

Rehfeldt, Daniel

Erfassung der Lehrqualität naturwissenschaftlicher Experimentalpraktika

Berlin : Logos Verlag 2017, 295 S. - (Studien zum Physik- und Chemielernen; 246) - (Dissertation, Freie Universität Berlin, 2017)



Quellenangabe/ Reference:

Rehfeldt, Daniel: Erfassung der Lehrqualität naturwissenschaftlicher Experimentalpraktika. Berlin : Logos Verlag 2017, 295 S. - (Studien zum Physik- und Chemielernen; 246) - (Dissertation, Freie Universität Berlin, 2017) - URN: urn:nbn:de:0111-pedocs-333770 - DOI: 10.25656/01:33377; 10.5281/zenodo.1183547

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-333770>

<https://doi.org/10.25656/01:33377>

in Kooperation mit / in cooperation with:



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

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz: <http://creativecommons.org/licenses/by-nc-nd/4.0/deed.de> - Sie dürfen das Werk bzw. den Inhalt unter folgenden Bedingungen vervielfältigen, verbreiten und öffentlich zugänglich machen: Sie müssen den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen. Dieses Werk bzw. dieser Inhalt darf nicht für kommerzielle Zwecke verwendet werden und es darf nicht bearbeitet, abgewandelt oder in anderer Weise verändert werden.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-License: <http://creativecommons.org/licenses/by-nc-nd/4.0/deed.en> - You may copy, distribute and transmit, adapt or exhibit the work in the public as long as you attribute the work in the manner specified by the author or licensor. You are not allowed to make commercial use of the work or its contents. You are not allowed to alter, transform, or change this work in any other way.

By using this particular document, you accept the above-stated conditions of use.



Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Studien zum Physik- und Chemielernen

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

246

Daniel Rehfeldt

Erfassung der Lehrqualität naturwissenschaftlicher Experimentalpraktika



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

Studien zum Physik- und Chemielernen

Band 246

Daniel Rehfeldt

Erfassung der Lehrqualität
naturwissenschaftlicher Experimentalpraktika

Logos Verlag Berlin



Studien zum Physik- und Chemielernen

Hans Niedderer, Helmut Fischler, Elke Sumfleth [Hrsg.]

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 2017

Alle Rechte vorbehalten.

ISBN 978-3-8325-4590-1



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

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

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

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

Dissertation
zur Erlangung des Grades
eines Doktors der Naturwissenschaften
am Fachbereich Physik der
Freien Universität Berlin

**Erfassung der Lehrqualität
naturwissenschaftlicher
Experimentalpraktika**

vorgelegt von
Daniel Rehfeldt, M. Ed.
aus
Berlin

Berlin, 2017

Erster Gutachter: Prof. Dr. Volkhard Nordmeier
Zweiter Gutachter: Prof. Dr. Rainer Watermann

Aus dieser Dissertation hervorgegangene Veröffentlichungen

Erstautorenschaften

- **REHFELDT, D.** & **NORDMEIER, V.** (2016). Theoretically and empirically based Evaluation of laboratory courses – PraQ questionnaire. In E. Dębowska & T. Greczyło (Hrsg.), *GIREP-MPTL International Conference on Teaching/Learning Physics: Key Competences in Physics Teaching and Learning*.
- **REHFELDT, D.** & **NORDMEIER, V.** (2016). Skalen zur Messung von Praktikumsqualität: Konfirmatorische Analyse der Struktur und Konstrukte. *PhyDid B – Didaktik der Physik: Beiträge zur DPG-Frühjahrstagung*, Berlin.
- **REHFELDT, D.**, **MÜHLENBRUCH, T.** & **NORDMEIER, V.** (2017). Praktikumsqualität messen – Validierungsstudie. In J. Stiller & C. Laschke (Hrsg.), *Berlin-Brandenburger Beiträge zur Bildungsforschung 2016*. Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien, Peter Lang.
- **REHFELDT, D.**, **MÜHLENBRUCH, T.** & **NORDMEIER, V.** (2016). Fragebogen zur Erfassung von Praktikumsqualität (PraQ): Faktorielle Validierung. In S. Bernholt (Hrsg.), *Authentizität und Lernen - das Fach in der Fachdidaktik: Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung 2015*. Kiel, IPN.
- **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: Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung 2014* (S. 417–419). Kiel, IPN.
- **REHFELDT, D.**, **BRÜGGEMANN, V.** & **NORDMEIER, V.** (2015). Der PraQ: Ein Instrument zur Einschätzung von Praktikumsqualität – Skriptqualität. *PhyDid B – Didaktik der Physik: Beiträge zur DPG-Frühjahrstagung*, Berlin.
- **REHFELDT, D.**, **MÜHLENBRUCH, T.** & **NORDMEIER, V.** (2015). TSL: Ergebnisse einer quantitativen Studie zur Problemanalyse im Physikalischen Praktikum für Naturwissenschaftler. In J. Stiller & C. Laschke (Hrsg.), *Berlin-Brandenburger Beiträge zur Bildungsforschung 2015: Herausforderungen, Befunde und Perspektiven interdisziplinärer Bildungsforschung* (S. 203–228). Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien, Peter Lang.
- **REHFELDT, D.**, **GUTZLER, T.** & **NORDMEIER, V.** (2014). TSL: Quantitative Problemanalyse im Nebenfachpraktikum. In S. Bernholt (Hrsg.), *Naturwissenschaftliche Bildung zwischen Science- und Fachunterricht: Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung 2013*. Münster, LIT.
- **REHFELDT, D.**, **MÜHLENBRUCH, T.** & **NORDMEIER, V.** (2014). TSL: Quantitative Bedarfs- und Problemanalyse im Physikalischen Praktikum für NaturwissenschaftlerInnen. *PhyDid B – Didaktik der Physik: Beiträge zur DPG-Frühjahrstagung*, Berlin.

