). New York: W.W. Norton & Co.
- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists* (Multivariate applications book series). Mahwah, N.J.: Erlbaum.
- Emereole, H. U. (2009). Learners' and teachers' conceptual knowledge of science processes: the case study of botswana. *International Journal of Science and Mathematics Education*, 7, 1033–1056.
- Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39 (2), 175–191.
- Feyerabend, P. (1983). *Wider den Methodenzwang* (Weisses Programm im 33. Jahr Suhrkamp, 2. Aufl.). Frankfurt/M.: Suhrkamp.
- Field, A. (2009). *Discovering statistics using SPSS* (3. ed., reprinted 2009 (twice)). Los Angeles, Calif.: Sage.
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R. et al. (2014). Scientific Reasoning and Argumentation: Advancing an Interdisciplinary Research Agenda in Education. *Frontline Learning Research*, 5, 28–45.
- Frey, A., Hartig, J. & Rupp, A. A. (2009). An NCME Instructional Module on Booklet Designs in Large-Scale Assessments of Student Achievement: Theory and Practice. *Educational Measurement: Issues and Practice*, 28 (3), 39–53.
- Giere, R. N., Bickle, J. & Mauldin, R. F. (2006). *Understanding scientific reasoning* (5th ed). Belmont, CA: Thomson/Wadsworth.
- GÓMEZ-BENITO, J. & Navas-ara, M. J. (2000). A Comparison of χ^2 , RFA and IRT Based Procedures in the Detection of DIF. *Quality and Quantity*, 34 (1), 17–31.
- Gonzales, E. & Rutkowski, L. (2010). Principles of multiple matrix booklet designs and parameter recovery in largescale assessments. In M. v. Davier & D. Hastedt (Hrsg.), *Issues and Methodologies in Large-Scale Assessments* (IERI Monograph Series, Bd. 3, S. 125–156). Princeton, Hamburg: IERInstitute. Zugriff am 22.05.2015. Verfügbar unter http://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_03_Chapter_6.pdf
- Gott, R. & Duggan, S. (1995). *Investigative Work in the Science Curriculum. Developing Science and Technology Education*. Buckingham: Open University Press.

- Grossman, P. L., Wilson, S. M. & Shulman, L. (1989). Teacher of Substance: Subject Matter Knowledge for Teaching. In M. C. Reynolds (Hrsg.), *Knowledge Base for the Beginning Teacher* (S. 23–36). Oxford: Pergamon Press.
- Grube, C. (2010). *Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung. Untersuchung der Struktur und Entwicklung des wissenschaftlichen Denkens bei Schülerinnen und Schülern der Sekundarstufe I*. Dissertation, Universität Kassel. Kassel.
- Gut-Glanzmann, C. (2012). *Modellierung und Messung experimenteller Kompetenz. Analyse eines large-scale Experimentiertests*. Dissertation, Universität Basel. Basel. Zugriff am 04.08.2015. Verfügbar unter https://www.phzh.ch/MAPortrait_Data/158541/1/Gut_2012_Experimentelle_Kompetenz.pdf
- Hamman, M. (2007). Das Scientific Discovery as Dual Search-Modell. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biogiedidaktischen Forschung. Ein Handbuch für Lehramtsstudenten und Doktoranden* (S. 187–196). Berlin: Springer.
- Hannover, B. & Kessels, U. (2002). Challenge the science-stereotype. Der Einfluss von Technik-Freizeitkursen auf das Naturwissenschaften-Stereotyp von Schülerinnen und Schülern. In M. Prenzel & J. Doll (Hrsg.) *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen. Zeitschrift für Pädagogik*. 45, 341–358 [Themenheft]. Weinheim: Beltz.
- Hartig, J. (2008). Kompetenzen als Ergebnisse von Bildungsprozessen. In Bundesministerium für Bildung und Forschung (Hrsg.), *Kompetenzerfassung in pädagogischen Handlungsfeldern. Theorien, Konzepte und Methoden* (S. 15–25). Bonn: BMBF.
- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Springer-Lehrbuch* (S. 143–171). Berlin: Springer.
- Hartig, J. & Höhler, J. (2010). Modellierung von Kompetenzen mit mehrdimensionalen IRT-Modellen. Projekt MIRT. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (S. 189–198). Weinheim: Beltz.
- Hartig, J. & Jude, N. (2007). Empirische Erfassung von Kompetenzen und psychometrische Kompetenzmodelle. In J. Hartig & E. Klieme (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik. Eine Expertise im Auftrag des Bundesministeriums für Bildung und Forschung* (Bildungsforschung, Bd. 20, S. 17–30). Bonn: BMBF.
- Hartig, J. & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik. Mit 18 Tabellen* (S. 127–143). Heidelberg: Springer Medizin.
- Hartmann, S., Mathesius, S., Stiller, J., Straube, P., Upmeyer zu Belzen, A. & Krüger, D. (2015). Kompetenzen der naturwissenschaftlichen Erkenntnisgewinnung als Teil des Professionswissens zukünftiger Lehrkräfte: Das Projekt Ko-WADiS. In B. Koch-Priewe, A. Köker, J. Seifried & E. Wuttke (Hrsg.), *Kompetenzerwerb an Hochschulen: Modellierung und Messung. Zur Profes-*

- nalisierung angehender Lehrerinnen und Lehrer sowie fröhpädagogischer Fachkräfte*. Bad Heilbrunn: Klinkhart.
- Hartmann, S., Upmeier zu Belzen, A., Krüger, D. & Pant, H. A. (2015). Scientific Reasoning in Higher Education. *Zeitschrift für Psychologie*, 223 (1), 47–53.
- Hofstein, A., Navon, O., Kipnis, M. & Mamlok-Naaman, R. (2005). Developing students' ability to ask more and better questions resulting from inquiry-type chemistry laboratories. *Journal of Research in Science Teaching*, 42 (7), 791–806.
- Huber, L. (2008). 'Kompetenzen' prüfen? In S. Dany, B. Szczyrba & J. Wildt (Hrsg.), *Prüfungen auf die Agenda! Hochschuldidaktische Perspektiven auf Reformen im Prüfungswesen* (1. Aufl.). s.l: Bertelsmann W. Verlag.
- Hunger, E. (1964). *Grundbegriffe physikalischen Denkens*. Frankfurt a.M.: Hirschgraben Verlag.
- IBM Corp. (2013) IBM SPSS Statistics for Windows, Version 22.0 [Computer software].
- Justi, R. & Gilbert, J. (2003). Teachers' views on the nature of models. *International Journal of Science Education*, 25 (11), 1369–1386.
- Kane, M. T. (1992). The Assessment of Professional Competence. *Evaluation & the Health Professions*, 15 (2), 163–182.
- Kircher, E. (1995). *Studien zur Physikdidaktik. Erkenntnis- und wissenschaftstheoretische Grundlagen* (IPN, Bd. 145). Kiel: IPN.
- Kircher, E. & Dittmer, A. (2004). Lehren und Lernen über die Natur der Naturwissenschaften - ein Überblick. In C. Höbke, D. Höttecke & E. Kircher (Hrsg.), *Lehren und Lernen über die Natur der Naturwissenschaften* (S. 1–22). Baltmannsweiler: Schneider-Verlag Hohengehren.
- Kirschner, S. (2013). *Modellierung und Analyse des Professionswissens von Physiklehrkräften* (Studien zum Physik- und Chemielernen, Bd. 161).
- Klahr, D. & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12 (1), 1–48.
- Klahr, D. (2000). *Exploring science. The cognition and development of discovery processes*. Cambridge, Massachusetts: MIT Press.
- Klahr, D. & Simon, H. A. (2001). What Have Psychologists (And Others) Discovered about the Process of Scientific Discovery? *Current Directions in Psychological Science*, 10 (3), 75–79. Verfügbar unter <http://www.jstor.org/stable/20182703>
- Klemm, A. (in Vorbereitung). *Evaluierung von Items zur naturwissenschaftlichen Erkenntnisgewinnung [Arbeitstitel]*. Masterarbeit, Freie Universität Berlin. Berlin.
- Klieme, E. & Hartig, J. (2008). Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. In M. Prenzel, I. Gogolin & H. Krüger (Hrsg.), *Kompetenzdiagnostik: Zeitschrift für Erziehungswissenschaft. Sonderheft 8 2007* (S. 11–29). VS Verlag für Sozialwissenschaften GmbH. Zugriff am 15.12.2014. Zugriff am 15.12.2014.

- Klieme, E. & Leutner, D. (2006). *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Überarbeitete Fassung des Antrags an die DFG auf Einrichtung eines Schwerpunktprogramms*. Zugriff am 30.08.2013. Verfügbar unter <http://kompetenzmodelle.dipf.de/pdf/rahmenantrag>
- Klieme, E., Maag-Merki, K. & Hartig, J. (2007). Kompetenzbegriff und Bedeutung von Kompetenzen im Bildungswesen. In J. Hartig & E. Klieme (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik. Eine Expertise im Auftrag des Bundesministeriums für Bildung und Forschung* (Bildungsforschung, Bd. 20, S. 5–15). Bonn: BMBF.
- Klos, S., Henke, C., Kieren, C., Walpulski, M. & Sumfleth, E. (2008). Naturwissenschaftliches Experimentieren und chemisches Fachwissen – zwei verschiedene Kompetenzen. *Zeitschrift für Pädagogik*, 54 (3), 304–312.
- Köller, O. (2014). Naturwissenschaftliche Leistungen, demographische Veränderungen und Lehrerbildung. *Zeitschrift für Didaktik der Naturwissenschaften*, 20 (1), 3–9.
- Konferenz der europäischen Hochschulministerinnen und -minister. (2003). *Den Europäischen Hochschulraum verwirklichen. Communiqué der Konferenz der europäischen Hochschulministerinnen und -minister am 19. September 2003 in Berlin*. Zugriff am 03.03.2014. Verfügbar unter http://www.bologna-berlin2003.de/pdf/Communique_dt.pdf
- Koslowski, B. (1996). *Theory and evidence. The development of scientific reasoning* (Learning, development, and conceptual change). Cambridge, Massachusetts: MIT Press.
- Kreisel, G. (1980). Modell. In J. Speck (Hrsg.), *Handbuch wissenschaftstheoretischer Begriffe* (UTB, Bd. 967, Bd. 2, S. 437–440). Göttingen: Vandenhoeck & Ruprecht.
- Kröger, J., Neumann, K. & Petersen, S. (2014). Erfassung des Professionswissens angehender Physiklehrkräfte im Rahmen des Projekts KiL. In S. Bernholt (Hrsg.), *Naturwissenschaftliche Bildung zwischen Science- und Fachunterricht* (S. 117–119). Kiel: GDGP. Zugriff am 24.04.2015. Verfügbar unter http://www.gdcp.de/images/tb2014/TB2014_117_Kroeger.pdf
- Kröger, J., Neumann, K. & Petersen, S. (2015). Struktur und Entwicklung des Professionswissens angehender Physiklehrkräfte. In S. Bernholt (Hrsg.), *Heterogenität und Diversität - Vielfalt der Voraussetzungen im naturwissenschaftlichen Unterricht* (S. 106–108). Kiel: IPN. Zugriff am 24.05.2015. Verfügbar unter http://www.gdcp.de/images/tb2015/TB2015_106_Kroeger.pdf
- Kuhn, D., Amsel, E. & O'Loughlin, M. (1988). *The development of scientific thinking skills* (Developmental psychology series). San Diego: Academic Press.
- Kuhn, T. S. (1974a). Logik der Forschung oder psychologie der wissenschaftlichen Arbeit. In I. Lakatos & A. Musgrave (Hrsg.), *Kritik und Erkenntnisfortschritt* (Its Abhandlungen, Bd. 4, S. 1–24). Braunschweig: Vieweg.
- Kuhn, T. S. (1976b). *Die Struktur wissenschaftlicher Revolutionen* (Suhrkamp Taschenbuch Wissenschaft, Bd. 25, 2. Aufl.). Frankfurt am Main: Suhrkamp.
- Kultusministerkonferenz. (2005). *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss (Jahrgangsstufe 10)*. München: Wolter Kluwer (Beschlüsse der Kultusministerkonferenz).

- Kultusministerkonferenz. (2010). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung*. München: Wolter Kluwer (Beschlüsse der Kultusministerkonferenz)
- Lakatos, I. (1974). Falsifikation und die Methodologie wissenschaftlicher Forschungsprogramme. In I. Lakatos & A. Musgrave (Hrsg.), *Kritik und Erkenntnisfortschritt* (Its Abhandlungen, Bd. 4, S. 89–182). Braunschweig: Vieweg.
- Lawson, A. E. (2002). Sound and faulty arguments generated by preservice biology teachers when testing hypotheses involving unobservable entities. *Journal of Research in Science Teaching*, 39 (3), 237–252.
- Leisner-Bodenthin, A. (2006). Zur Entwicklung von Modellkompetenz im Physikunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 91–109.
- Liepmann, D., Beauducel, A., Brocke, B. & Nettelstroth, W. (2012). *Intelligenz-Struktur-Test – Screening*. Göttingen: Hogrefe.
- Linacre, J. M. (1994). Sample Size and Item Calibration [or Person Measure] Stability. *Rasch Measurement Transactions*, 7 (4), 328. Zugriff am 25.05.2015. Verfügbar unter <http://www.rasch.org/rmt/rmt74m.htm>
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In C. Allemann-Ghionda & E. Terhart (Hrsg.) *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern*. *Zeitschrift für Pädagogik*, 51, 47–70 [Themenheft]. Weinheim: Beltz.
- Majer, U. (1980). Experiment in den Naturwissenschaften. In J. Speck (Hrsg.), *Handbuch wissenschaftstheoretischer Begriffe* (UTB, Bd. 967, Bd. 1, S. 208–210). Göttingen: Vandenhoeck & Ruprecht.
- Mathesius, S. (in Vorbereitung). *Wissenschaftliches Denken im Hochschulstudium von Biologiestudierenden*. Dissertation, Freie Universität Berlin. Berlin.
- Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung. Ein Handbuch für Lehramtsstudenten und Doktoranden* (S. 177–184). Berlin: Springer.
- Mayer, J., Grube, C. & Möller, A. (2008). Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung. In R. Klee & U. Harms (Hrsg.), *Ausbildung und Professionalisierung von Lehrkräften. Internationale Tagung der Fachsektion Didaktik der Biologie im VBiO, Essen 2007* (Forschungen zur Fachdidaktik, Bd. 10, S. 63–79). Innsbruck: StudienVerl.
- Mayer, J. & Wellnitz, N. (2014). Die Entwicklung von Kompetenzstrukturmodellen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 19–29). Berlin, Heidelberg: Springer.
- Mayer, R. E. (2012). Problem Solving. In V. S. Ramachandran (Hrsg.), *Encyclopedia of Human Behavior* (S. 181–186). Oxford: Elsevier.

- Menne, A. (1980). *Einführung in die Methodologie: elementare allgemeine wissenschaftliche Denkmethode im Überblick*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50 (9), 741–749.
- Microsoft. (2010) Excel 2010 [Computer software].
- Mittelstrass, J. (1973). Das praktische Fundament der Wissenschaft und die Aufgabe der Philosophie. In F. Kambartel & J. Mittelstrass (Hrsg.), *Zum normativen Fundament der Wissenschaft* (Wissenschaftliche Paperbacks. Grundlagenforschung : Studien, Bd. 1, S. 1–69). Frankfurt (M.): Athenäum-Verlag.
- Monseur, C. & Adams, R. (2009). Plausible Values: How to deal with their Limitations. *Journal of applied Measurement*, 10 (3), 1–15.
- Moosbrugger, H. (2012). Item-Response-Theory (IRT). In H. Moosbrugger & A. Kelava (Hrsg.), *Springer-Lehrbuch* (S. 227–274). Berlin: Springer.
- Moosbrugger, H. & Kelava, A. (2012a). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In H. Moosbrugger & A. Kelava (Hrsg.), *Springer-Lehrbuch* (S. 7–26). Berlin: Springer.
- Moosbrugger, H. & Kelava, A. (2012b). *Testtheorie und Fragebogenkonstruktion* (Springer-Lehrbuch, 2. Aufl.). Berlin: Springer.
- Morrison, J. A. (2014). SCIENTISTS' PARTICIPATION IN TEACHER PROFESSIONAL DEVELOPMENT: THE IMPACT ON FOURTH TO EIGHTH GRADE TEACHERS' UNDERSTANDING AND IMPLEMENTATION OF INQUIRY SCIENCE. *International Journal of Science and Mathematics Education*, 12, 793–816.
- Neumann, K. (2014). Rasch-Analyse naturwissenschaftsbezogener Leistungstests. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 355–369). Berlin, Heidelberg: Springer.
- Nowak, K. H., Nehring, A., Tiemann, R. & Upmeyer zu Belzen, A. (2013). Assessing students' abilities in processes of scientific inquiry in biology using a paper-and-pencil test. *Journal of Biological Education*, 47 (3), 182–188.
- OECD. (2009). *PISA data analysis manual. SPSS* (2nd ed). Paris: OECD.
- Oh, P. S. & Oh, S. J. (2011). What Teachers of Science Need to Know about Models: An overview. *International Journal of Science Education*, 33 (8), 1109–1130.
- Oser, F., Curcio, G.-P. & Düggeli, A. (2007). Kompetenzmessung in der Lehrerbildung als Notwendigkeit - Fragen und Zugänge. *Beiträge zur Lehrerbildung*, 25 (1), 14–26.
- Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T. & Pöhlmann, C. (Hrsg.). (2013). *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I*. Münster: Waxmann.

- PhysLab. (2014). *Radioaktivität, die natürlichste Sache der Welt?*, Freie Universität Berlin. Verfügbar unter http://www.physik.fu-berlin.de/studium/schulkontakte/physlab/labor/img/Radioaktivitaet_2014_03_18_ohne_Anhang_ohne-Logarithmieren.pdf
- Pierce, C. A., Block, R. A. & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA Designs. *Educational and Psychological Measurement*, 64 (6), 916–924. Verfügbar unter <http://www.montana.edu/rblock/documents/papers/PierceBlockAguinas2004.pdf>
- Pietschmann, H. (1996). *Phänomenologie der Naturwissenschaft. Wissenschaftstheoretische und philosophische Probleme der Physik*. Berlin: Springer.
- Popper, K. R. (1974). Die Normalwissenschaft und ihre Gefahren. In I. Lakatos & A. Musgrave (Hrsg.), *Kritik und Erkenntnisfortschritt* (Its Abhandlungen, Bd. 4, S. 51–57). Braunschweig: Vieweg.
- Popper, K. R. (1984). *Logik der Forschung* (Einheit der Gesellschaftswissenschaften, Bd. 4, 8. Aufl.). Tübingen: J.C.B. Mohr (Paul Siebeck).
- Popper, K. R. (2010). *Lesebuch. Ausgewählte Texte zur Erkenntnistheorie, Philosophie der Naturwissenschaften, Metaphysik, Sozialphilosophie* (UTB, Bd. 2000, 2. Auflage). Stuttgart: UTB GmbH.
- Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2014). *Quantitative Methoden II. Einführung in die Statistik für Psychologen und Sozialwissenschaftler* (Springer-Lehrbuch, Bd. 2, 4. Aufl., 2 Bände). Berlin: Springer.
- Rehfeldt, D., Mühlenbruch, T. & Nordmeier, V. (2015). Fragebogen zu Praktikumskompetenzen (PraKo): Erforschung naturwissenschaftlicher Praktika. In S. Bernholt (Hrsg.), *Heterogenität und Diversität - Vielfalt der Voraussetzungen im naturwissenschaftlichen Unterricht* (S. 420–422). Kiel: IPN. Zugriff am 01.06.2015. Verfügbar unter http://www.gdcp.de/images/tb2015/TB2015_420_Rehfeldt.pdf
- Rehfeldt, D., Straube, P., Stiller, J., Kratschmar, J. & Nordmeier, V. (in Vorbereitung). Validierung PraQ (Arbeitstitel).
- Ricco, R. B. & Overton, W. F. (2012). Reasoning. In V. S. Ramachandran (Hrsg.), *Encyclopedia of Human Behavior* (S. 257–264). Oxford: Elsevier.
- Riese, J. (2009). *Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkräften* (Studien zum Physik- und Chemielernen, Bd. 97). Berlin: Logos.
- Riese, J. & Reinhold, P. (2012). Die professionelle Kompetenz angehender Physiklehrkräfte in verschiedenen Ausbildungsformen. *Zeitschrift für Erziehungswissenschaft*, 15 (1), 111–143.
- Riese, J. & Reinhold, P. (2014). Entwicklung eines Leistungstests für fachdidaktisches Wissen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 257–267). Berlin, Heidelberg: Springer.
- Rost, J. (2004). *Lehrbuch Testtheorie-Testkonstruktion* (2. Aufl.). Bern: Verlag Hans Huber.

- Roth, W.-M., McGinn, M. K. & Bowen, G. M. (1998). How prepared are preservice teachers to teach scientific inquiry? Levels of performance in scientific representation practices. *Journal of science teacher education*, 9 (1), 25–48.
- Sadler, P. M. (1998). Psychometric Models of Student Conceptions in Science: Reconciling Qualitative Studies and Distractor-Driven Assessment Instruments. *Journal of Research in Science Teaching*, 35 (3), 265–296.
- Schecker, H. & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 45–66.
- Schmiemann, P. & Lücken, M. (2014). Validität – Misst mein Test, was er soll? In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 107–118). Berlin, Heidelberg: Springer.
- Schnell, R., Hill, P. B. & Esser, E. (2011). *Methoden der empirischen Sozialforschung* (9., aktualis. Auflage). München: Oldenbourg, R.
- Schreiber, N., Theyßen, H. & Schecker, H. (2009). Experimentelle Kompetenz messen?! *Physik und Didaktik in Schule und Hochschule*, 8 (3), 92–101.
- Schroeders, U., Penk, C., Jansen, M. & Pant, H. A. (2013). Geschlechtsbezogene Disparitäten. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 249–274). Münster: Waxmann.
- Schwartz, R. S., Lederman, N. G. & Crawford, B. A. (2004). Developing Views of Nature of Science in an Authentic Context: An Explicit Approach to Bridging the Gap Between Nature of Science and Scientific Inquiry. *Science Teacher Education*, 600–645.
- Shamos, M. H. (1959). *Great Experiments in Physics*. New York: Holt, Rinehart and Winston.
- Shavelson, R. J. (2013). An approach to testing & modelling competence. In S. Blömeke, O. Zlatkin-Troitschanskaia, C. Kuhn & J. Fege (Hrsg.), *Modeling and measuring competencies in higher education. Tasks and challenges* (Professional and VET learning, Bd. 1, S. 29–43). Rotterdam: Sense Publishers.
- Shulman, L. (1986a). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15, 4–14.
- Shulman, L. (1987b). Knowledge and teaching: Foundations of the new reform. *Harvard educational review*, 57 (1), 1–23.
- Sippel, S. (2009). Zur Relevanz von Assessment-Feedback in der Hochschullehre. *Zeitschrift für Hochschulentwicklung*, 4 (1), 1–22.
- Stecher, B. M. & Klein, S. P. (1997). The Cost of Science Performance Assessments in Large-Scale Testing Programs. *Educational Evaluation and Policy*, 19, 1–14.

- Stiller, J. (in Vorbereitung). *Analyse von Struktur und Entwicklung der Kompetenzen Studierender des Lehramts im Bereich Scientific Reasoning. Eine Querschnittsstudie im Fach Chemie*. Dissertation, Humboldt Universität zu Berlin. Berlin.
- Stiller, J., Straube, P., Hartmann, S., Nordmeier, V. & Tiemann, R. (2015). Erkenntnisgewinnungskompetenz Chemie und Physik-Lehramtsstudierender: Untersuchungen zu Domänenspezifität. In J. Stiller & C. Laschke (Hrsg.), *Berlin-Brandenburger Beiträge zur Bildungsforschung 2015. Herausforderungen, Befunde und Perspektiven interdisziplinärer Bildungsforschung* (S. 179–202). Frankfurt a.M.: Peter Lang.
- Stöckler, M. (1995). Modell, Idealisierung und Realität. Theoretische Modelle im Lichte der Wissenschaftstheorie. *Praxis der Naturwissenschaften - Physik*, 44 (1), 16–21.
- Straube, P. & Nordmeier, V. (2014). Ko - WADiS - Erkenntnisgewinnungskompetenz von Lehramtsstudierenden. In V. Nordmeier & H. Grötzebauch (Hrsg.), *PhyDid B, Didaktik der Physik, Beiträge zur DPG-Frühjahrstagung*. Berlin: DPG. Zugriff am 13.04.2015. Verfügbar unter <http://www.phydid.de/index.php/phydid-b/article/download/583/643>
- Strelow, S., Straube, P. & Nordmeier, V. (2014). Falsifikation als naturwissenschaftliche Methode der Erkenntnisgewinnung. *Praxis der Naturwissenschaften - Physik in der Schule*, 63 (8), 22–26.
- Sun, X. (2012). *Conquest Notes 7 - Using Mantel-Haenszel Statistics*. Zugriff am 05.12.2014. Verfügbar unter <http://www.acer.edu.au/files/Conquest-Notes-7-UsingMantel-HaenszelStatistics.pdf>
- Taylor, J. A. & Dana, T. M. (2003). Secondary school physics teachers' conceptions of scientific evidence: An exploratory case study. *Journal of Research in Science Teaching*, 40 (8), 721–736.
- Tepner, O., Borowski, A., Dollny, S., Fischer, H. E., Jüttner, M., Kirschner, S. et al. (2012). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 7–28.
- Upmeyer zu Belzen, A. & Krüger, D. (2010). Modellkompetenz im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 41–57.
- van Driel, J., Berry, A. & Meirink, J. (2014). Research on Science Teacher Knowledge. In S. K. Abell & N. G. Lederman (Hrsg.), *Handbook of research on science education* (2. Aufl., S. 848–870). London: Routledge.
- Van Driel, Jan H. & Verloop, N. (1999). Teachers' knowledge of models and modelling in science. *International Journal of Science Education*, 21 (11), 1141–1153.
- Vogelsang, C. (2014). *Validierung eines Instruments zur Erfassung der professionellen Handlungskompetenz von (angehenden) Physiklehrkräften. Zusammenhangsanalysen zwischen Lehrerkompetenz und Lehrerperformance* (Studien zum Physik- und Chemielernen, Bd. 174). Berlin: Logos.
- Walpulski, M., Kauertz, A., Kampa, N., Fischer, H. E., Mayer, J., Sumfleth, E. et al. (2010). ESNaS - Evaluation der Standards für die Naturwissenschaften in der Sekundarstufe I. In A. Gehrmann (Hrsg.), *Bildungsstandards und Kompetenzmodelle. Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht* (S. 171–184). Bad Heilbrunn: Klinkhardt.

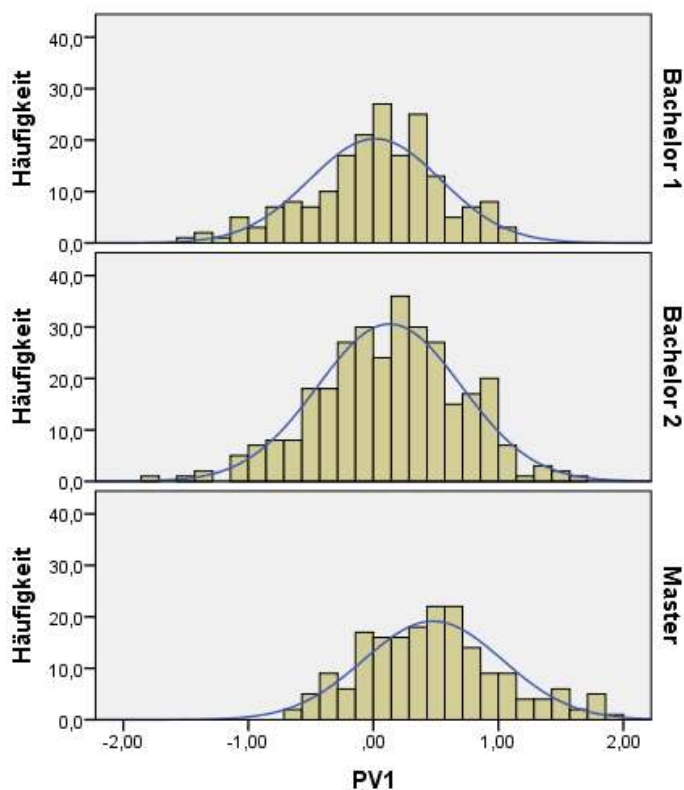
- Weinert, F. E. (2001a). Concept of Competence - a conceptual clarification. In D. S. Rychen & L. H. Salganik (Hrsg.), *Defining and selecting key competencies* (S. 45–65). Seattle: Hogrefe&Huber.
- Weinert, F. E. (2001b). *Leistungsmessungen in Schulen* (Beltz-Pädagogik). Weinheim [u.a.]: Beltz.
- Wellnitz, N. (2012). *Kompetenzstruktur und -niveaus von Methoden naturwissenschaftlicher Erkenntnisgewinnung* (Biologie lernen und lehren, Bd. 2). Berlin: Logos.
- Wellnitz, N., Fischer, H. E., Kauertz, A., Mayer, J., Neumann, I., Pant, H. A. et al. (2012). Evaluation der Bildungsstandards - eine fächerübergreifende Testkonzeption für den Kompetenzbereich Erkenntnisgewinnung. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 261–291.
- Wiltsche, H. A. (2013). *Einführung in die Wissenschaftstheorie* (UTB: Philosophie, Bd. 3936). Göttingen: Vandenhoeck et Ruprecht.
- Windschitl, M. (2003). Inquiry projects in science teacher education: What can investigative experiences reveal about teacher thinking and eventual classroom practice? *Science Teacher Education*, 87 (1), 112–143.
- Wissenschaftsrat. (2008). *Empfehlungen zur Qualitätsverbesserung von Lehre und Studium* (Drs. 8639-08). Zugriff am 03.04.2016. Verfügbar unter <http://www.wissenschaftsrat.de/download/archiv/8639-08.pdf>
- Woitkowski, D. (2015). *Fachliches Wissen Physik in der Hochschulausbildung*. Dissertation, Universität Paderborn. Paderborn.
- Woitkowski, D., Riese, J. & Reinhold, P. (2011). Modellierung fachwissenschaftlicher Kompetenz angehender Physiklehrkräfte. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 289–313.
- Woitkowski, D., Riese, J. & Reinhold, P. (2014). Prospective physicists' and physics teachers' content knowledge. First Results of a Germany-Wide Study. In C. P. Constantinou, N. Papadouris & A. Hadjigeorgiou (Hrsg.), *E-Book Proceedings of the ESERA 2013 Conference: Science Education Research For Evidence-based Teaching and Coherence in Learning* (S. 28–37). Nicosia: European Science Education Research Association. Zugriff am 22.04.2015. Verfügbar unter http://www.esera.org/media/eBook_2013/Strand%2013/ESERA_eBook_Part_13.pdf
- Worral, J. (1980). Wie die Methodologie der wissenschaftlichen Forschungsprogramme die Poppersche Methodologie verbessert. In G. Radnitzky, H. Albert & G. Andersson (Hrsg.), *Fortschritt und Rationalität der Wissenschaft* (S. 51–73). Tübingen: Mohr.
- Wright, B. D. & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8 (3), 370. Zugriff am 12.05.2015. Verfügbar unter <http://www.rasch.org/rmt/rmt83b.htm>
- Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31 (2-3), 114–128.
- Wu, M. & Adams, R. (2007). *Applying the Rasch Model to psycho-social measurement. a practical approach*, Educational Measurement Solutions. Zugriff am 12.05.2015. Verfügbar unter http://www.edmeasurement.com.au/_publications/RaschMeasurement_Complete.pdf

- Wu, M., Adams, R., Wilson, M. & Haldane, S. (2007). *Acer Conquest 2.0. Generalised Item response modelling software*. Camberwell: Acerpress.
- Yip, D. Y. (2001). Assessing and Developing the concept of Assumptions in Science Teachers. *Journal of Science Education and Technology*, 10 (2), 173–179.
- Zimmerman, C. (2000). The Development of Scientific Reasoning Skills. *Developmental Review*, 20 (1), 99–149.
- Zlatkin-Troitschanskaia, O. & Kuhn, C. (2010). Messung akademisch vermittelter Fertigkeiten und Kenntnisse von Studierenden bzw. Hochschulabsolventen. Analyse zum Forschungsstand. *Arbeitspapiere WP* (56), 1–33.
- Zwick, R., Thayer, D. T. & Lewis, C. (1999). An Empirical Bayes Approach to Mantel-Haenszel DIF Analysis. *Journal of Educational Measurement*, 36 (1), 1–28.

Das Literaturverzeichnis wurde mit Hilfe der Software Citavi 4 erstellt.

11. Anlagen

Test auf Normalverteilung Fragestellungen 2-5



$D(187)=0,061$, n.s.

$S(187)=0,982$, **$p=0,014$**

$v=-0,414$ (0,178) $Z_S = 2,32$

$w=0,046$ (0,354) $Z_K = 0,13$

$D(308)=0,041$, n.s.

$S(308)=0,995$, n.s.

$v=-0,208$ (0,139) $Z_S = 1,50$

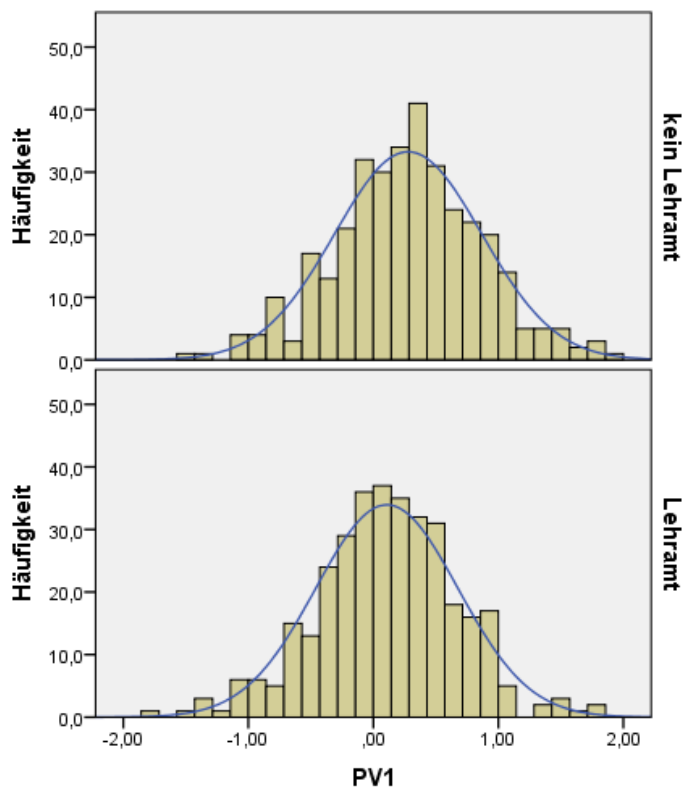
$w=0,149$ (0,277) $Z_K = 0,54$

$D(187)=0,065$, n.s.

$S(187)=0,982$, **$p=0,016$**

$v=0,366$ (0,178) $Z_S = 2,01$

$w=-0,168$ (0,354) $Z_K = 0,47$



$D(343)=0,040$, n.s.

$S(343)=0,996$, n.s.

$v= 0,021$ (0,132) $Z_S = 0,16$

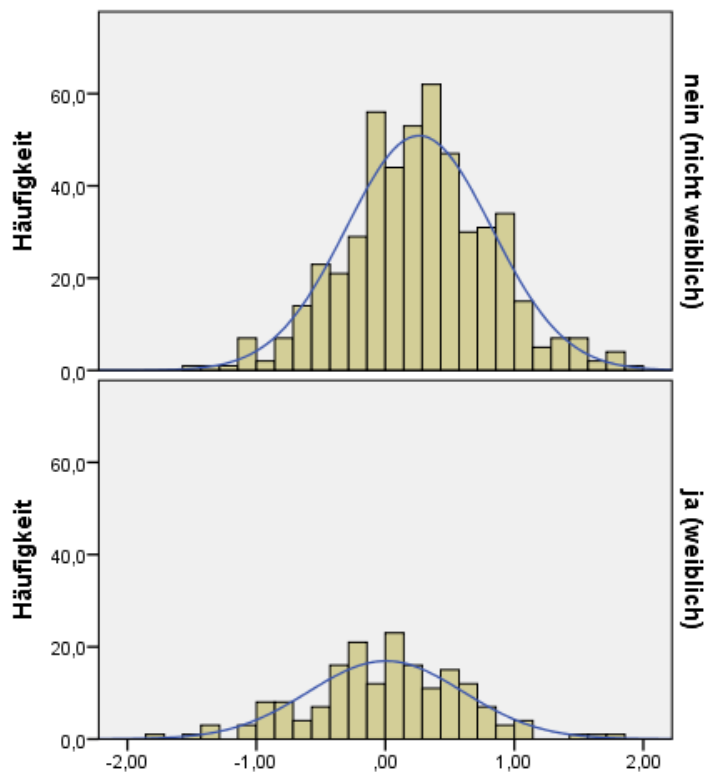
$w=0,161$ (0,263) $Z_K = 0,61$

$D(339)=0,044$, n.s.

$S(339)=0,995$, n.s.

$v=-0,107$ (0,132) $Z_S = 0,81$

$w=0,495$ (0,264) $Z_K = 1,88$

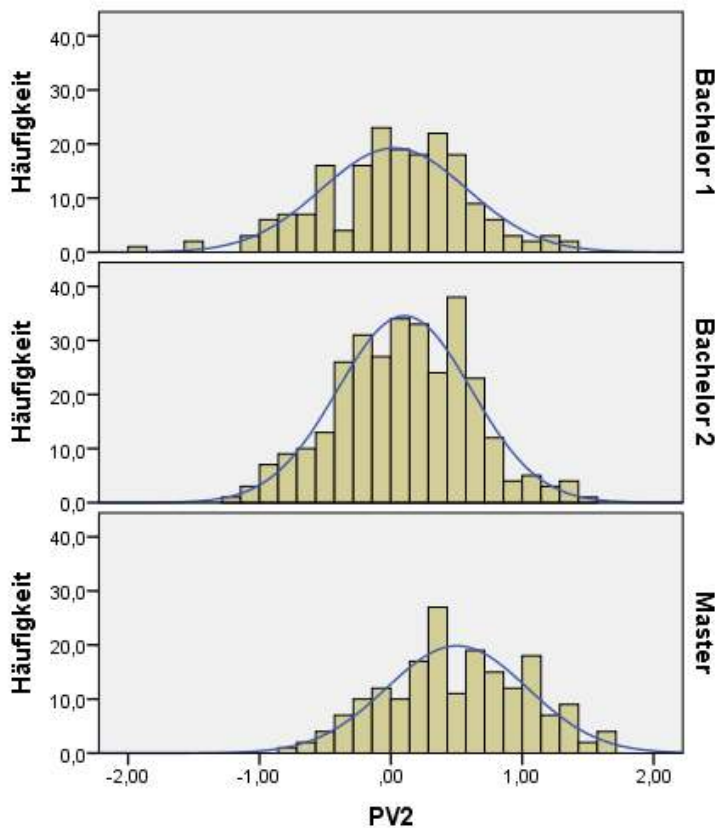


$D(504)=0,039$, n.s.
 $S(504)=0,997$, n.s.
 $v= 0,033$ (0,109) $Z_S = 0,30$
 $w=0,187$ (0,217) $Z_K = 0,86$

$D(178)=0,056$, n.s.
 $S(178)=0,992$, n.s.
 $v=-0,111$ (0,182) $Z_K = 0,61$
 $w=0,488$ (0,362) $Z_K = 1,35$

PV2:

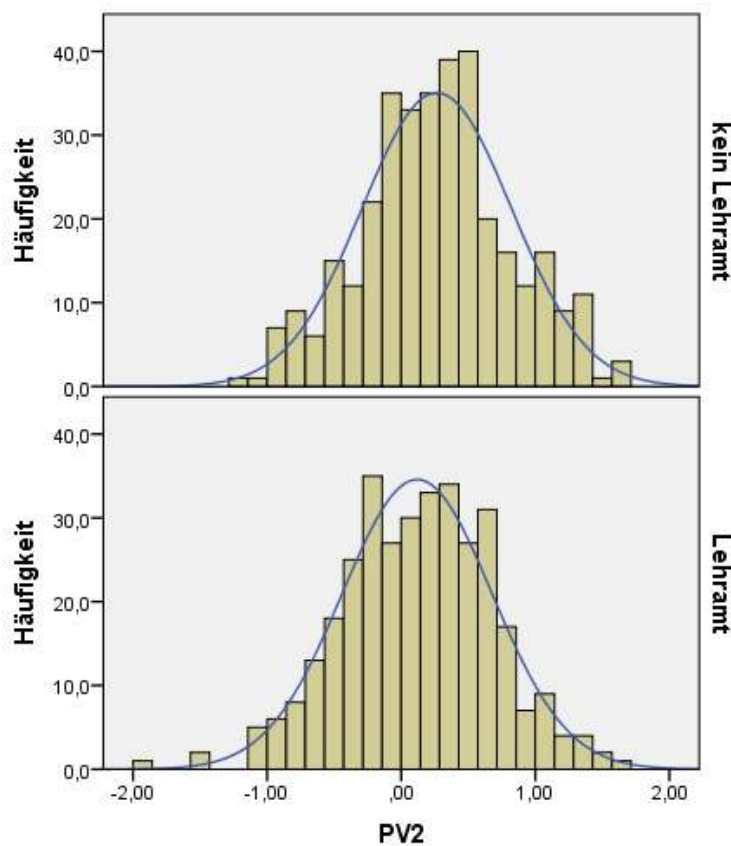
PV1



$D(187)=0,058$, n.s.
 $S(187)=0,988$, n.s.
 $v=-0,342$ (0,178) $Z_S = 1,92$
 $w=0,451$ (0,354) $Z_K = 1,27$

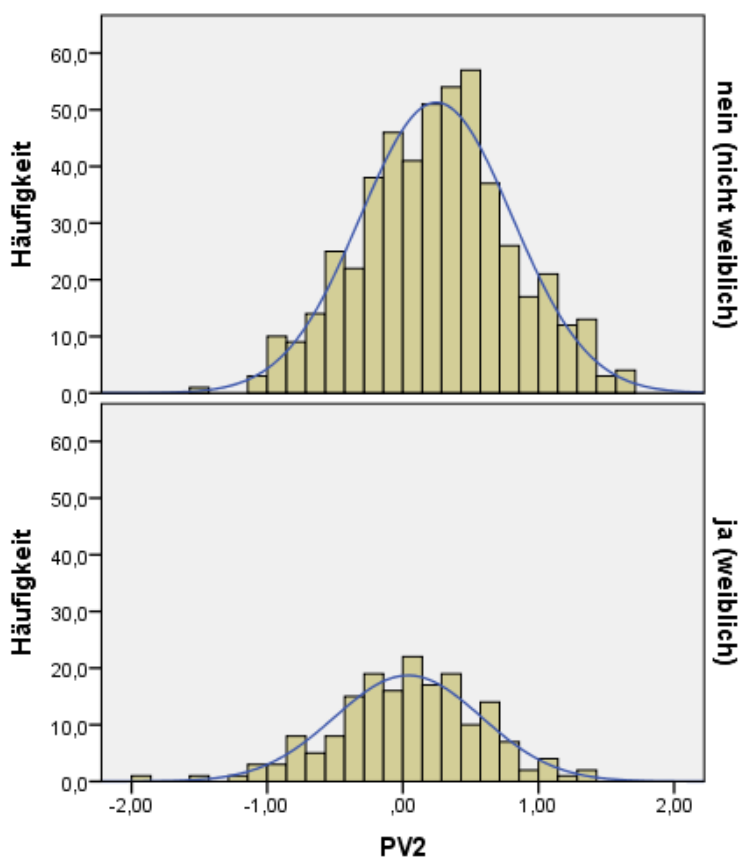
$D(308)=0,041$, n.s.
 $S(308)=0,995$, n.s.
 $v=-0,008$ (0,139) $Z_S = 0,06$
 $w=0,142$ (0,277) $Z_K = 0,23$

$D(187)=0,044$, n.s.
 $S(187)=0,992$, n.s.
 $v=-0,034$ (0,178) $Z_S = 0,19$
 $w=-0,544$ (0,354) $Z_K = 1,54$



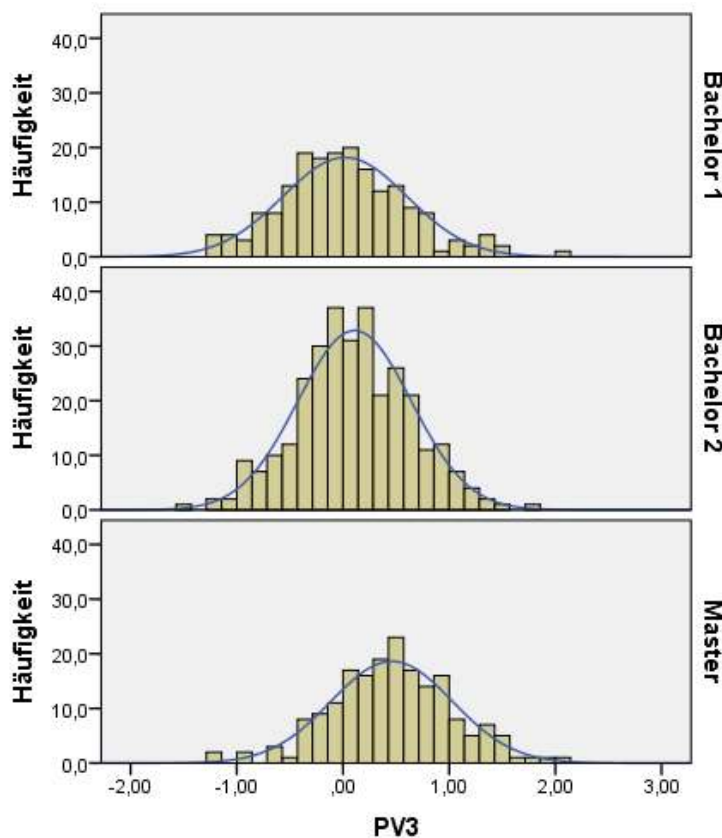
$D(343)=0,038$, n.s.
 $S(343)=0,995$, n.s.
 $v=-0,021$ (0,132) $Z_S = 0,16$
 $w=-0,146$ (0,263) $Z_K = 0,55$

$D(339)=0,34$, n.s.
 $S(339)=0,997$, n.s.
 $v=-0,132$ (0,132) $Z_S = 1,00$
 $w=0,179$ (0,264) $Z_K = 0,68$



$D(504)=0,031$, n.s.
 $S(504)=0,996$, n.s.
 $v=0,005$ (0,109) $Z_S = 0,05$
 $w=-0,145$ (0,217) $Z_K = 0,67$

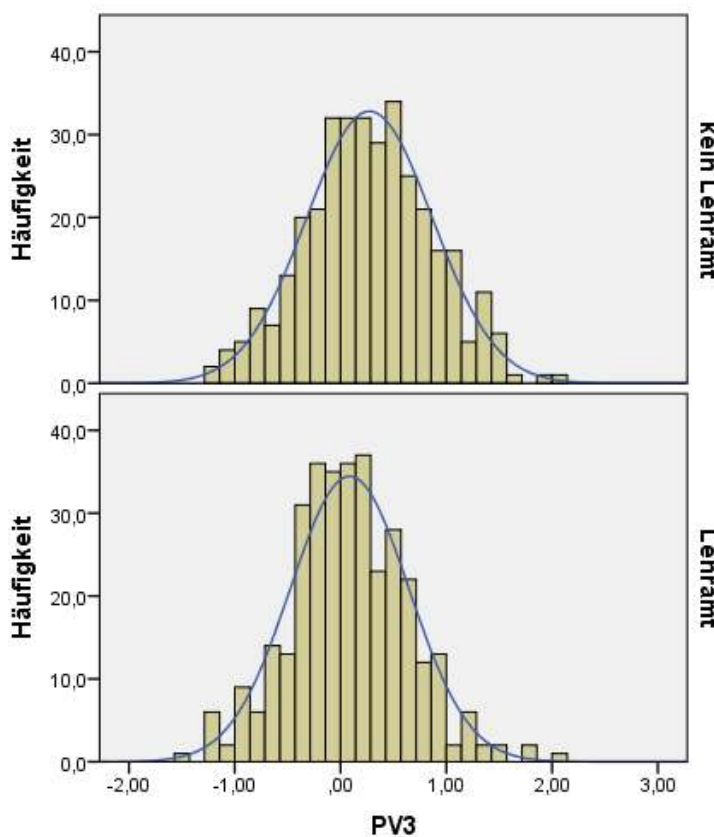
$D(178)=0,035$, n.s.
 $S(178)=0,994$, n.s.
 $v=-0,262$ (0,182) $Z_S = 1,44$
 $w=0,438$ (0,362) $Z_K = 1,21$



Bachelor 1
 $D(187)=0,042$, n.s.
 $S(187)=0,988$, n.s.
 $\nu=0,388$ (0,178) $Z_S = 2,18$
 $\mu=-0,385$ (0,354) $Z_K = 1,08$

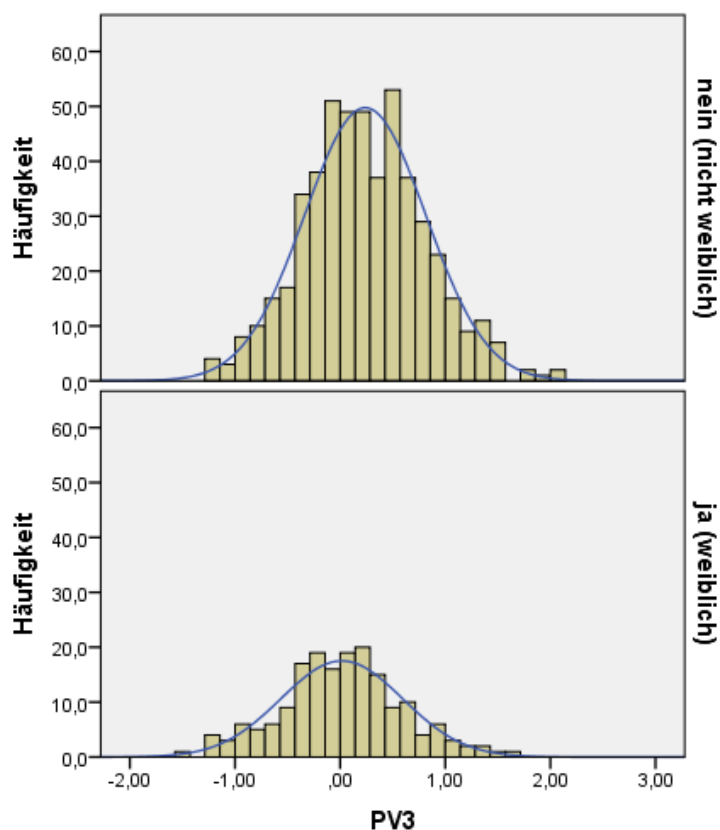
Bachelor 2
 $D(308)=0,032$, n.s.
 $S(308)=0,998$, n.s.
 $\nu=0,032$ (0,139) $Z_S = 0,23$
 $\mu=0,051$ (0,277) $Z_K = 0,18$

Master
 $D(187)=0,035$, n.s.
 $S(187)=0,994$, n.s.
 $\nu=-0,26$ (0,178) $Z_S = 1,46$
 $\mu=0,418$ (0,354) $Z_K = 1,18$



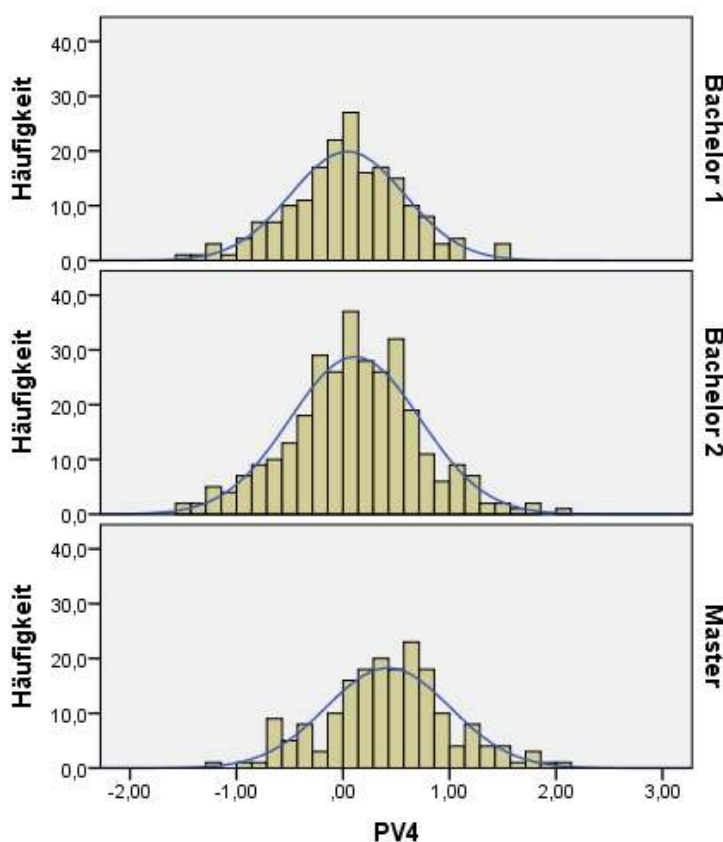
kein Lehramt
 $D(343)=0,026$, n.s.
 $S(343)=0,997$, n.s.
 $\nu=0,037$ (0,132) $Z_S = 0,28$
 $\mu=-0,129$ (0,263) $Z_K = 0,49$

Lehramt
 $D(339)=0,038$, n.s.
 $S(339)=0,993$, n.s.
 $\nu=0,216$ (0,132) $Z_S = 1,64$
 $\mu=0,528$ (0,264) $Z_K = 2,00$



D(504)=0,030, n.s.
 S(504)=0,996, n.s.
 $\nu=0,169$ (0,109) $Z_S = 1,55$
 $\eta=0,090$ (0,217) $Z_K = 0,41$

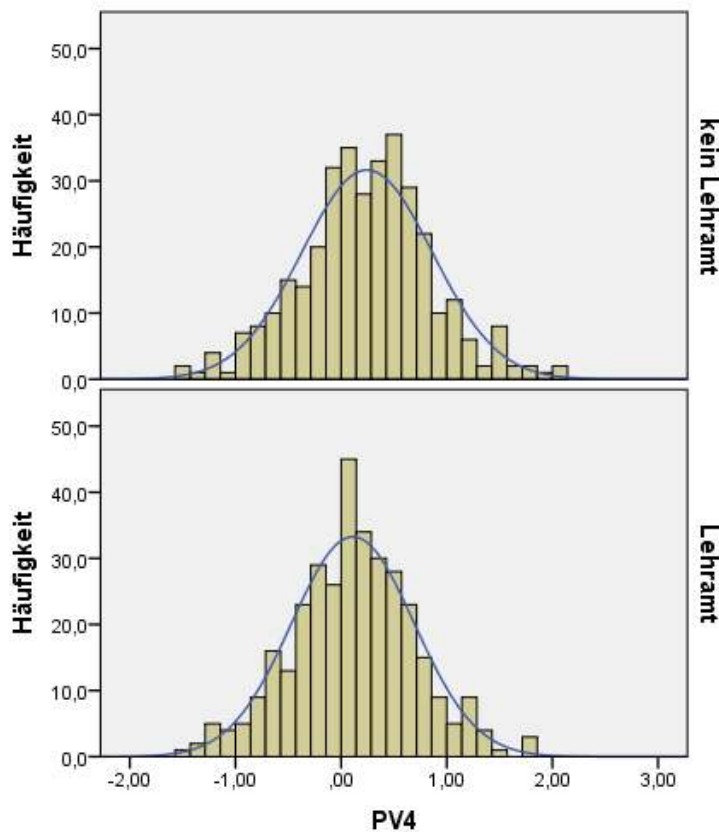
D(178)=0,040, n.s.
 S(178)=0,995, n.s.
 $\nu=0,114$ (0,182) $Z_S = 0,63$
 $\eta=0,132$ (0,362) $Z_K = 0,36$



D(187)=0,059, n.s.
 S(187)=0,993, n.s.
 $\nu=-0,090$ (0,178) $Z_S = 0,51$
 $\eta=0,331$ (0,354) $Z_K = 0,94$

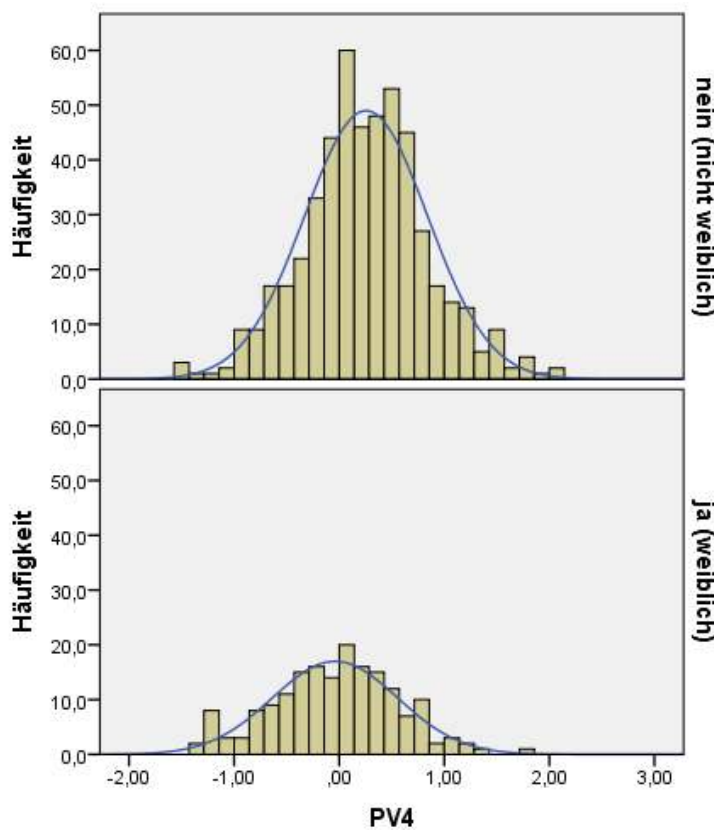
D(308)=0,040, n.s.
 S(308)=0,995, n.s.
 $\nu=0,025$ (0,139) $Z_K = 0,18$
 $\eta=0,343$ (0,277) $Z_K = 1,24$

D(187)=0,058, n.s.
 S(187)=0,991, n.s.
 $\nu=0,074$ (0,178) $Z_S = 0,42$
 $\eta=0,221$ (0,354) $Z_K = 0,62$



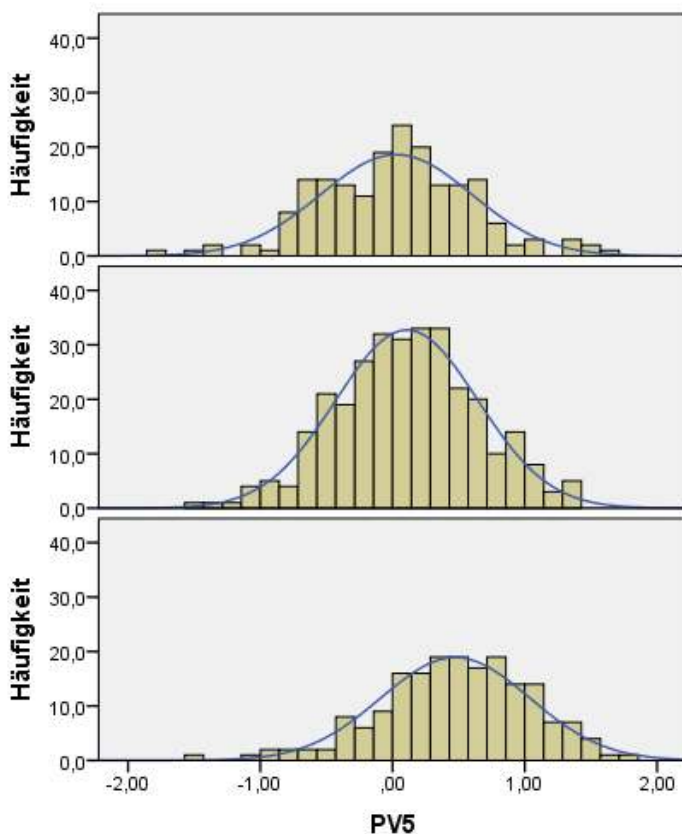
$D(343)=0,040$, n.s.
 $S(343)=0,994$, n.s.
 $\nu=0,028$ (0,132) $Z_S = 0,21$
 $\mu=0,321$ (0,264) $Z_K = 1,22$

$D(339)=0,033$, n.s.
 $S(339)=0,996$, n.s.
 $\nu=0,014$ (0,132) $Z_S = 0,11$
 $\mu=0,301$ (0,264) $Z_K = 1,14$



$D(504)=0,040$, n.s.
 $S(504)=0,992$, n.s.
 $\nu=0,102$ (0,109) $Z_S = 0,94$
 $\mu=0,425$ (0,217) $Z_K = 1,96$

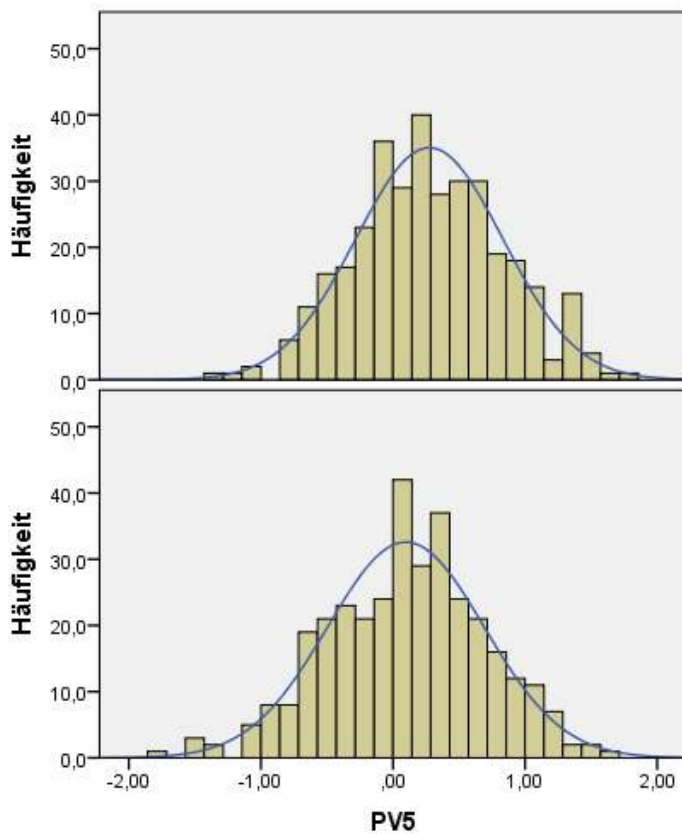
$D(178)=0,040$, n.s.
 $S(178)=0,992$, n.s.
 $\nu=-0,029$ (0,182) $Z_S = 0,16$
 $\mu=-0,049$ (0,362) $Z_K = 0,14$



D(187)=0,045, n.s.
 S(187)=0,992, n.s.
 $\nu = -0,009$ (0,178) $Z_S = 0,05$
 $w = 0,470$ (0,354) $Z_K = 1,38$

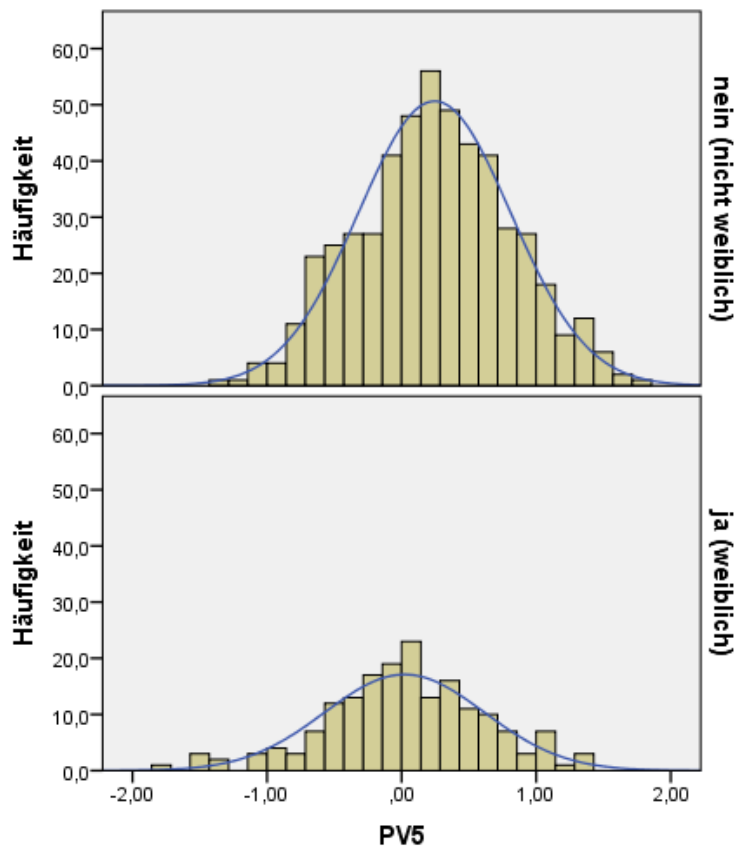
D(308)=0,023, n.s.
 S(308)=0,996, n.s.
 $\nu = -0,046$ (0,139) $Z_S = 0,33$
 $w = -0,145$ (0,277) $Z_K = 0,52$

D(187)=0,044, n.s.
 S(187)=0,989, n.s.
 $\nu = 0,401$ (0,178) $Z_S = 2,25$
 $w = 0,306$ (0,354) $Z_K = 0,86$



D(343)=0,036, n.s.
 S(343)=0,996, n.s.
 $\nu = 0,082$ (0,132) $Z_S = 0,62$
 $w = -0,215$ (0,263) $Z_K = 0,82$

D(339)=0,039, n.s.
 S(339)=0,996, n.s.
 $\nu = -0,172$ (0,132) $Z_K = 1,30$
 $w = -0,055$ (0,264) $Z_K = 0,21$



$D(504)=0,025$, n.s.
 $S(504)=0,997$, n.s.
 $v= 0,020$ (0,109) $Z_S = 0,18$
 $w=-0,326$ (0,217) $Z_K = 1,50$

$D(178)=0,042$, n.s.
 $S(178)=0,989$, n.s.
 $v=-0,260$ (0,182) $Z_S = 1,43$
 $w=0,325$ (0,362) $Z_K = 0,90$

Deskriptive Statistiken zur ANOVA (Fragestellung 2-5)

Abhängige Variable: PV1

StuPhase			Mittelwert	Standard- abweichung	H
Bachelor 1	nein	nein (nicht weiblich)	,1649	,49052	108
		ja (weiblich)	-,0704	,48948	27
		Gesamtsumme	,1179	,49754	135
	ja	nein (nicht weiblich)	-,2129	,44648	38
		ja (weiblich)	-,3664	,64924	14
		Gesamtsumme	-,2542	,50675	52
	Gesamt- summe	nein (nicht weiblich)	,0666	,50606	146
		ja (weiblich)	-,1715	,55940	41
		Gesamtsumme	,0144	,52601	187
Bachelor 2	nein	nein (nicht weiblich)	,2544	,58848	109
		ja (weiblich)	,0696	,58564	26
		Gesamtsumme	,2188	,59030	135
	ja	nein (nicht weiblich)	,1636	,48832	107
		ja (weiblich)	-,1091	,60965	66
		Gesamtsumme	,0595	,55232	173
	Gesamt- summe	nein (nicht weiblich)	,2094	,54185	216
		ja (weiblich)	-,0586	,60519	92
		Gesamtsumme	,1294	,57382	308
Master	nein	nein (nicht weiblich)	,6972	,56208	68
		ja (weiblich)	,5000	,47308	5
		Gesamtsumme	,6837	,55582	73
	ja	nein (nicht weiblich)	,3899	,50439	74
		ja (weiblich)	,2633	,53744	40
		Gesamtsumme	,3454	,51742	114
	Gesamt- summe	nein (nicht weiblich)	,5370	,55279	142
		ja (weiblich)	,2896	,53106	45
		Gesamtsumme	,4775	,55643	187
Gesamtsumme	nein	nein (nicht weiblich)	,3261	,58456	285
		ja (weiblich)	,0416	,54786	58
		Gesamtsumme	,2780	,58754	343
	ja	nein (nicht weiblich)	,1747	,52625	219
		ja (weiblich)	-,0150	,62385	120
		Gesamtsumme	,1076	,56911	339
	Gesamt- summe	nein (nicht weiblich)	,2603	,56445	504
		ja (weiblich)	,0034	,59918	178
		Gesamtsumme	,1933	,58429	682

Abhängige Variable: PV2

StuPhase			Mittelwert	Standard- abweichung	H
Bachelor 1	nein	nein (nicht weiblich)	,1364	,50472	108
		ja (weiblich)	,0470	,48059	27
		Gesamtsumme	,1185	,49951	135
	ja	nein (nicht weiblich)	-,1558	,54370	38
		ja (weiblich)	-,3707	,78305	14
		Gesamtsumme	-,2137	,61646	52
	Gesamtsumme	nein (nicht weiblich)	,0603	,52911	146
		ja (weiblich)	-,0956	,62420	41
		Gesamtsumme	,0261	,55338	187
Bachelor 2	nein	nein (nicht weiblich)	,2336	,51399	109
		ja (weiblich)	-,0608	,49809	26
		Gesamtsumme	,1769	,52229	135
	ja	nein (nicht weiblich)	,0989	,50077	107
		ja (weiblich)	-,0383	,46436	66
		Gesamtsumme	,0465	,49041	173
	Gesamtsumme	nein (nicht weiblich)	,1669	,51078	216
		ja (weiblich)	-,0447	,47147	92
		Gesamtsumme	,1037	,50794	308
Master	nein	nein (nicht weiblich)	,6926	,52094	68
		ja (weiblich)	,7660	,35260	5
		Gesamtsumme	,6977	,50969	73
	ja	nein (nicht weiblich)	,4353	,52720	74
		ja (weiblich)	,2758	,49229	40
		Gesamtsumme	,3793	,51870	114
	Gesamtsumme	nein (nicht weiblich)	,5585	,53805	142
		ja (weiblich)	,3302	,50039	45
		Gesamtsumme	,5036	,53690	187
Gesamtsumme	nein	nein (nicht weiblich)	,3063	,55607	285
		ja (weiblich)	,0607	,52283	58
		Gesamtsumme	,2648	,55752	343
	ja	nein (nicht weiblich)	,1684	,55682	219
		ja (weiblich)	,0276	,55314	120
		Gesamtsumme	,1185	,55878	339
	Gesamtsumme	nein (nicht weiblich)	,2463	,56004	504
		ja (weiblich)	,0384	,54219	178
		Gesamtsumme	,1921	,56252	682

Abhängige Variable: PV3

StuPhase			Mittelwert	Standard- abweichung	H
Bachelor 1	nein	nein (nicht weiblich)	,1665	,56358	108
		ja (weiblich)	,0326	,63046	27
		Gesamtsumme	,1397	,57761	135
	ja	nein (nicht weiblich)	-,3063	,44033	38
		ja (weiblich)	-,2736	,60668	14
		Gesamtsumme	-,2975	,48446	52
	Gesamt- summe	nein (nicht weiblich)	,0434	,57200	146
		ja (weiblich)	-,0720	,63212	41
		Gesamtsumme	,0181	,58591	187
Bachelor 2	nein	nein (nicht weiblich)	,2281	,54045	109
		ja (weiblich)	,0781	,53016	26
		Gesamtsumme	,1992	,53980	135
	ja	nein (nicht weiblich)	,1401	,52861	107
		ja (weiblich)	-,1265	,46737	66
		Gesamtsumme	,0384	,52118	173
	Gesamt- summe	nein (nicht weiblich)	,1845	,53520	216
		ja (weiblich)	-,0687	,49175	92
		Gesamtsumme	,1089	,53455	308
Master	nein	nein (nicht weiblich)	,6572	,56160	68
		ja (weiblich)	,6740	,78005	5
		Gesamtsumme	,6584	,57212	73
	ja	nein (nicht weiblich)	,3943	,49027	74
		ja (weiblich)	,2165	,60335	40
		Gesamtsumme	,3319	,53683	114
	Gesamt- summe	nein (nicht weiblich)	,5202	,54007	142
		ja (weiblich)	,2673	,63176	45
		Gesamtsumme	,4594	,57208	187
Gesamtsumme	nein	nein (nicht weiblich)	,3071	,58687	285
		ja (weiblich)	,1083	,61521	58
		Gesamtsumme	,2735	,59553	343
	ja	nein (nicht weiblich)	,1485	,55306	219
		ja (weiblich)	-,0293	,55803	120
		Gesamtsumme	,0856	,56051	339
	Gesamt- summe	nein (nicht weiblich)	,2382	,57725	504
		ja (weiblich)	,0155	,57916	178
		Gesamtsumme	,1801	,58556	682

Modellierung und Erfassung von Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung

Abhängige Variable: PV4

StuPhase			Mittelwert	Standard- abweichung	H
Bachelor 1	nein	nein (nicht weiblich)	,1662	,52821	108
		ja (weiblich)	-,0719	,54880	27
		Gesamtsumme	,1186	,53885	135
	ja	nein (nicht weiblich)	,0042	,40605	38
		ja (weiblich)	-,5936	,38302	14
		Gesamtsumme	-,1567	,47822	52
	Gesamt- summe	nein (nicht weiblich)	,1240	,50304	146
		ja (weiblich)	-,2500	,55334	41
		Gesamtsumme	,0420	,53590	187
Bachelor 2	nein	nein (nicht weiblich)	,2365	,65602	109
		ja (weiblich)	-,0431	,58892	26
		Gesamtsumme	,1827	,65101	135
	ja	nein (nicht weiblich)	,1527	,56424	107
		ja (weiblich)	-,0939	,56009	66
		Gesamtsumme	,0586	,57375	173
	Gesamt- summe	nein (nicht weiblich)	,1950	,61229	216
		ja (weiblich)	-,0796	,56558	92
		Gesamtsumme	,1130	,61092	308
Master	nein	nein (nicht weiblich)	,6062	,57820	68
		ja (weiblich)	,4340	,38371	5
		Gesamtsumme	,5944	,56674	73
	ja	nein (nicht weiblich)	,3697	,52991	74
		ja (weiblich)	,2100	,63244	40
		Gesamtsumme	,3137	,57037	114
	Gesamt- summe	nein (nicht weiblich)	,4830	,56417	142
		ja (weiblich)	,2349	,61072	45
		Gesamtsumme	,4233	,58380	187
Gesamtsumme	nein	nein (nicht weiblich)	,2981	,61518	285
		ja (weiblich)	-,0153	,56516	58
		Gesamtsumme	,2451	,61753	343
	ja	nein (nicht weiblich)	,2003	,54256	219
		ja (weiblich)	-,0509	,61402	120
		Gesamtsumme	,1114	,58058	339
	Gesamt- summe	nein (nicht weiblich)	,2556	,58618	504
		ja (weiblich)	-,0393	,59717	178
		Gesamtsumme	,1786	,60273	682

Abhängige Variable: PV5

StuPhase			Mittelwert	Standard- abweichung	H
Bachelor 1	nein	nein (nicht weiblich)	,1347	,52692	108
		ja (weiblich)	,1522	,56831	27
		Gesamtsumme	,1382	,53331	135
	ja	nein (nicht weiblich)	-,1487	,50631	38
		ja (weiblich)	-,5564	,66792	14
		Gesamtsumme	-,2585	,57710	52
	Gesamt- summe	nein (nicht weiblich)	,0610	,53466	146
		ja (weiblich)	-,0898	,68605	41
		Gesamtsumme	,0279	,57270	187
Bachelor 2	nein	nein (nicht weiblich)	,2581	,52117	109
		ja (weiblich)	,0327	,43601	26
		Gesamtsumme	,2147	,51219	135
	ja	nein (nicht weiblich)	,1022	,53738	107
		ja (weiblich)	-,1029	,52716	66
		Gesamtsumme	,0240	,54127	173
	Gesamt- summe	nein (nicht weiblich)	,1809	,53377	216
		ja (weiblich)	-,0646	,50448	92
		Gesamtsumme	,1076	,53631	308
Master	nein	nein (nicht weiblich)	,6540	,52765	68
		ja (weiblich)	,4020	,62799	5
		Gesamtsumme	,6367	,53394	73
	ja	nein (nicht weiblich)	,4158	,53709	74
		ja (weiblich)	,2850	,59796	40
		Gesamtsumme	,3699	,56008	114
	Gesamt- summe	nein (nicht weiblich)	,5299	,54396	142
		ja (weiblich)	,2980	,59511	45
		Gesamtsumme	,4741	,56388	187
Gesamtsumme	nein	nein (nicht weiblich)	,3058	,56090	285
		ja (weiblich)	,1202	,51896	58
		Gesamtsumme	,2744	,55767	343
	ja	nein (nicht weiblich)	,1647	,56653	219
		ja (weiblich)	-,0265	,62177	120
		Gesamtsumme	,0970	,59287	339
	Gesamt- summe	nein (nicht weiblich)	,2445	,56713	504
		ja (weiblich)	,0213	,59279	178
		Gesamtsumme	,1862	,58183	682

Ergebnisse der ANOVA Fragestellungen 1-5

Abhängige Variable:		PV1		
Quelle	Typ III Quadratsumme	df	Quadratischer Mittelwert	F
Korrigiertes Modell	39,946	6	6,658	23,340
Konstanter Term	8,908	1	8,908	31,230
StuPhase	28,557	2	14,278	50,056
lehramt	8,867	1	8,867	31,084
female	5,385	1	5,385	18,877
StuPhase * lehramt	1,627	2	,813	2,851
Fehler	192,542	675	,285	
Gesamtsumme	257,966	682		
Korrigierter Gesamtwert	232,487	681		

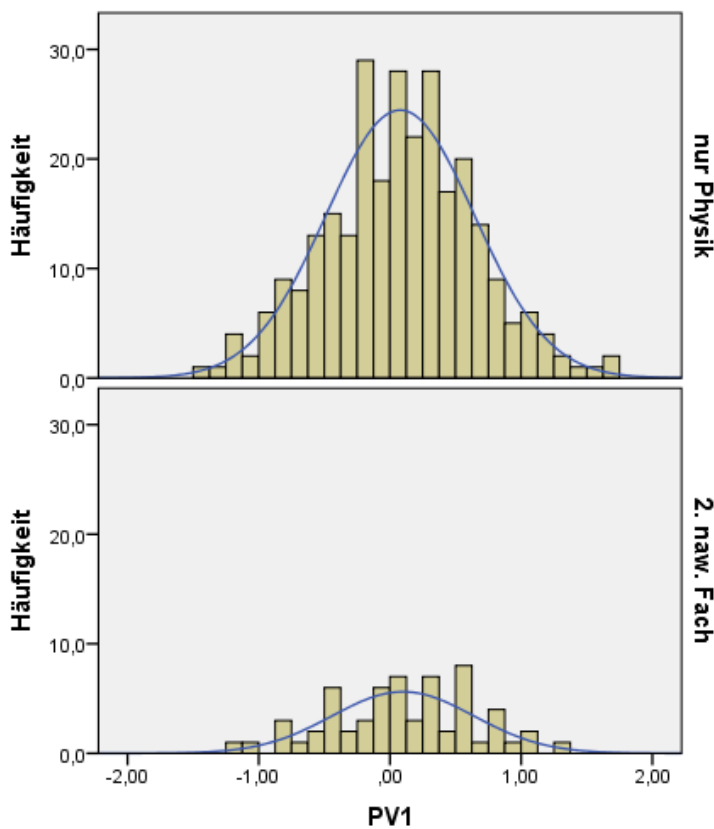
Abhängige Variable:		PV2		
Quelle	Typ III Quadratsumme	df	Quadratischer Mittelwert	F
Korrigiertes Modell	38,769	6	6,462	24,681
Konstanter Term	11,368	1	11,368	43,421
StuPhase	31,000	2	15,500	59,205
lehramt	7,475	1	7,475	28,554
female	3,126	1	3,126	11,940
StuPhase * lehramt	1,544	2	,772	2,948
Fehler	176,716	675	,262	
Gesamtsumme	240,644	682		
Korrigierter Gesamtwert	215,485	681		

Abhängige Variable:		PV3		
Quelle	Typ III Quadratsumme	df	Quadratischer Mittelwert	F
Korrigiertes Modell	38,523	6	6,420	22,227
Konstanter Term	8,042	1	8,042	27,842
StuPhase	28,061	2	14,031	48,573
lehramt	10,721	1	10,721	37,116
female	3,592	1	3,592	12,436
StuPhase * lehramt	2,292	2	1,146	3,968
Fehler	194,979	675	,289	
Gesamtsumme	255,621	682		
Korrigierter Gesamtwert	233,502	681		

Abhängige Variable:		PV4		
Quelle	Typ III Quadratsumme	df	Quadratischer Mittelwert	F
Korrigiertes Modell	31,842	6	5,307	16,618
Konstanter Term	6,261	1	6,261	19,607
StuPhase	19,278	2	9,639	30,184
lehramt	4,515	1	4,515	14,138
female	8,315	1	8,315	26,038
StuPhase * lehramt	1,000	2	,500	1,565
Fehler	215,554	675	,319	
Gesamtsumme	269,152	682		
Korrigierter Gesamtwert	247,396	681		

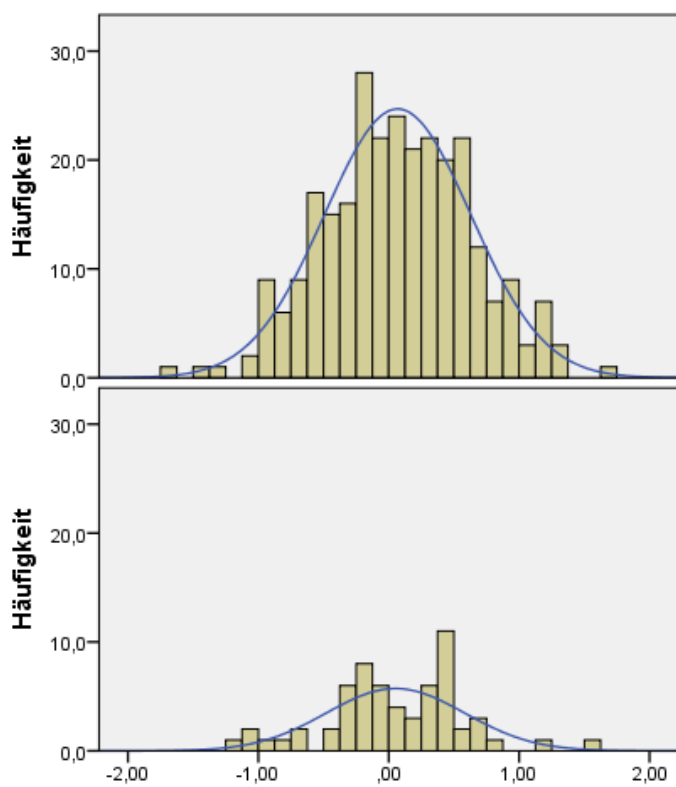
Abhängige Variable:		PV5		
Quelle	Typ III Quadratsumme	df	Quadratischer Mittelwert	F
Korrigiertes Modell	37,626	6	6,271	21,943
Konstanter Term	9,084	1	9,084	31,786
StuPhase	27,497	2	13,748	48,107
lehramt	8,953	1	8,953	31,327
female	3,709	1	3,709	12,978
StuPhase * lehramt	1,293	2	,647	2,262
Fehler	192,907	675	,286	
Gesamtsumme	254,183	682		
Korrigierter Gesamtwert	230,533	681		

Tests auf Normalverteilung Fragestellung 6



$D(278)=0,027$, n.s.
 $S(278)=0,997$, n.s.
 $v=0,034$ (0,146) $Z_S = 0,23$
 $w=0,012$ (0,291) $Z_K = 0,04$

$D(61)=0,060$, n.s.
 $S(61)=0,990$, n.s.
 $v=-0,161$ (0,306) $Z_S = 0,53$
 $w=-0,371$ (0,604) $Z_K = 0,61$

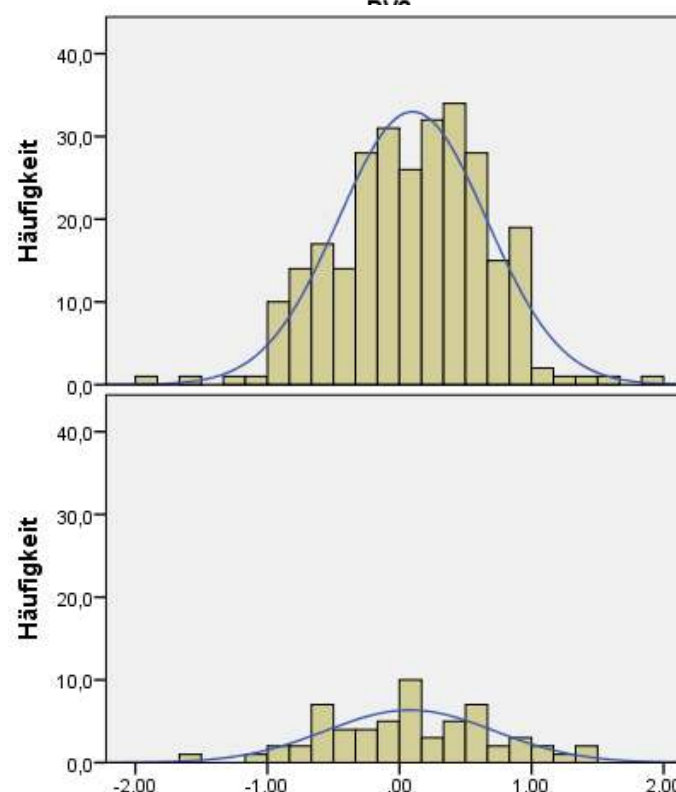


nur Physik

$D(278)=0,026$, n.s.
 $S(278)=0,997$, n.s.
 $\nu=-0,039$ (0,146) $Z_S = 0,27$
 $\mu=-0,053$ (0,291) $Z_K = 0,18$

2. naw. Fach

$D(61)=0,079$, n.s.
 $S(61)=0,973$, n.s.
 $\nu=-0,004$ (0,306) $Z_S = 0,01$
 $\mu=-0,683$ (0,604) $Z_K = 1,13$



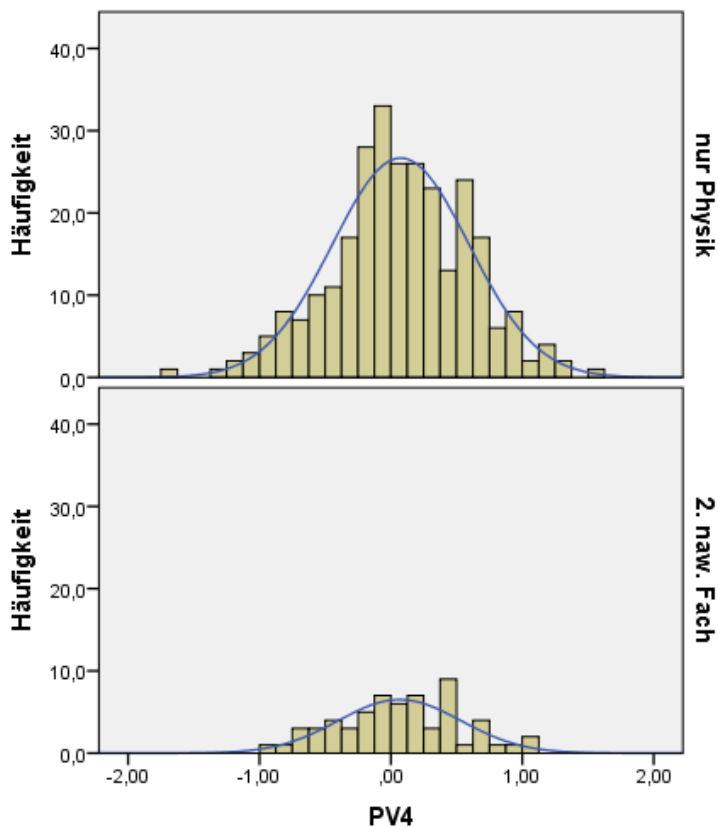
nur Physik

$D(278)=0,044$, n.s.
 $S(278)=0,992$, n.s.
 $\nu=-0,237$ (0,146) $Z_S = 1,62$
 $\mu=0,248$ (0,291) $Z_K = 0,85$

2. naw. Fach

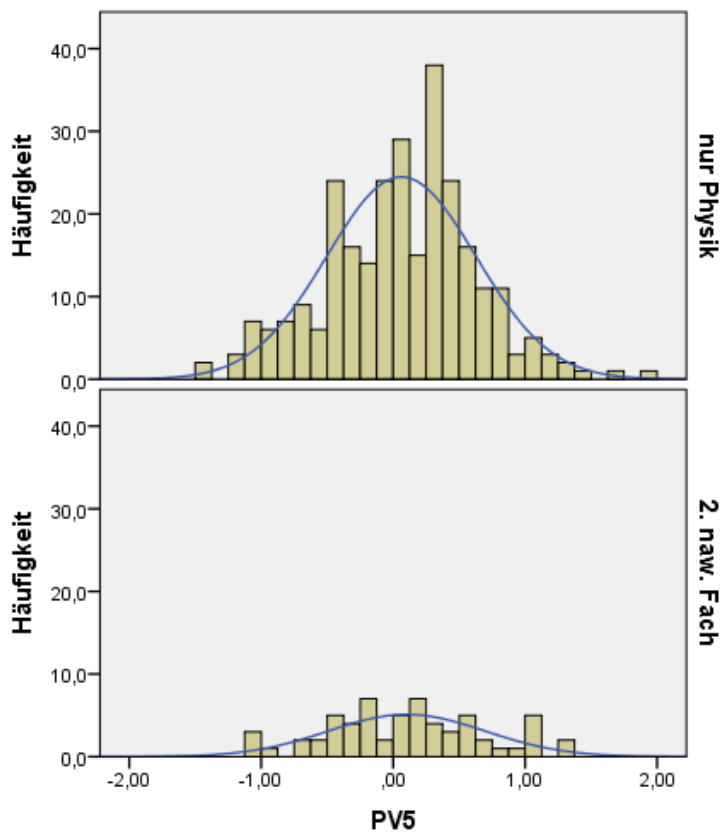
$D(61)=0,070$, n.s.
 $S(61)=0,992$, n.s.
 $\nu=0,015$ (0,306) $Z_S = 0,05$
 $\mu=-0,220$ (0,604) $Z_K = 0,36$

PV3



$D(278)=0,048$, n.s.
 $S(278)=0,994$, n.s.
 $\nu=-0,048$ (0,146) $Z_S = 0,33$
 $w=0,183$ (0,362) $Z_K = 0,50$

$D(61)=0,064$, n.s.
 $S(61)=0,984$, n.s.
 $\nu=0,075$ (0,182) $Z_S = 0,41$
 $w=-0,500$ (0,604) $Z_K = 0,83$



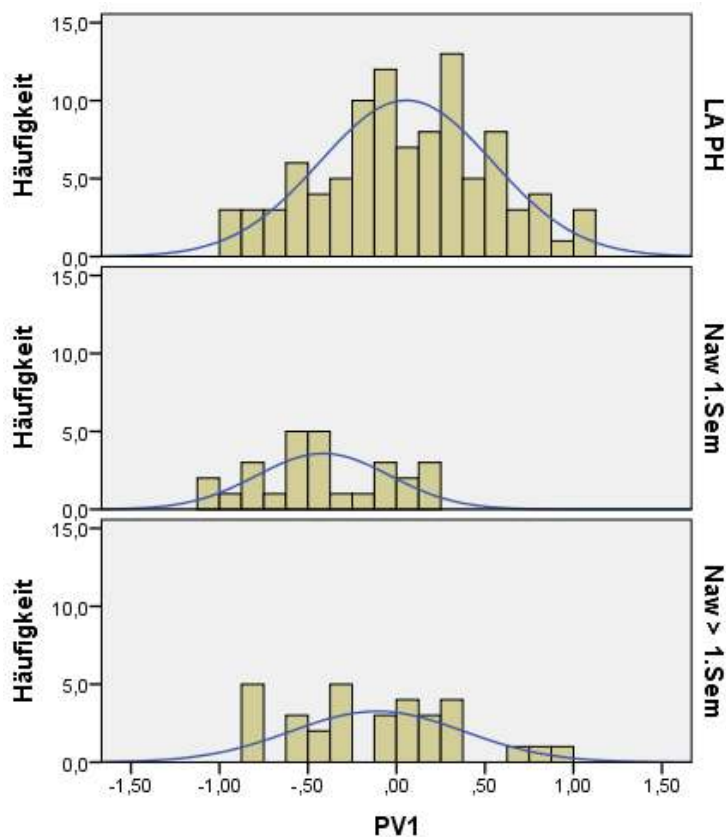
$D(278)=0,045$, n.s.
 $S(278)=0,995$, n.s.
 $\nu=-0,158$ (0,146) $Z_S = 1,08$
 $w=0,301$ (0,291) $Z_K = 1,03$

$D(61)=0,055$, n.s.
 $S(61)=0,989$, n.s.
 $\nu=0,057$ (0,306) $Z_S = 0,18$
 $w=-0,401$ (0,604) $Z_K = 0,66$

Deskriptive Statistik Fragestellung 6

	Nur Physik		Physik + 2. nawi. Fach	
	M	SD	M	SD
PV1	0,0779	0,56668	0,0989	0,54113
PV2	0,068	0,56151	0,0511	0,53182
PV3	0,0983	0,56017	0,0983	0,56017
PV4	0,074	0,51968	0,0659	0,46761
PV5	0,0616	0,56622	0,102	0,59566

Test auf Normalverteilung Fragestellung 7 und 8



$D(98)=0,053$, n.s.

$S(98)=0,988$, n.s.

$\nu=-0,089$ (0,244) $Z_S = 0,36$

$w=-0,475$ (0,483) $Z_K = 0,98$

$D(27)=0,100$, n.s.

$S(27)=0,958$, n.s.

$\nu=-0,141$ (0,448) $Z_S = 0,31$

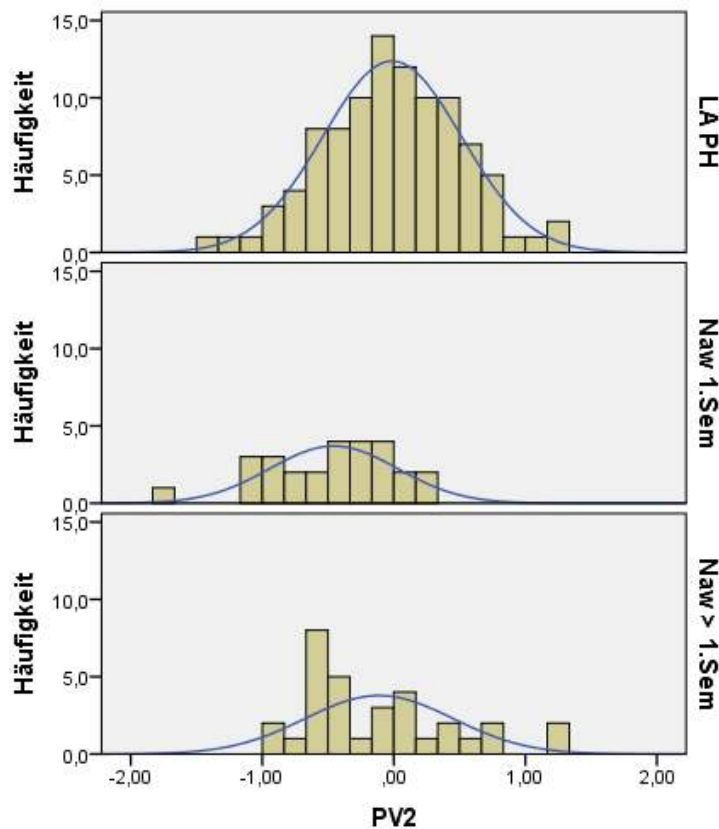
$w=-0,933$ (0,872) $Z_K = 1,07$

$D(32)=0,090$, n.s.

$S(32)=0,960$, n.s.

$\nu=0,300$ (0,414) $Z_S = 0,72$

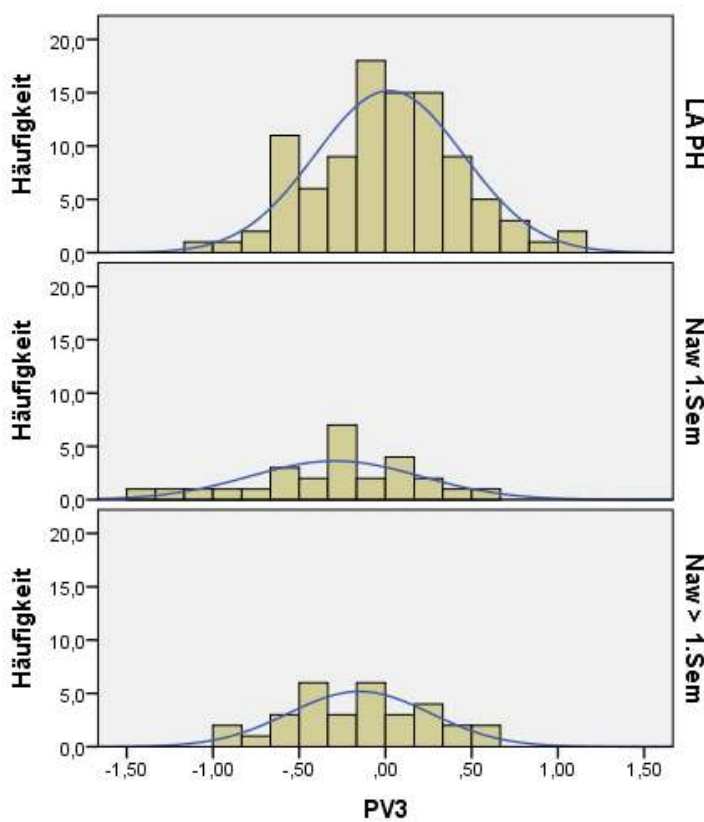
$w=-0,476$ (0,809) $Z_K = 0,59$



$D(98)=0,043$, n.s.
 $S(98)=0,995$, n.s.
 $\nu=-0,097$ (0,244) $Z_S = 0,40$
 $w=-0,048$ (0,483) $Z_K = 0,10$

$D(27)=0,124$, n.s.
 $S(27)=0,966$, n.s.
 $\nu=-0,458$ (0,448) $Z_S = 1,02$
 $w=-0,061$ (0,827) $Z_K = 0,07$

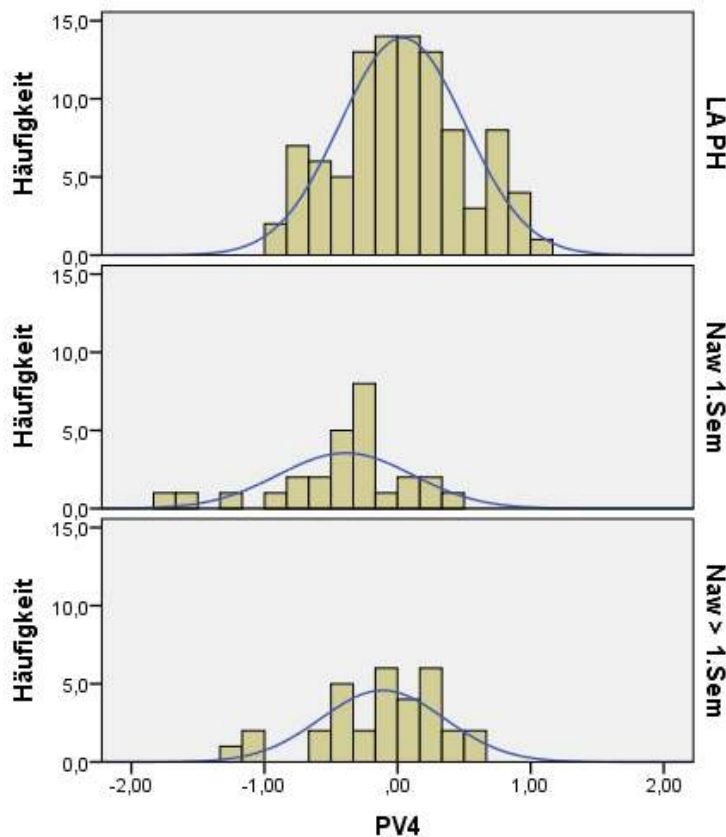
$D(32)=0,164$, $p<0,05^*$
 $S(32)=0,925$, $p<0,05^*$
 $\nu=0,896$ (0,414) $Z_S = 2,16^*$
 $w=0,232$ (0,809) $Z_K = 0,29$



$D(98)=0,059$, n.s.
 $S(98)=0,991$, n.s.
 $\nu=0,104$ (0,244) $Z_S = 0,43$
 $w=0,114$ (0,483) $Z_K = 0,27$

$D(27)=0,102$, n.s.
 $S(27)=0,968$, n.s.
 $\nu=-0,559$ (0,448) $Z_S = 1,25$
 $w=0,243$ (0,872) $Z_K = 0,28$

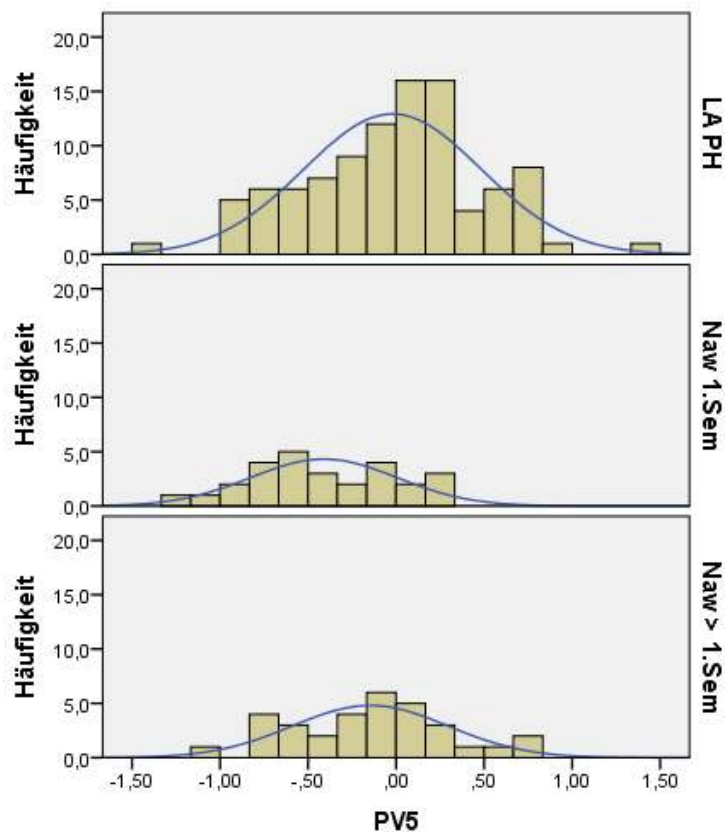
$D(32)=0,090$, n.s.
 $S(32)=0,983$, n.s.
 $\nu=-0,117$ (0,414) $Z_S = 0,28$
 $w=-0,586$ (0,809) $Z_K = 0,72$



D(98)=0,053, n.s.
 S(98)=0,981, n.s.
 $v=0,079$ (0,244) $Z_S = 0,32$
 $w=-0,602$ (0,483) $Z_K = 1,25$

D(27)=0,163, n.s.
S(27)=0,921, p<0,05*
 $v=-0,968$ (0,448) $Z_S = 2,16^*$
 $w=1,189$ (0,872) $Z_K = 1,36$

D(32)=0,109, n.s.
 S(32)=0,949, n.s.
 $v=-0,709$ (0,414) $Z_S = 1,71$
 $w=0,122$ (0,809) $Z_K = 0,15$



D(98)=0,066, n.s.
 S(98)=0,987, n.s.
 $v=-0,111$ (0,244) $Z_S = 0,45$
 $w=-0,018$ (0,483) $Z_K = 0,04$

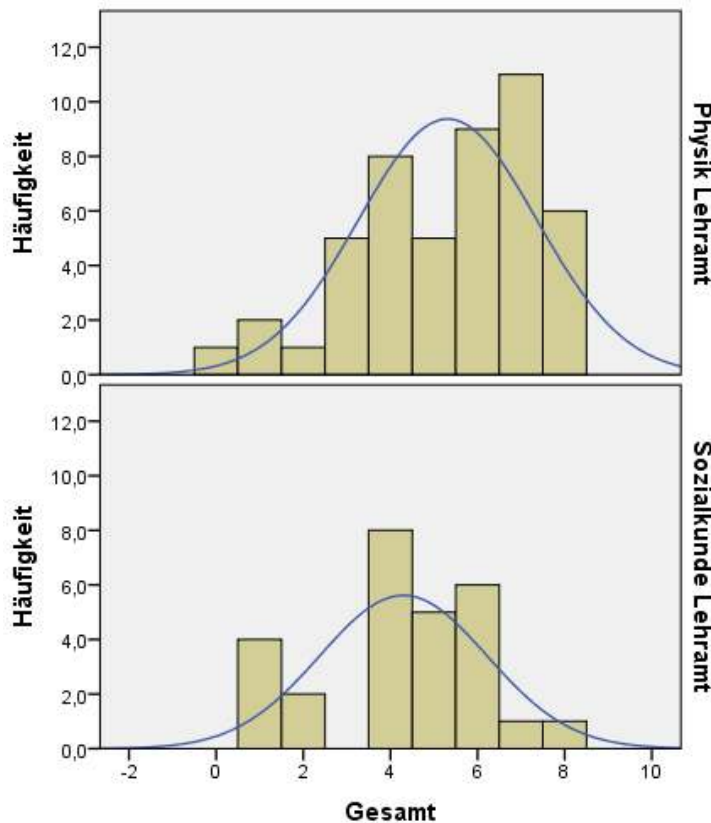
D(27)=0,106, n.s.
 S(27)=0,977, n.s.
 $v=-0,077$ (0,448) $Z_S = 0,17$
 $w=-0,688$ (0,872) $Z_K = 0,79$

D(32)=0,087, n.s.
 S(32)=0,979, n.s.
 $v=-0,022$ (0,414) $Z_S = 0,05$
 $w=-0,443$ (0,809) $Z_K = 0,55$

Deskriptive Statistik Fragestellung 7 und 8

	Lehramt Physik		Integrierte Naturwissenschaften (1.Sem.)		Integrierte Naturwissenschaften (< 1.Sem.)	
	M	SD	M	SD	M	SD
PV1	0,0551	0,48789	-0,4141	0,37600	-0,1084	0,48954
PV2	-0,0131	0,52653	-0,4567	0,48528	-0,1125	0,56261
PV3	0,0223	0,42914	-0,2919	0,49495	-0,1503	0,41083
PV4	0,0369	0,46856	-0,3919	0,50563	-0,1119	0,46498
PV5	-0,0202	0,50405	-0,4107	0,41567	-0,1475	0,44285

Test auf Normalverteilung Fragestellung 9



<p>D(48)=0,173, p=0,001*** S(48)=0,925, p=0,005** $v=-0,662 (0,343) Z_S = 1,93$ $w=-0,181 (0,674) Z_K = 0,27$</p>
<p>D(27)=0,216, p=0,002** S(27)=0,915, p=0,030* $v=-0,387 (0,448) Z_S = 0,86$ $w=-0,438 (0,872) Z_K = 0,50$</p>

12. Abbildungsverzeichnis

Abbildung 1: Strukturmodell zum Scientific Reasoning (aus Mayer 2007)	39
Abbildung 2: Aufgabenbeispiel "Hypothese" (offen)	73
Abbildung 3: Aufgabenbeispiel "Hypothese" (geschlossen)	73
Abbildung 4: ICC zum Aufgabenbeispiel „Hypothese“	75
Abbildung 5: Aufgabenbeispiel "Planung und Durchführung" (offen)	77
Abbildung 6: Aufgabenbeispiel "Planung und Durchführung" (geschlossen)	78
Abbildung 7: ICC des Aufgabenbeispiels "Planung und Durchführung"	80
Abbildung 8: Aufgabenbeispiel „Auswertung und Interpretation“ (offen)	82
Abbildung 9: Aufgabenbeispiel „Auswertung und Interpretation“ (geschlossen)	83
Abbildung 10: ICC des Aufgabenbeispiels „Auswertung und Interpretation“	85
Abbildung 11: Aufgabenbeispiel „Zweck von Modellen" (offen)	86
Abbildung 12: Aufgabenbeispiel "Zweck von Modellen" (geschlossen)	87
Abbildung 13: ICC des Aufgabenbeispiels "Zweck von Modellen"	89
Abbildung 14: ICC Item 73 (Underfit)	101
Abbildung 15: ICC Item 9 (Underfit)	102
Abbildung 16: Wright-Map der Normierung	104
Abbildung 17: Kompetenzstand Fach- und Lehramtsstudierender (Fragestellung 2)	109
Abbildung 18: Kompetenzstand Studierende zu verschiedenen Phasen des Studiums (Fragestellung 3)	109
Abbildung 19: Kompetenzstand Lehramts- und Fachstudierender zu verschiedenen Phasen des Studiums (Fragestellung 4)	111
Abbildung 20: Kompetenzstand nicht- weiblicher und weiblicher Studierender (Fragestellung 5)	112
Abbildung 21: Kompetenzstand Lehramtsstudierender der Physik und Lehramtsstudierende der Physik mit einem weiteren naturwissenschaftlichen Fach (Fragestellung 6)	114
Abbildung 22: Kompetenzstand Grundschullehramtsstudierender (Integrierte Naturwissenschaften) zu verschiedenen Zeiten des Studiums mit Vergleichsgruppe (Fragestellungen 7 und 8)	116
Abbildung 23: Logit-Werte der Aufgaben von vier Testheften	119

13. Tabellenverzeichnis

Tabelle 1: Übersicht über Strukturmodelle wissenschaftlicher Erkenntnisgewinnung (verändert und erweitert nach Wellnitz et al., 2012)	38
Tabelle 2: Kompetenzstrukturmodell zur Naturwissenschaftlichen Erkenntnisgewinnung	40
Tabelle 3: Erläuterungen zu den MC-Optionen und Beispielantworten der Studierenden zum Aufgabenbeispiel „Hypothese“	74
Tabelle 4: Item-Fit-Werte zum Aufgabenbeispiel „Hypothese"	75
Tabelle 5: Erläuterungen zu den MC-Optionen und Beispielantworten der Studierenden zum Aufgabenbeispiel "Planung und Durchführung"	79
Tabelle 6: Item-Fit-Werte des Aufgabenbeispiels "Planung und Durchführung"	80
Tabelle 7: Erläuterungen zu den MC-Optionen und Beispielantworten der Studierenden zum Aufgabenbeispiel „Auswertung und Interpretation“	84
Tabelle 8: Item-Fit-Werte des Aufgabenbeispiels „Auswertung und Interpretation“	85
Tabelle 9: Erläuterungen zu den MC-Optionen und Beispielantworten der Studierenden zum Aufgabenbeispiel „Zweck von Modellen“	88
Tabelle 10: Item-Fit-Werte des Aufgabenbeispiels "Zweck von Modellen"	89
Tabelle 11: Übersicht über die Items (B=Biologie, C=Chemie, P=Physik).....	90
Tabelle 12: Übersicht über das Rotationsdesign (P=Physik, C=Chemie, B=Biologie, U=Untersuchungen, M=Modelle)	91
Tabelle 13: Verteilung der Proband/inn/en der Normierungsstichprobe nach Fächern..	96
Tabelle 14: Verteilung der Proband/inn/en der Normierungsstichprobe nach Universitäten.....	96
Tabelle 15: Prüfung der Modellpassung Hypothese H1a.....	97
Tabelle 16: Prüfung der Modellpassung Hypothese H1b	98
Tabelle 17: Prüfung der Modellpassung Hypothese H1c.....	99
Tabelle 18: Verteilung der Items in die DIF-Kategorien	103
Tabelle 19: Aufteilung der Studienphasen.....	105
Tabelle 20: Verteilung der Proband/inn/en Fragestellungen 2 bis 5	105
Tabelle 21: Unstandardisierte Regressionsgewichte (in Logits) des Hintergrundmodells Fragestellungen 2 bis 5	106
Tabelle 22: Kombinierte ANOVA Fragestellungen 2 bis 5.....	107
Tabelle 23: Kombinierte Deskriptive Statistik Fragestellungen 2 bis 5	108
Tabelle 24: T-Test: Bachelor Phase 1 gegen Bachelor Phase 2	110
Tabelle 25: T-Test: Bachelor Phase 2 gegen Master	110
Tabelle 26: Unstandardisierte Regressionsgewichte (in Logits) des Hintergrundmodells	113
Tabelle 27: T-Test: Lehramtsstudierende der Physik gegen Lehramtsstudierende der Physik mit einem weiterem naturwissenschaftlichen Fach	114

Tabelle 28: Verteilung der Proband/inn/en 115
Tabelle 29: Unstandardisierte Regressionsgewichte (in Logits) des Hintergrundmodells 115
Tabelle 30: Deskriptive Statistik Fragestellungen 7 und 8..... 116
Tabelle 31: T-Test Physik-Lehramtsstudierende gegen Studierende des Faches Integrierte
Naturwissenschaften 117
Tabelle 32: T-Test: Studierende des Faches Integrierte Naturwissenschaften 1. Semester
gegen Semester <1..... 118
Tabelle 33: Reliabilitätsanalyse Physik-Lehramtsstudierende 121
Tabelle 34: Reliabilitätsanalyse Studierende der Sozialkunde 121

14. Kurzfassung

Im Rahmen der Einführung der Bachelor- und Masterstruktur in den Deutschen Universitäten wurde eine Kompetenzorientierung der Studiengänge etabliert. Entsprechend wird auch für das Lehramts- und Fachstudium der Physik eine auf den Kompetenzerwerb ausgerichtete Lehre eingefordert. Zur Kontrolle des Erfolges der Kompetenzförderung im Rahmen des Studiums bedarf es fachdidaktisch und psychologisch fundierter Messinstrumente, die bislang aber noch fehlen (Zlatkin-Troitschanskaia & Kuhn, 2010).

Im Projekt „Ko-WADiS“, in dessen Rahmen diese Dissertation angesiedelt ist, wurde ein Kompetenztest zur Naturwissenschaftlichen Erkenntnisgewinnung entwickelt. Erkenntnisgewinnung wird dabei als ein komplexer Problemlöseprozess im Sinne des Scientific Reasoning angesehen. Auf Grundlage dieser theoretischen Überlegungen wurde zunächst ein Kompetenzmodell mit den beiden Dimensionen „Naturwissenschaftliche Untersuchungen“ (Mayer, 2007) und „Modelle Nutzen“ (Upmeier zu Belzen & Krüger, 2010) adaptiert. Auf Basis dieses Kompetenzmodells wurden Items generiert, pilotiert und schließlich in einer querschnittlichen Untersuchung bei Studierenden der Physik eingesetzt.

In dieser Dissertation werden die theoretischen Grundlagen, die schließlich in das Kompetenzmodell münden, dargestellt. Dabei wird auf wissenschaftstheoretische Betrachtungen der Erkenntnisgewinnung genauso wie auf kognitionswissenschaftliche Modelle Bezug genommen. Anhand von Beispielen wird die Konstruktion der Physikitems beschrieben. Die Testgüte, die durch verschiedene Verfahren abgesichert wird, wird ebenfalls diskutiert.

Auf Grundlage der empirischen Ergebnisse zeigt sich die beste Passung eines eindimensionalen Modells. Auf Basis dieses Modells wurden weitere Analysen vorgenommen. Hierbei wird deutlich, dass erwartungskonform Masterstudierende einen höheren Kompetenzstand als Bachelorstudierende haben. Gleichzeitig gibt es Evidenz für einen signifikanten Vorteil der Fachstudierenden im Vergleich zu den Lehramtsstudierenden. Weiterhin zeigt sich, dass männliche Studierende einen leichten Vorteil vor weiblichen Studierenden haben. Keine Vorteile zeigen sich hingegen für Lehramtsstudierende mit einem weiteren naturwissenschaftlichen Fach.

Die Ergebnisse werden ausführlich diskutiert und Desiderate für weitere Forschungsprojekte werden formuliert.

15. Abstract

Since the conversion of the European and German university system, an orientation on competences was established. Therefore competence orientated studies are also demanded for physicists and for pre-service physics teachers. The success of this paradigm change needs to be evaluated. For this purpose psychological and didactical proven competence tests are necessary. However, these instruments have not been developed yet (Zlatkin-Troitschanskaia & Kuhn, 2010).

This dissertation is evolved within the project “Ko-WADiS”. In this project, a competence test for the processes of scientific inquiry has been developed. Inquiry is viewed as a complex problem solving process and is operationalized as scientific reasoning. Based on these theoretical considerations, a competence model was adapted. The competence model consists of two dimensions “investigations” (Mayer, 2007) and “using models” (Upmeier zu Belzen & Krüger, 2010). On this basis, a paper-and-pencil-test was developed. It was applied in a cross-sectional-study in order to investigate the competences of pre-service physics teachers and physics students.

In this dissertation, the theoretical background which leads to the model of competence is presented. Thereby different concepts of scientific inquiry (philosophy of science, cognitive science) are mentioned. The construction of the items is illustrated on different examples. The validity of this test is also discussed.

Empirical findings show a one dimensional model fits the data best. On the basis of a one dimensional model further analysis were carried out. As expected, students in the master degree program outperform students in the bachelor degree program. Physics students outperform pre-service physics teacher; moreover, there is evidence male students slightly outperform female students. In contrast, there is no evidence for an advantage of pre-service physics teachers with a second scientific subject (biology or chemistry) compared to pre-service teachers who study physics in combination with another subject (e.g. languages or social sciences).

The findings are discussed and implications for further investigations are formulated.

16. Vorveröffentlichungen aus dieser Dissertation

Straube, P. & Nordmeier, V. (2012). Ko-WADiS - Wohin geht es? In V. Nordmeier & H. Grötzebauch (Hrsg.), *PhyDid B, Didaktik der Physik, Beiträge zur DPG-Frühjahrstagung*. Zugriff am 27.04.2016. Verfügbar unter <http://phydid.physik.fu-berlin.de/index.php/phydid-b/article/download/409/548>

Straube, P. & Nordmeier, V. (2013). Längsschnitt zur SI-Kompetenzerfassung Physik(Lehramts)-Studierender. In S. Bernholt (Hrsg.), *Inquiry-based Learning – Forschendes Lernen* (S. 539–541). Kiel: IPN. Zugriff am 27.04.2016. Verfügbar unter http://www.gdcp.de/images/tb2013/TB2013_539_Straube.pdf

Straube, P. & Nordmeier, V. (2013). Ko-WADiS – Kompetenzmodell der Erkenntnisgewinnung. In V. Nordmeier & H. Grötzebauch (Hrsg.), *PhyDid B, Didaktik der Physik, Beiträge zur DPG-Frühjahrstagung*. Zugriff am 27.04.2016. Verfügbar unter <http://www.phydid.de/index.php/phydid-b/article/download/489/622>

Stiller, J., **Straube, P.**, Mathesius, S., Hartmann, S., Tiemann, R., Nordmeier, V. et al. (2014). Fachmethodische Kompetenzen im Lehramtsstudium - das Projekt Ko-WADiS. In S. Bernholt (Hrsg.), *Naturwissenschaftliche Bildung zwischen Science- und Fachunterricht* (S. 156–158). Kiel: GDGP. Zugriff am 13.04.2015. Verfügbar unter http://www.gdcp.de/images/tb2014/TB2014_156_Stiller.pdf

Straube, P., Tiemann, R., Upmeyer zu Belzen, A., Krüger, D. & Nordmeier, V. (2014). Ko-WADiS | Theoretische Grundlagen. In S. Bernholt (Hrsg.), *Naturwissenschaftliche Bildung zwischen Science- und Fachunterricht* (S. 159–161). Kiel: GDGP. Zugriff am 13.04.2015. Verfügbar unter http://www.gdcp.de/images/tb2014/TB2014_159_Straube.pdf

Straube, P., Stiller, J., Tiemann, R. & Nordmeier, V. (2014). Ko-WADiS | Aspekte der Itemkonstruktion. In S. Bernholt (Hrsg.), *Naturwissenschaftliche Bildung zwischen Science- und Fachunterricht* (S. 162–164). Kiel: GDGP. Zugriff am 13.04.2015. Verfügbar unter http://www.gdcp.de/images/tb2014/TB2014_162_Straube.pdf

Stiller, J., **Straube, P.**, Mathesius, S. & Hartmann, S. (2014). Ko-WADiS | Vorläufige Ergebnisse der Pilotierung. In S. Bernholt (Hrsg.), *Naturwissenschaftliche Bildung zwischen Science- und Fachunterricht* (S. 165–167). Kiel: GDGP. Zugriff am 13.04.2015. Verfügbar unter http://www.gdcp.de/images/tb2014/TB2014_165_Stiller.pdf

Straube, P. & Nordmeier, V. (2014). Ko - WADiS - Erkenntnisgewinnungskompetenz von Lehramtsstudierenden. In V. Nordmeier & H. Grötzebauch (Hrsg.), *PhyDid B, Didaktik der Physik, Beiträge zur DPG-Frühjahrstagung*. Berlin: DPG. Zugriff am 13.04.2015. Verfügbar unter <http://www.phydid.de/index.php/phydid-b/article/download/583/643>

Strelow, S., **Straube, P.** & Nordmeier, V. (2014). Falsifikation als naturwissenschaftliche Methode der Erkenntnisgewinnung. *Praxis der Naturwissenschaften - Physik in der Schule*, 63 (8), 22–26.

Straube, P. & Nordmeier, V. (2015). Erkenntnisgewinnungskompetenz im Lehramtsstudium. In S. Bernholt (Hrsg.), *Heterogenität und Diversität - Vielfalt der Voraussetzungen im naturwissenschaftlichen Unterricht*. Kiel: IPN. Zugriff am 27.04.2016. Verfügbar unter http://www.gdcp.de/images/tb2015/TB2015_121_Straube.pdf

Straube, P. & Nordmeier, V. (2016). Physikalische Erkenntnisgewinnung im Lehramtsstudium Ergebnisse eines Quasi-Längsschnitts. In C. Maurer (Hrsg.), *Authentizität und Lernen - das Fach in der Fachdidaktik* (S. 125–127). Regensburg: GDGP. Zugriff am 27.04.2016. Verfügbar unter http://www.gdcp.de/images/tb2016/TB2016_125_Straube.pdf

17. Lebenslauf

Mein Lebenslauf wird aus Gründen des Datenschutzes in der elektronischen Fassung meiner Arbeit nicht veröffentlicht.

Schluss

"It has been said that the purpose of inquiry is not to move from confusion to understanding, but to move from confusion to a higher state of confusion, one in which we clearly understand what we are confused about."

(Michael T. Kane)

Danksagung

Diese Arbeit konnte nur durch die unablässige Unterstützung von diversen Personen entstehen, die mich durch ihren seelischen und/oder fachlichen Beistand, bei der Bewältigung diverser Herausforderungen unterstützt haben.

Zunächst sei meinem Doktorvater Herrn Prof. Dr. Volkhard Nordmeier gedankt, ohne dessen Vertrauen, Zuspruch und Unterstützung ich diese Promotion niemals beginnen und durchführen hätte können. Den Kolleginnen und Kollegen aus dem „Ko-WADiS“-Projekt möchte ich danken, insbesondere Stefan Hartmann, der immer ansprechbar war, wenn es Fragen zur Statistik oder sonstigen Bereichen gab. Auch meinen Mit-Doktorand/inn/en Jurik Stiller und Sabrina Mathesius gilt mein ausdrücklicher Dank für viele interessante Gespräche und die gegenseitige Unterstützung in Fragen der Didaktik oder Methodik. Meinem studentischen Mitarbeiter Christian Rohrsen, der mich immer tatkräftig unterstützte und mir auch in fachlichen und didaktischen Fragen eine Quelle der Inspiration war, möchte ich an dieser Stelle sehr herzlich danken.

Prof. Dr. Hilde Köster danke ich für ihre Unterstützung in sämtlichen Belangen meiner Forschungsarbeit und für die nötigen Freiräume um diese Arbeit neben anderen Aufgaben fertigstellen zu können. Prof. Dr. Helmut Fischler danke ich für seine guten Ratschläge und Versorgung mit Literatur. Meinen beiden Arbeitsgruppen – der Physikdidaktik und dem Lernbereich Sachunterricht an der Freien Universität Berlin – danke ich für den fachlichen Austausch aber auch für viele lustige Momente in den Pausen und Abenden. Danke an Tobi, Steffi, Daniel, Joachim, Nikola, Luzy, René, Helen, Nadia, Philipp G., Tara und Jörg.

Dank gilt auch den weiteren Mitgliedern der Prüfungskommission: Prof. Dr. Katharina Franke, Prof. Dr. Andreas Borowski, Dr. Jürgen Kirstein und Tobias Mühlenbruch.

Schlussendlich danke ich meinen Freunden und meiner Familie, die mich in den letzten Jahren immer unterstützt und ermutigt haben. Insbesondere danke ich meinen Eltern und meinen Geschwistern, denen diese Arbeit gewidmet ist.

Bisher erschienene Bände der Reihe „*Studien zum Physik- und Chemielernen*“

ISSN 1614-8967 (vormals *Studien zum Physiklernen* ISSN 1435-5280)

- 1 Helmut Fischler, Jochen Peuckert (Hrsg.): Concept Mapping in fachdidaktischen Forschungsprojekten der Physik und Chemie
ISBN 978-3-89722-256-4 40.50 EUR
- 2 Anja Schoster: Bedeutungsentwicklungsprozesse beim Lösen algorithmischer Physikaufgaben. *Eine Fallstudie zu Lernprozessen von Schülern im Physikinachhilfeunterricht während der Bearbeitung algorithmischer Physikaufgaben*
ISBN 978-3-89722-045-4 40.50 EUR
- 3 Claudia von Aufschnaiter: Bedeutungsentwicklungen, Interaktionen und situatives Erleben beim Bearbeiten physikalischer Aufgaben
ISBN 978-3-89722-143-7 40.50 EUR
- 4 Susanne Haeberlen: Lernprozesse im Unterricht mit Wasserstromkreisen. *Eine Fallstudie in der Sekundarstufe I*
ISBN 978-3-89722-172-7 40.50 EUR
- 5 Kerstin Haller: Über den Zusammenhang von Handlungen und Zielen. *Eine empirische Untersuchung zu Lernprozessen im physikalischen Praktikum*
ISBN 978-3-89722-242-7 40.50 EUR
- 6 Michaela Horstendahl: Motivationale Orientierungen im Physikunterricht
ISBN 978-3-89722-227-4 50.00 EUR
- 7 Stefan Deylitz: Lernergebnisse in der Quanten-Atomphysik. *Evaluation des Bremer Unterrichtskonzepts*
ISBN 978-3-89722-291-5 40.50 EUR
- 8 Lorenz Hucke: Handlungsregulation und Wissenserwerb in traditionellen und computergestützten Experimenten des physikalischen Praktikums
ISBN 978-3-89722-316-5 50.00 EUR
- 9 Heike Theyßen: Ein Physikpraktikum für Studierende der Medizin. *Darstellung der Entwicklung und Evaluation eines adressatenspezifischen Praktikums nach dem Modell der Didaktischen Rekonstruktion*
ISBN 978-3-89722-334-9 40.50 EUR
- 10 Annette Schick: Der Einfluß von Interesse und anderen selbstbezogenen Kognitionen auf Handlungen im Physikunterricht. *Fallstudien zu Interessenhandlungen im Physikunterricht*
ISBN 978-3-89722-380-6 40.50 EUR
- 11 Roland Berger: Moderne bildgebende Verfahren der medizinischen Diagnostik. *Ein Weg zu interessanterem Physikunterricht*
ISBN 978-3-89722-445-2 40.50 EUR

- 12 Johannes Werner: Vom Licht zum Atom. *Ein Unterrichtskonzept zur Quantenphysik unter Nutzung des Zeigermodells*
ISBN 978-3-89722-471-1 40.50 EUR
- 13 Florian Sander: Verbindung von Theorie und Experiment im physikalischen Praktikum. *Eine empirische Untersuchung zum handlungsbezogenen Vorverständnis und dem Einsatz grafikorientierter Modellbildung im Praktikum*
ISBN 978-3-89722-482-7 40.50 EUR
- 14 Jörn Gerdes: Der Begriff der physikalischen Kompetenz. *Zur Validierung eines Konstruktes*
ISBN 978-3-89722-510-7 40.50 EUR
- 15 Malte Meyer-Arndt: Interaktionen im Physikpraktikum zwischen Studierenden und Betreuern. *Feldstudie zu Bedeutungsentwicklungsprozessen im physikalischen Praktikum*
ISBN 978-3-89722-541-1 40.50 EUR
- 16 Dietmar Höttecke: Die Natur der Naturwissenschaften historisch verstehen. *Fachdidaktische und wissenschaftshistorische Untersuchungen*
ISBN 978-3-89722-607-4 40.50 EUR
- 17 Gil Gabriel Mavanga: Entwicklung und Evaluation eines experimentell- und phänomenorientierten Optikcurriculums. *Untersuchung zu Schülervorstellungen in der Sekundarstufe I in Mosambik und Deutschland*
ISBN 978-3-89722-721-7 40.50 EUR
- 18 Meike Ute Zastrow: Interaktive Experimentieranleitungen. *Entwicklung und Evaluation eines Konzeptes zur Vorbereitung auf das Experimentieren mit Messgeräten im Physikalischen Praktikum*
ISBN 978-3-89722-802-3 40.50 EUR
- 19 Gunnar Friege: Wissen und Problemlösen. *Eine empirische Untersuchung des wissenszentrierten Problemlösens im Gebiet der Elektrizitätslehre auf der Grundlage des Experten-Novizen-Vergleichs*
ISBN 978-3-89722-809-2 40.50 EUR
- 20 Erich Starauschek: Physikunterricht nach dem Karlsruher Physikkurs. *Ergebnisse einer Evaluationsstudie*
ISBN 978-3-89722-823-8 40.50 EUR
- 21 Roland Paatz: Charakteristika analogiebasierten Denkens. *Vergleich von Lernprozessen in Basis- und Zielbereich*
ISBN 978-3-89722-944-0 40.50 EUR
- 22 Silke Mikelskis-Seifert: Die Entwicklung von Metakonzepten zur Teilchenvorstellung bei Schülern. *Untersuchung eines Unterrichts über Modelle mithilfe eines Systems multipler Repräsentationsebenen*
ISBN 978-3-8325-0013-9 40.50 EUR
- 23 Brunhild Landwehr: Distanzen von Lehrkräften und Studierenden des Sachunterrichts zur Physik. *Eine qualitativ-empirische Studie zu den Ursachen*
ISBN 978-3-8325-0044-3 40.50 EUR

- 24 Lydia Murmann: Physiklernen zu Licht, Schatten und Sehen. *Eine phänomenografische Untersuchung in der Primarstufe*
ISBN 978-3-8325-0060-3 40.50 EUR
- 25 Thorsten Bell: Strukturprinzipien der Selbstregulation. *Komplexe Systeme, Elementarisierungen und Lernprozessstudien für den Unterricht der Sekundarstufe II*
ISBN 978-3-8325-0134-1 40.50 EUR
- 26 Rainer Müller: Quantenphysik in der Schule
ISBN 978-3-8325-0186-0 40.50 EUR
- 27 Jutta Roth: Bedeutungsentwicklungsprozesse von Physikerinnen und Physikern in den Dimensionen Komplexität, Zeit und Inhalt
ISBN 978-3-8325-0183-9 40.50 EUR
- 28 Andreas Saniter: Spezifika der Verhaltensmuster fortgeschrittener Studierender der Physik
ISBN 978-3-8325-0292-8 40.50 EUR
- 29 Thomas Weber: Kumulatives Lernen im Physikunterricht. *Eine vergleichende Untersuchung in Unterrichtsgängen zur geometrischen Optik*
ISBN 978-3-8325-0316-1 40.50 EUR
- 30 Markus Rehm: Über die Chancen und Grenzen moralischer Erziehung im naturwissenschaftlichen Unterricht
ISBN 978-3-8325-0368-0 40.50 EUR
- 31 Marion Budde: Lernwirkungen in der Quanten-Atom-Physik. *Fallstudien über Resonanzen zwischen Lernangeboten und SchülerInnen-Vorstellungen*
ISBN 978-3-8325-0483-0 40.50 EUR
- 32 Thomas Reyer: Oberflächenmerkmale und Tiefenstrukturen im Unterricht. *Exemplarische Analysen im Physikunterricht der gymnasialen Sekundarstufe*
ISBN 978-3-8325-0488-5 40.50 EUR
- 33 Christoph Thomas Müller: Subjektive Theorien und handlungsleitende Kognitionen von Lehrern als Determinanten schulischer Lehr-Lern-Prozesse im Physikunterricht
ISBN 978-3-8325-0543-1 40.50 EUR
- 34 Gabriela Jonas-Ahrend: Physiklehrvorstellungen zum Experiment im Physikunterricht
ISBN 978-3-8325-0576-9 40.50 EUR
- 35 Dimitrios Stavrou: Das Zusammenspiel von Zufall und Gesetzmäßigkeiten in der nicht-linearen Dynamik. *Didaktische Analyse und Lernprozesse*
ISBN 978-3-8325-0609-4 40.50 EUR
- 36 Katrin Engeln: Schülerlabors: authentische, aktivierende Lernumgebungen als Möglichkeit, Interesse an Naturwissenschaften und Technik zu wecken
ISBN 978-3-8325-0689-6 40.50 EUR
- 37 Susann Hartmann: Erklärungsvielfalt
ISBN 978-3-8325-0730-5 40.50 EUR

- 38 Knut Neumann: Didaktische Rekonstruktion eines physikalischen Praktikums für Physiker
ISBN 978-3-8325-0762-6 40.50 EUR
- 39 Michael Späth: Kontextbedingungen für Physikunterricht an der Hauptschule. *Möglichkeiten und Ansatzpunkte für einen fachübergreifenden, handlungsorientierten und berufsorientierten Unterricht*
ISBN 978-3-8325-0827-2 40.50 EUR
- 40 Jörg Hirsch: Interesse, Handlungen und situatives Erleben von Schülerinnen und Schülern beim Bearbeiten physikalischer Aufgaben
ISBN 978-3-8325-0875-3 40.50 EUR
- 41 Monika Hüther: Evaluation einer hypermedialen Lernumgebung zum Thema Gasgesetze. *Eine Studie im Rahmen des Physikpraktikums für Studierende der Medizin*
ISBN 978-3-8325-0911-8 40.50 EUR
- 42 Maïke Tesch: Das Experiment im Physikunterricht. *Didaktische Konzepte und Ergebnisse einer Videostudie*
ISBN 978-3-8325-0975-0 40.50 EUR
- 43 Nina Nicolai: Skriptgeleitete Eltern-Kind-Interaktion bei Chemiehausaufgaben. *Eine Evaluationsstudie im Themenbereich Säure-Base*
ISBN 978-3-8325-1013-8 40.50 EUR
- 44 Antje Leisner: Entwicklung von Modellkompetenz im Physikunterricht
ISBN 978-3-8325-1020-6 40.50 EUR
- 45 Stefan Rumann: Evaluation einer Interventionsstudie zur Säure-Base-Thematik
ISBN 978-3-8325-1027-5 40.50 EUR
- 46 Thomas Wilhelm: Konzeption und Evaluation eines Kinematik/Dynamik-Lehrgangs zur Veränderung von Schülervorstellungen mit Hilfe dynamisch ikonischer Repräsentationen und graphischer Modellbildung – mit CD-ROM
ISBN 978-3-8325-1046-6 45.50 EUR
- 47 Andrea Maier-Richter: Computerunterstütztes Lernen mit Lösungsbeispielen in der Chemie. *Eine Evaluationsstudie im Themenbereich Löslichkeit*
ISBN 978-3-8325-1046-6 40.50 EUR
- 48 Jochen Peuckert: Stabilität und Ausprägung kognitiver Strukturen zum Atombegriff
ISBN 978-3-8325-1104-3 40.50 EUR
- 49 Maik Walpuski: Optimierung von experimenteller Kleingruppenarbeit durch Strukturierungshilfen und Feedback
ISBN 978-3-8325-1184-5 40.50 EUR
- 50 Helmut Fischler, Christiane S. Reiners (Hrsg.): Die Teilchenstruktur der Materie im Physik- und Chemieunterricht
ISBN 978-3-8325-1225-5 34.90 EUR
- 51 Claudia Eysel: Interdisziplinäres Lehren und Lernen in der Lehrerbildung. *Eine empirische Studie zum Kompetenzerwerb in einer komplexen Lernumgebung*
ISBN 978-3-8325-1238-5 40.50 EUR

- 52 Johannes Günther: Lehrerfortbildung über die Natur der Naturwissenschaften. *Studien über das Wissenschaftsverständnis von Grundschullehrkräften*
ISBN 978-3-8325-1287-3 40.50 EUR
- 53 Christoph Neugebauer: Lernen mit Simulationen und der Einfluss auf das Problemlösen in der Physik
ISBN 978-3-8325-1300-9 40.50 EUR
- 54 Andreas Schnirch: Gendergerechte Interessen- und Motivationsförderung im Kontext naturwissenschaftlicher Grundbildung. *Konzeption, Entwicklung und Evaluation einer multimedial unterstützten Lernumgebung*
ISBN 978-3-8325-1334-4 40.50 EUR
- 55 Hilde Köster: Freies Explorieren und Experimentieren. *Eine Untersuchung zur selbstbestimmten Gewinnung von Erfahrungen mit physikalischen Phänomenen im Sachunterricht*
ISBN 978-3-8325-1348-1 40.50 EUR
- 56 Eva Heran-Dörr: Entwicklung und Evaluation einer Lehrerfortbildung zur Förderung der physikdidaktischen Kompetenz von Sachunterrichtslehrkräften
ISBN 978-3-8325-1377-1 40.50 EUR
- 57 Agnes Szabone Varnai: Unterstützung des Problemlösens in Physik durch den Einsatz von Simulationen und die Vorgabe eines strukturierten Kooperationsformats
ISBN 978-3-8325-1403-7 40.50 EUR
- 58 Johannes Rethfeld: Aufgabenbasierte Lernprozesse in selbstorganisationsoffenem Unterricht der Sekundarstufe I zum Themengebiet ELEKTROSTATIK. *Eine Feldstudie in vier 10. Klassen zu einer kartenbasierten Lernumgebung mit Aufgaben aus der Elektrostatik*
ISBN 978-3-8325-1416-7 40.50 EUR
- 59 Christian Henke: Experimentell-naturwissenschaftliche Arbeitsweisen in der Oberstufe. *Untersuchung am Beispiel des HIGHSEA-Projekts in Bremerhaven*
ISBN 978-3-8325-1515-7 40.50 EUR
- 60 Lutz Kasper: Diskursiv-narrative Elemente für den Physikunterricht. *Entwicklung und Evaluation einer multimedialen Lernumgebung zum Erdmagnetismus*
ISBN 978-3-8325-1537-9 40.50 EUR
- 61 Thorid Rabe: Textgestaltung und Aufforderung zu Selbsterklärungen beim Physiklernen mit Multimedia
ISBN 978-3-8325-1539-3 40.50 EUR
- 62 Ina Glemnitz: Vertikale Vernetzung im Chemieunterricht. *Ein Vergleich von traditionellem Unterricht mit Unterricht nach Chemie im Kontext*
ISBN 978-3-8325-1628-4 40.50 EUR
- 63 Erik Einhaus: Schülerkompetenzen im Bereich Wärmelehre. *Entwicklung eines Testinstruments zur Überprüfung und Weiterentwicklung eines normativen Modells fachbezogener Kompetenzen*
ISBN 978-3-8325-1630-7 40.50 EUR

- 64 Jasmin Neuroth: Concept Mapping als Lernstrategie. *Eine Interventionsstudie zum Chemielernen aus Texten*
ISBN 978-3-8325-1659-8 40.50 EUR
- 65 Hans Gerd Hegeler-Burkhart: Zur Kommunikation von Hauptschülerinnen und Hauptschülern in einem handlungsorientierten und fächerübergreifenden Unterricht mit physikalischen und technischen Inhalten
ISBN 978-3-8325-1667-3 40.50 EUR
- 66 Karsten Rincke: Sprachentwicklung und Fachlernen im Mechanikunterricht. *Sprache und Kommunikation bei der Einführung in den Kraftbegriff*
ISBN 978-3-8325-1699-4 40.50 EUR
- 67 Nina Strehle: Das Ion im Chemieunterricht. *Alternative Schülervorstellungen und curriculare Konsequenzen*
ISBN 978-3-8325-1710-6 40.50 EUR
- 68 Martin Hopf: Problemorientierte Schülerexperimente
ISBN 978-3-8325-1711-3 40.50 EUR
- 69 Anne Beerenwinkel: Fostering conceptual change in chemistry classes using expository texts
ISBN 978-3-8325-1721-2 40.50 EUR
- 70 Roland Berger: Das Gruppenpuzzle im Physikunterricht der Sekundarstufe II. *Eine empirische Untersuchung auf der Grundlage der Selbstbestimmungstheorie der Motivation*
ISBN 978-3-8325-1732-8 40.50 EUR
- 71 Giuseppe Colicchia: Physikunterricht im Kontext von Medizin und Biologie. *Entwicklung und Erprobung von Unterrichtseinheiten*
ISBN 978-3-8325-1746-5 40.50 EUR
- 72 Sandra Winheller: Geschlechtsspezifische Auswirkungen der Lehrer-Schüler-Interaktion im Chemieanfangsunterricht
ISBN 978-3-8325-1757-1 40.50 EUR
- 73 Isabel Wahser: Training von naturwissenschaftlichen Arbeitsweisen zur Unterstützung experimenteller Kleingruppenarbeit im Fach Chemie
ISBN 978-3-8325-1815-8 40.50 EUR
- 74 Claus Brell: Lernmedien und Lernerfolg - reale und virtuelle Materialien im Physikunterricht. *Empirische Untersuchungen in achten Klassen an Gymnasien (Laborstudie) zum Computereinsatz mit Simulation und IBE*
ISBN 978-3-8325-1829-5 40.50 EUR
- 75 Rainer Wackermann: Überprüfung der Wirksamkeit eines Basismodell-Trainings für Physiklehrer
ISBN 978-3-8325-1882-0 40.50 EUR
- 76 Oliver Tepner: Effektivität von Aufgaben im Chemieunterricht der Sekundarstufe I
ISBN 978-3-8325-1919-3 40.50 EUR

- 77 Claudia Geyer: Museums- und Science-Center-Besuche im naturwissenschaftlichen Unterricht aus einer motivationalen Perspektive. *Die Sicht von Lehrkräften und Schülerinnen und Schülern*
ISBN 978-3-8325-1922-3 40.50 EUR
- 78 Tobias Leonhard: Professionalisierung in der Lehrerbildung. *Eine explorative Studie zur Entwicklung professioneller Kompetenzen in der Lehrererstausbildung*
ISBN 978-3-8325-1924-7 40.50 EUR
- 79 Alexander Kauertz: Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben
ISBN 978-3-8325-1925-4 40.50 EUR
- 80 Regina Hübinger: Schüler auf Weltreise. *Entwicklung und Evaluation von Lehr-/Lernmaterialien zur Förderung experimentell-naturwissenschaftlicher Kompetenzen für die Jahrgangsstufen 5 und 6*
ISBN 978-3-8325-1932-2 40.50 EUR
- 81 Christine Waltner: Physik lernen im Deutschen Museum
ISBN 978-3-8325-1933-9 40.50 EUR
- 82 Torsten Fischer: Handlungsmuster von Physiklehrkräften beim Einsatz neuer Medien. *Fallstudien zur Unterrichtspraxis*
ISBN 978-3-8325-1948-3 42.00 EUR
- 83 Corinna Kieren: Chemiehausaufgaben in der Sekundarstufe I des Gymnasiums. *Fragebogenerhebung zur gegenwärtigen Praxis und Entwicklung eines optimierten Hausaufgabendesigns im Themenbereich Säure-Base*
978-3-8325-1975-9 37.00 EUR
- 84 Marco Thiele: Modelle der Thermohalinen Zirkulation im Unterricht. *Eine empirische Studie zur Förderung des Modellverständnisses*
ISBN 978-3-8325-1982-7 40.50 EUR
- 85 Bernd Zinn: Physik lernen, um Physik zu lehren. *Eine Möglichkeit für interessanteren Physikunterricht*
ISBN 978-3-8325-1995-7 39.50 EUR
- 86 Esther Klaes: Außerschulische Lernorte im naturwissenschaftlichen Unterricht. *Die Perspektive der Lehrkraft*
ISBN 978-3-8325-2006-9 43.00 EUR
- 87 Marita Schmidt: Kompetenzmodellierung und -diagnostik im Themengebiet Energie der Sekundarstufe I. *Entwicklung und Erprobung eines Testinventars*
ISBN 978-3-8325-2024-3 37.00 EUR
- 88 Gudrun Franke-Braun: Aufgaben mit gestuften Lernhilfen. *Ein Aufgabenformat zur Förderung der sachbezogenen Kommunikation und Lernleistung für den naturwissenschaftlichen Unterricht*
ISBN 978-3-8325-2026-7 38.00 EUR
- 89 Silke Klos: Kompetenzförderung im naturwissenschaftlichen Anfangsunterricht. *Der Einfluss eines integrierten Unterrichtskonzepts*
ISBN 978-3-8325-2133-2 37.00 EUR

- 90 Ulrike Elisabeth Burkard: Quantenphysik in der Schule. *Bestandsaufnahme, Perspektiven und Weiterentwicklungsmöglichkeiten durch die Implementation eines Medienservers*
ISBN 978-3-8325-2215-5 43.00 EUR
- 91 Ulrike Gromadecki: Argumente in physikalischen Kontexten. *Welche Geltungsgründe halten Physikanfänger für überzeugend?*
ISBN 978-3-8325-2250-6 41.50 EUR
- 92 Jürgen Bruns: Auf dem Weg zur Förderung naturwissenschaftsspezifischer Vorstellungen von zukünftigen Chemie-Lehrenden
ISBN 978-3-8325-2257-5 43.50 EUR
- 93 Cornelius Marsch: Räumliche Atomvorstellung. *Entwicklung und Erprobung eines Unterrichtskonzeptes mit Hilfe des Computers*
ISBN 978-3-8325-2293-3 82.50 EUR
- 94 Maja Brückmann: Sachstrukturen im Physikunterricht. *Ergebnisse einer Videostudie*
ISBN 978-3-8325-2272-8 39.50 EUR
- 95 Sabine Fechner: Effects of Context-oriented Learning on Student Interest and Achievement in Chemistry Education
ISBN 978-3-8325-2343-5 36.50 EUR
- 96 Clemens Nagel: eLearning im Physikalischen Anfängerpraktikum
ISBN 978-3-8325-2355-8 39.50 EUR
- 97 Josef Riese: Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkräften
ISBN 978-3-8325-2376-3 39.00 EUR
- 98 Sascha Bernholt: Kompetenzmodellierung in der Chemie. *Theoretische und empirische Reflexion am Beispiel des Modells hierarchischer Komplexität*
ISBN 978-3-8325-2447-0 40.00 EUR
- 99 Holger Christoph Stawitz: Auswirkung unterschiedlicher Aufgabenprofile auf die Schülerleistung. *Vergleich von Naturwissenschafts- und Problemlöseaufgaben der PISA 2003-Studie*
ISBN 978-3-8325-2451-7 37.50 EUR
- 100 Hans Ernst Fischer, Elke Sumfleth (Hrsg.): nwu-essen – 10 Jahre Essener Forschung zum naturwissenschaftlichen Unterricht
ISBN 978-3-8325-3331-1 40.00 EUR
- 101 Hendrik Härtig: Sachstrukturen von Physikschulbüchern als Grundlage zur Bestimmung der Inhaltsvalidität eines Tests
ISBN 978-3-8325-2512-5 34.00 EUR
- 102 Thomas Grüß-Niehaus: Zum Verständnis des Löslichkeitskonzeptes im Chemieunterricht. *Der Effekt von Methoden progressiver und kollaborativer Reflexion*
ISBN 978-3-8325-2537-8 40.50 EUR
- 103 Patrick Bronner: Quantenoptische Experimente als Grundlage eines Curriculums zur Quantenphysik des Photons
ISBN 978-3-8325-2540-8 36.00 EUR

- 104 Adrian Voßkühler: Blickbewegungsmessung an Versuchsaufbauten. *Studien zur Wahrnehmung, Verarbeitung und Usability von physikbezogenen Experimenten am Bildschirm und in der Realität*
ISBN 978-3-8325-2548-4 47.50 EUR
- 105 Verena Tobias: Newton'sche Mechanik im Anfangsunterricht. *Die Wirksamkeit einer Einführung über die zweidimensionale Dynamik auf das Lehren und Lernen*
ISBN 978-3-8325-2558-3 54.00 EUR
- 106 Christian Rogge: Entwicklung physikalischer Konzepte in aufgabenbasierten Lernumgebungen
ISBN 978-3-8325-2574-3 45.00 EUR
- 107 Mathias Ropohl: Modellierung von Schülerkompetenzen im Basiskonzept Chemische Reaktion. *Entwicklung und Analyse von Testaufgaben*
ISBN 978-3-8325-2609-2 36.50 EUR
- 108 Christoph Kulgemeyer: Physikalische Kommunikationskompetenz. *Modellierung und Diagnostik*
ISBN 978-3-8325-2674-0 44.50 EUR
- 109 Jennifer Olszewski: The Impact of Physics Teachers' Pedagogical Content Knowledge on Teacher Actions and Student Outcomes
ISBN 978-3-8325-2680-1 33.50 EUR
- 110 Annika Ohle: Primary School Teachers' Content Knowledge in Physics and its Impact on Teaching and Students' Achievement
ISBN 978-3-8325-2684-9 36.50 EUR
- 111 Susanne Mannel: Assessing scientific inquiry. *Development and evaluation of a test for the low-performing stage*
ISBN 978-3-8325-2761-7 40.00 EUR
- 112 Michael Plomer: Physik physiologisch passend praktiziert. *Eine Studie zur Lernwirksamkeit von traditionellen und adressatenspezifischen Physikpraktika für die Physiologie*
ISBN 978-3-8325-2804-1 34.50 EUR
- 113 Alexandra Schulz: Experimentierspezifische Qualitätsmerkmale im Chemieunterricht. *Eine Videostudie*
ISBN 978-3-8325-2817-1 40.00 EUR
- 114 Franz Boczianowski: Eine empirische Untersuchung zu Vektoren im Physikunterricht der Mittelstufe
ISBN 978-3-8325-2843-0 39.50 EUR
- 115 Maria Ploog: Internetbasiertes Lernen durch Textproduktion im Fach Physik
ISBN 978-3-8325-2853-9 39.50 EUR
- 116 Anja Dhein: Lernen in Explorier- und Experimentiersituationen. *Eine explorative Studie zu Bedeutungsentwicklungsprozessen bei Kindern im Alter zwischen 4 und 6 Jahren*
ISBN 978-3-8325-2859-1 45.50 EUR

- 117 Irene Neumann: Beyond Physics Content Knowledge. *Modeling Competence Regarding Nature of Scientific Inquiry and Nature of Scientific Knowledge*
ISBN 978-3-8325-2880-5 37.00 EUR
- 118 Markus Emden: Prozessorientierte Leistungsmessung des naturwissenschaftlich-experimentellen Arbeitens. *Eine vergleichende Studie zu Diagnoseinstrumenten zu Beginn der Sekundarstufe I*
ISBN 978-3-8325-2867-6 38.00 EUR
- 119 Birgit Hofmann: Analyse von Blickbewegungen von Schülern beim Lesen von physikbezogenen Texten mit Bildern. *Eye Tracking als Methodenwerkzeug in der physikdidaktischen Forschung*
ISBN 978-3-8325-2925-3 59.00 EUR
- 120 Rebecca Knobloch: Analyse der fachinhaltlichen Qualität von Schüleräußerungen und deren Einfluss auf den Lernerfolg. *Eine Videostudie zu kooperativer Kleingruppenarbeit*
ISBN 978-3-8325-3006-8 36.50 EUR
- 121 Julia Hostenbach: Entwicklung und Prüfung eines Modells zur Beschreibung der Bewertungskompetenz im Chemieunterricht
ISBN 978-3-8325-3013-6 38.00 EUR
- 122 Anna Windt: Naturwissenschaftliches Experimentieren im Elementarbereich. *Evaluation verschiedener Lernsituationen*
ISBN 978-3-8325-3020-4 43.50 EUR
- 123 Eva Kölbach: Kontexteinflüsse beim Lernen mit Lösungsbeispielen
ISBN 978-3-8325-3025-9 38.50 EUR
- 124 Anna Lau: Passung und vertikale Vernetzung im Chemie- und Physikunterricht
ISBN 978-3-8325-3021-1 36.00 EUR
- 125 Jan Lamprecht: Ausbildungswege und Komponenten professioneller Handlungskompetenz. *Vergleich von Quereinsteigern mit Lehramtsabsolventen für Gymnasien im Fach Physik*
ISBN 978-3-8325-3035-8 38.50 EUR
- 126 Ulrike Böhm: Förderung von Verstehensprozessen unter Einsatz von Modellen
ISBN 978-3-8325-3042-6 41.00 EUR
- 127 Sabrina Dollny: Entwicklung und Evaluation eines Testinstruments zur Erfassung des fachspezifischen Professionswissens von Chemielehrkräften
ISBN 978-3-8325-3046-4 37.00 EUR
- 128 Monika Zimmermann: Naturwissenschaftliche Bildung im Kindergarten. *Eine integrative Längsschnittstudie zur Kompetenzentwicklung von Erzieherinnen*
ISBN 978-3-8325-3053-2 54.00 EUR
- 129 Ulf Saballus: Über das Schlussfolgern von Schülerinnen und Schülern zu öffentlichen Kontroversen mit naturwissenschaftlichem Hintergrund. *Eine Fallstudie*
ISBN 978-3-8325-3086-0 39.50 EUR
- 130 Olaf Krey: Zur Rolle der Mathematik in der Physik. *Wissenschaftstheoretische Aspekte und Vorstellungen Physiklernender*
ISBN 978-3-8325-3101-0 46.00 EUR

- 131 Angelika Wolf: Zusammenhänge zwischen der Eigenständigkeit im Physikunterricht, der Motivation, den Grundbedürfnissen und dem Lernerfolg von Schülern
ISBN 978-3-8325-3161-4 45.00 EUR
- 132 Johannes Börlin: Das Experiment als Lerngelegenheit. *Vom interkulturellen Vergleich des Physikunterrichts zu Merkmalen seiner Qualität*
ISBN 978-3-8325-3170-6 45.00 EUR
- 133 Olaf Uhden: Mathematisches Denken im Physikunterricht. *Theorieentwicklung und Problemanalyse*
ISBN 978-3-8325-3170-6 45.00 EUR
- 134 Christoph Gut: Modellierung und Messung experimenteller Kompetenz. *Analyse eines large-scale Experimentiertests*
ISBN 978-3-8325-3213-0 40.00 EUR
- 135 Antonio Rueda: Lernen mit ExploMultimedial in kolumbianischen Schulen. *Analyse von kurzzeitigen Lernprozessen und der Motivation beim länderübergreifenden Einsatz einer deutschen computergestützten multimedialen Lernumgebung für den naturwissenschaftlichen Unterricht*
ISBN 978-3-8325-3218-5 45.50 EUR
- 136 Krisztina Berger: Bilder, Animationen und Notizen. *Empirische Untersuchung zur Wirkung einfacher visueller Repräsentationen und Notizen auf den Wissenserwerb in der Optik*
ISBN 978-3-8325-3238-3 41.50 EUR
- 137 Antony Crossley: Untersuchung des Einflusses unterschiedlicher physikalischer Konzepte auf den Wissenserwerb in der Thermodynamik der Sekundarstufe I
ISBN 978-3-8325-3275-8 40.00 EUR
- 138 Tobias Viering: Entwicklung physikalischer Kompetenz in der Sekundarstufe I. *Validierung eines Kompetenzentwicklungsmodells für das Energiekonzept im Bereich Fachwissen*
ISBN 978-3-8325-3277-2 37.00 EUR
- 139 Nico Schreiber: Diagnostik experimenteller Kompetenz. *Validierung technologiegestützter Testverfahren im Rahmen eines Kompetenzstrukturmodells*
ISBN 978-3-8325-3284-0 39.00 EUR
- 140 Sarah Hundertmark: Einblicke in kollaborative Lernprozesse. *Eine Fallstudie zur reflektierenden Zusammenarbeit unterstützt durch die Methoden Concept Mapping und Lernbegleitbogen*
ISBN 978-3-8325-3251-2 43.00 EUR
- 141 Ronny Scherer: Analyse der Struktur, Messinvarianz und Ausprägung komplexer Problemlösekompetenz im Fach Chemie. *Eine Querschnittstudie in der Sekundarstufe I und am Übergang zur Sekundarstufe II*
ISBN 978-3-8325-3312-0 43.00 EUR
- 142 Patricia Heitmann: Bewertungskompetenz im Rahmen naturwissenschaftlicher Problemlöseprozesse. *Modellierung und Diagnose der Kompetenzen Bewertung und analytisches Problemlösen für das Fach Chemie*
ISBN 978-3-8325-3314-4 37.00 EUR

- 143 Jan Fleischhauer: Wissenschaftliches Argumentieren und Entwicklung von Konzepten beim Lernen von Physik
ISBN 978-3-8325-3325-0 35.00 EUR
- 144 Nermin Özcan: Zum Einfluss der Fachsprache auf die Leistung im Fach Chemie. *Eine Förderstudie zur Fachsprache im Chemieunterricht*
ISBN 978-3-8325-3328-1 36.50 EUR
- 145 Helena van Vorst: Kontextmerkmale und ihr Einfluss auf das Schülerinteresse im Fach Chemie
ISBN 978-3-8325-3321-2 38.50 EUR
- 146 Janine Cappell: Fachspezifische Diagnosekompetenz angehender Physiklehrkräfte in der ersten Ausbildungsphase
ISBN 978-3-8325-3356-4 38.50 EUR
- 147 Susanne Bley: Förderung von Transferprozessen im Chemieunterricht
ISBN 978-3-8325-3407-3 40.50 EUR
- 148 Cathrin Blaes: Die übungsgestützte Lehrerpräsentation im Chemieunterricht der Sekundarstufe I. *Evaluation der Effektivität*
ISBN 978-3-8325-3409-7 43.50 EUR
- 149 Julia Suckut: Die Wirksamkeit von piko-OWL als Lehrerfortbildung. Eine Evaluation zum Projekt *Physik im Kontext* in Fallstudien
ISBN 978-3-8325-3440-0 45.00 EUR
- 150 Alexandra Dorschu: Die Wirkung von Kontexten in Physikkompetenztestaufgaben
ISBN 978-3-8325-3446-2 37.00 EUR
- 151 Jochen Scheid: Multiple Repräsentationen, Verständnis physikalischer Experimente und kognitive Aktivierung: *Ein Beitrag zur Entwicklung der Aufgabenkultur*
ISBN 978-3-8325-3449-3 49.00 EUR
- 152 Tim Plasa: Die Wahrnehmung von Schülerlaboren und Schülerforschungszentren
ISBN 978-3-8325-3483-7 35.50 EUR
- 153 Felix Schoppmeier: Physikkompetenz in der gymnasialen Oberstufe. *Entwicklung und Validierung eines Kompetenzstrukturmodells für den Kompetenzbereich Umgang mit Fachwissen*
ISBN 978-3-8325-3502-5 36.00 EUR
- 154 Katharina Groß: Experimente alternativ dokumentieren. *Eine qualitative Studie zur Förderung der Diagnose- und Differenzierungskompetenz in der Chemielehrerbildung*
ISBN 978-3-8325-3508-7 43.50 EUR
- 155 Barbara Hank: Konzeptwandelprozesse im Anfangsunterricht Chemie. *Eine quasiexperimentelle Längsschnittstudie*
ISBN 978-3-8325-3519-3 38.50 EUR

- 156 Katja Freyer: Zum Einfluss von Studieneingangsvoraussetzungen auf den Studienerfolg Erstsemesterstudierender im Fach Chemie
ISBN 978-3-8325-3544-5 38.00 EUR
- 157 Alexander Rachel: Auswirkungen instruktionaler Hilfen bei der Einführung des (Ferro-)Magnetismus. *Eine Vergleichsstudie in der Primar- und Sekundarstufe*
ISBN 978-3-8325-3548-3 43.50 EUR
- 158 Sebastian Ritter: Einfluss des Lerninhalts Nanogrößeneffekte auf Teilchen- und Teilchenmodellvorstellungen von Schülerinnen und Schülern
ISBN 978-3-8325-3558-2 36.00 EUR
- 159 Andrea Harbach: Problemorientierung und Vernetzung in kontextbasierten Lernaufgaben
ISBN 978-3-8325-3564-3 39.00 EUR
- 160 David Obst: Interaktive Tafeln im Physikunterricht. *Entwicklung und Evaluation einer Lehrerfortbildung*
ISBN 978-3-8325-3582-7 40.50 EUR
- 161 Sophie Kirschner: Modellierung und Analyse des Professionswissens von Physik Lehrkräften
ISBN 978-3-8325-3601-5 35.00 EUR
- 162 Katja Stief: Selbstregulationsprozesse und Hausaufgabenmotivation im Chemieunterricht
ISBN 978-3-8325-3631-2 34.00 EUR
- 163 Nicola Meschede: Professionelle Wahrnehmung der inhaltlichen Strukturierung im naturwissenschaftlichen Grundschulunterricht. *Theoretische Beschreibung und empirische Erfassung*
ISBN 978-3-8325-3668-8 37.00 EUR
- 164 Johannes Maximilian Barth: Experimentieren im Physikunterricht der gymnasialen Oberstufe. *Eine Rekonstruktion übergeordneter Einbettungsstrategien*
ISBN 978-3-8325-3681-7 39.00 EUR
- 165 Sandra Lein: Das Betriebspraktikum in der Lehrerbildung. *Eine Untersuchung zur Förderung der Wissenschafts- und Technikbildung im allgemeinbildenden Unterricht*
ISBN 978-3-8325-3698-5 40.00 EUR
- 166 Veranika Maiseyenko: Modellbasiertes Experimentieren im Unterricht. *Praxistauglichkeit und Lernwirkungen*
ISBN 978-3-8325-3708-1 38.00 EUR
- 167 Christoph Stolzenberger: Der Einfluss der didaktischen Lernumgebung auf das Erreichen geforderter Bildungsziele am Beispiel der W- und P-Seminare im Fach Physik
ISBN 978-3-8325-3708-1 38.00 EUR
- 168 Pia Altenburger: Mehrebenenregressionsanalysen zum Physiklernen im Sachunterricht der Primarstufe. *Ergebnisse einer Evaluationsstudie.*
ISBN 978-3-8325-3717-3 37.50 EUR

- 169 Nora Ferber: Entwicklung und Validierung eines Testinstruments zur Erfassung von Kompetenzentwicklung im Fach Chemie in der Sekundarstufe I
ISBN 978-3-8325-3727-2 39.50 EUR
- 170 Anita Stender: Unterrichtsplanung: Vom Wissen zum Handeln. Theoretische Entwicklung und empirische Überprüfung des Transformationsmodells der Unterrichtsplanung
ISBN 978-3-8325-3750-0 41.50 EUR
- 171 Jenna Koenen: Entwicklung und Evaluation von experimentunterstützten Lösungsbeispielen zur Förderung naturwissenschaftlich-experimenteller Arbeitsweisen
ISBN 978-3-8325-3785-2 43.00 EUR
- 172 Teresa Henning: Empirische Untersuchung kontextorientierter Lernumgebungen in der Hochschuldidaktik. *Entwicklung und Evaluation kontextorientierter Aufgaben in der Studieneingangsphase für Fach- und Nebenfachstudierende der Physik*
ISBN 978-3-8325-3801-9 43.00 EUR
- 173 Alexander Pusch: Fachspezifische Instrumente zur Diagnose und individuellen Förderung von Lehramtsstudierenden der Physik
ISBN 978-3-8325-3829-3 38.00 EUR
- 174 Christoph Vogelsang: Validierung eines Instruments zur Erfassung der professionellen Handlungskompetenz von (angehenden) Physiklehrkräften. *Zusammenhangsanalysen zwischen Lehrerkompetenz und Lehrerperformanz*
ISBN 978-3-8325-3846-0 50.50 EUR
- 175 Ingo Brebeck: Selbstreguliertes Lernen in der Studieneingangsphase im Fach Chemie
ISBN 978-3-8325-3859-0 37.00 EUR
- 176 Axel Eghtessad: Merkmale und Strukturen von Professionalisierungsprozessen in der ersten und zweiten Phase der Chemielehrerbildung. *Eine empirisch-qualitative Studie mit niedersächsischen Fachleiter_innen der Sekundarstufenlehrämter*
ISBN 978-3-8325-3861-3 45.00 EUR
- 177 Andreas Nehring: Wissenschaftliche Denk- und Arbeitsweisen im Fach Chemie. Eine kompetenzorientierte Modell- und Testentwicklung für den Bereich der Erkenntnisgewinnung
ISBN 978-3-8325-3872-9 39.50 EUR
- 178 Maïke Schmidt: Professionswissen von Sachunterrichtslehrkräften. Zusammenhangsanalyse zur Wirkung von Ausbildungshintergrund und Unterrichtserfahrung auf das fachspezifische Professionswissen im Unterrichtsinhalt „Verbrennung“
ISBN 978-3-8325-3907-8 38.50 EUR
- 179 Jan Winkelmann: Auswirkungen auf den Fachwissenszuwachs und auf affektive Schülermerkmale durch Schüler- und Demonstrationsexperimente im Physikunterricht
ISBN 978-3-8325-3915-3 41.00 EUR

- 180 Iwen Kobow: Entwicklung und Validierung eines Testinstrumentes zur Erfassung der Kommunikationskompetenz im Fach Chemie
ISBN 978-3-8325-3927-6 34.50 EUR
- 181 Yvonne Gramzow: Fachdidaktisches Wissen von Lehramtsstudierenden im Fach Physik. Modellierung und Testkonstruktion
ISBN 978-3-8325-3931-3 42.50 EUR
- 182 Evelin Schröter: Entwicklung der Kompetenzerwartung durch Lösen physikalischer Aufgaben einer multimedialen Lernumgebung
ISBN 978-3-8325-3975-7 54.50 EUR
- 183 Inga Kallweit: Effektivität des Einsatzes von Selbsteinschätzungsbögen im Chemieunterricht der Sekundarstufe I. *Individuelle Förderung durch selbstreguliertes Lernen*
ISBN 978-3-8325-3965-8 44.00 EUR
- 184 Andrea Schumacher: Paving the way towards authentic chemistry teaching. *A contribution to teachers' professional development*
ISBN 978-3-8325-3976-4 48.50 EUR
- 185 David Woitkowski: Fachliches Wissen Physik in der Hochschulausbildung. *Konzeptualisierung, Messung, Niveaubildung*
ISBN 978-3-8325-3988-7 53.00 EUR
- 186 Marianne Korner: Cross-Age Peer Tutoring in Physik. *Evaluation einer Unterrichtsmethode*
ISBN 978-3-8325-3979-5 38.50 EUR
- 187 Simone Nakoinz: Untersuchung zur Verknüpfung submikroskopischer und makroskopischer Konzepte im Fach Chemie
ISBN 978-3-8325-4057-9 38.50 EUR
- 188 Sandra Anus: Evaluation individueller Förderung im Chemieunterricht. *Adaptivität von Lerninhalten an das Vorwissen von Lernenden am Beispiel des Basiskonzeptes Chemische Reaktion*
ISBN 978-3-8325-4059-3 43.50 EUR
- 189 Thomas Roßbegalle: Fachdidaktische Entwicklungsforschung zum besseren Verständnis atmosphärischer Phänomene. *Treibhauseffekt, saurer Regen und stratosphärischer Ozonabbau als Kontexte zur Vermittlung von Basiskonzepten der Chemie*
ISBN 978-3-8325-4059-3 45.50 EUR
- 190 Kathrin Steckenmesser-Sander: Gemeinsamkeiten und Unterschiede physikbezogener Handlungs-, Denk- und Lernprozesse von Mädchen und Jungen
ISBN 978-3-8325-4066-1 38.50 EUR

- 191 Cornelia Geller: Lernprozessorientierte Sequenzierung des Physikunterrichts im Zusammenhang mit Fachwissenserwerb. *Eine Videostudie in Finnland, Deutschland und der Schweiz*
ISBN 978-3-8325-4082-1 35.50 EUR
- 192 Jan Hofmann: Untersuchung des Kompetenzaufbaus von Physiklehrkräften während einer Fortbildungsmaßnahme
ISBN 978-3-8325-4104-0 38.50 EUR
- 193 Andreas Dickhäuser: Chemiespezifischer Humor. *Theoriebildung, Materialentwicklung, Evaluation*
ISBN 978-3-8325-4108-8 37.00 EUR
- 194 Stefan Korte: Die Grenzen der Naturwissenschaft als Thema des Physikunterrichts
ISBN 978-3-8325-4112-5 57.50 EUR
- 195 Carolin Hülsmann: Kurswahlmotive im Fach Chemie. Eine Studie zum Wahlverhalten und Erfolg von Schülerinnen und Schülern in der gymnasialen Oberstufe
ISBN 978-3-8325-4144-6 49.00 EUR
- 196 Caroline Körbs: Mindeststandards im Fach Chemie am Ende der Pflichtschulzeit
ISBN 978-3-8325-4148-4 34.00 EUR
- 197 Andreas Vorholzer: Wie lassen sich Kompetenzen des experimentellen Denkens und Arbeitens fördern? *Eine empirische Untersuchung der Wirkung eines expliziten und eines impliziten Instruktionsansatzes*
ISBN 978-3-8325-4194-1 37.50 EUR
- 198 Anna Katharina Schmitt: Entwicklung und Evaluation einer Chemielehrerfortbildung zum Kompetenzbereich Erkenntnisgewinnung
ISBN 978-3-8325-4228-3 39.50 EUR
- 199 Christian Maurer: Strukturierung von Lehr-Lern-Sequenzen
ISBN 978-3-8325-4247-4 36.50 EUR
- 201 Simon Zander: Lehrerfortbildung zu Basismodellen und Zusammenhänge zum Fachwissen
ISBN 978-3-8325-4248-1 35.00 EUR
- 202 Kerstin Arndt: Experimentierkompetenz erfassen. *Analyse von Prozessen und Mustern am Beispiel von Lehramtsstudierenden der Chemie*
ISBN 978-3-8325-4266-5 45.00 EUR
- 203 Christian Lang: Kompetenzorientierung im Rahmen experimentalchemischer Praktika
ISBN 978-3-8325-4268-9 42.50 EUR
- 204 Eva Cauet: Testen wir relevantes Wissen? *Zusammenhang zwischen dem Professionswissen von Physiklehrkräften und gutem und erfolgreichem Unterrichten*
ISBN 978-3-8325-4276-4 39.50 EUR

- 205 Patrick Löffler: Modell-anwendung in Problemlöseaufgaben. *Wie wirkt Kontext?*
ISBN 978-3-8325-4303-7 35.00 EUR
- 206 Carina Gehlen: Kompetenzstruktur naturwissenschaftlicher Erkenntnisgewinnung
im Fach Chemie
ISBN 978-3-8325-4318-1 43.00 EUR
- 208 Jennifer Petersen: Zum Einfluss des Merkmals Humor auf die Gesundheitsförderung
im Chemieunterricht der Sekundarstufe I. *Eine Interventionsstudie zum Thema Sonnenschutz*
ISBN 978-3-8325-4348-8 40.00 EUR
- 209 Philipp Straube: Modellierung und Erfassung von Kompetenzen naturwissenschaftlicher
Erkenntnisgewinnung bei (Lehramts-) Studierenden im Fach Physik
ISBN 978-3-8325-4351-8 35.50 EUR
- 210 Martin Dickmann: Messung von Experimentierfähigkeiten. *Validierungsstudien zur
Qualität eines computerbasierten Testverfahrens*
ISBN 978-3-8325-4356-3 41.00 EUR

In Vorbereitung

- 207 Lars Oettinghaus Lehrerüberzeugungen und physikbezogenes Professionswissen
ISBN 978-3-8325-4319-8 38.50 EUR

Alle erschienenen Bücher können unter der angegebenen ISBN direkt online (<http://www.logos-verlag.de>) oder per Fax (030 - 42 85 10 92) beim Logos Verlag Berlin bestellt werden.

Studien zum Physik- und Chemielernen

Herausgegeben von Hans Niedderer, Helmut Fischler und Elke Sumfleth

Die Reihe umfasst inzwischen eine große Zahl von wissenschaftlichen Arbeiten aus vielen Arbeitsgruppen der Physik- und Chemiedidaktik und zeichnet damit ein gültiges Bild der empirischen physik- und chemiedidaktischen Forschung in Deutschland.

Die Herausgeber laden daher Interessenten zu neuen Beiträgen ein und bitten sie, sich im Bedarfsfall an den Logos-Verlag oder an ein Mitglied des Herausgeberteams zu wenden.

Kontaktadressen:

Prof. Dr. Hans Niedderer
Institut für Didaktik der Naturwissenschaften,
Abt. Physikdidaktik, FB Physik/Elektrotechnik,
Universität Bremen,
Postfach 33 04 40, 28334 Bremen
Tel. 0421-218 2484/4695, e-mail:
niedderer@physik.uni-bremen.de

Prof. Dr. Helmut Fischler
Didaktik der Physik, FB Physik, Freie Universität Berlin,
Arnimallee 14, 14195 Berlin
Tel. 030-838 56712/55966, e-mail:
fischler@physik.fu-berlin.de

Prof. Dr. Elke Sumfleth
Didaktik der Chemie,
Fachbereich Chemie,
Universität Duisburg-Essen,
Schützenbahn 70, 45127 Essen
Tel. 0201-183 3757/3761, e-mail:
elke.sumfleth@uni-essen.de

Für Studiengänge an deutschen Hochschulen wird heute meistens eine Orientierung an Kompetenzen eingefordert. Die zur Messung dieser Kompetenzen nötigen Testinstrumente fehlen aber weitestgehend. Im Kooperationsprojekt „Ko-WADiS“, in dem diese Studie angesiedelt war, wurde ein Kompetenztest zur naturwissenschaftlichen Erkenntnisgewinnung entwickelt. Erkenntnisgewinnung wird dabei als ein komplexer Problemlöseprozess im Sinne des Scientific Reasoning angesehen. Auf Grundlage dieser theoretischen Überlegungen wurden zunächst ein Kompetenzmodell adaptiert und Items entwickelt. Für diese Studie wurde der Kompetenztest in einer querschnittlichen Untersuchung bei Studierenden der Physik eingesetzt.

Die empirischen Ergebnisse deuten auf die beste Passung eines eindimensionalen Modells hin. Auf Basis dieses Modells wurden weitere Analysen vorgenommen. Hierbei wird deutlich, dass erwartungskonform Masterstudierende einen höheren Kompetenzstand als Bachelorstudierende haben. Gleichzeitig gibt es Evidenz für einen signifikanten Vorteil der Fachstudierenden im Vergleich zu den Lehramtsstudierenden. Keine Vorteile zeigen sich hingegen für Lehramtsstudierende der Physik mit einem weiteren naturwissenschaftlichen Fach.

Logos Verlag Berlin

ISBN 978-3-8325-4351-8