- **REHFELDT, D.**, GUTZLER, T. & NORDMEIER, V. (2013). TSL – Technology SUPPORTed Labs: Multimediale Unterstützung naturwissenschaftlicher Hochschulpraktika. *PhyDid B – Didaktik der Physik: Beiträge zur DPG-Frühjahrstagung*, Berlin.

Weitere Autorenschaften¹

- SCHILD, N., **REHFELDT, D.** & NORDMEIER, V. (2016). Mögliche Prädiktoren für den Studienerfolg im Lehramt und ihm Fach Physik. *PhyDid B – Didaktik der Physik: Beiträge zur DPG-Frühjahrstagung*, Berlin.
- SCHILD, N., **REHFELDT, D.** & NORDMEIER, V. (2016). Nicht-kognitive Prädiktoren für den Studienerfolg im Lehramt und im Fach Physik. In S. Bernholt (Hrsg.), *Authentizität und Lernen - das Fach in der Fachdidaktik: Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung 2015*. Kiel, IPN.
- MÜHLENBRUCH, T., **REHFELDT, D.** & NORDMEIER, V. (2015). TSL: Bedarfsanalyse im physikalischen Praktikum für Naturwissenschaftler: GRAFCET: Ein »neues« Werkzeug zur Strukturierung von Lehrveranstaltungen. In J. Stiller & C. Laschke (Hrsg.), *Berlin-Brandenburger Beiträge zur Bildungsforschung 2015: Herausforderungen, Befunde und Perspektiven interdisziplinärer Bildungsforschung* (S. 229–259). Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien, Peter Lang.
- SCHILD, N., HEINDEL, L., STRAUBE, P., **REHFELDT, D.** & NORDMEIER, V. (2015). Vorhersagemodell zum Studienerfolg im Fach und im Lehramt Physik: Erste Schritte zur Entwicklung eines Online-Self-Assessment. *PhyDid B – Didaktik der Physik: Beiträge zur DPG-Frühjahrstagung*, Berlin.
- HEINDEL, L., SCHILD, N., **REHFELDT, D.** & NORDMEIER, V. (2015). Entwicklung eines Online-Tools zur Studienfachwahl Physik/Lehramt Physik. *PhyDid B – Didaktik der Physik: Beiträge zur DPG-Frühjahrstagung*, Berlin.
- GUTZLER, T., **REHFELDT, D.** & NORDMEIER, V. (2014). Technology SUPPORTed Labs (TSL) – multimedial ergänztes Lernen im Praktikum. In N. Apostolopoulos, H. Hoffmann, U. Mußmann, W. Coy & A. Schwill (Hrsg.), *Grundfragen Multimedialen Lehrens und Lernens. Der Qualitätspakt E-Learning im Hochschulpakt 2020*. (Tagungsband), S. 224–236.
- GUTZLER, T., **REHFELDT, D.** & NORDMEIER, V. (2014). TSL: Bedarfsanalyse in Praktika: Ein »neues« Werkzeug zur Strukturierung. In S. Bernholt (Hrsg.), *Naturwissenschaftliche Bildung zwischen Science- und Fachunterricht: Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung 2013*. Münster, LIT.
- MÜHLENBRUCH, T., **REHFELDT, D.** & NORDMEIER, V. (2014). TSL: Interventionsgestaltung im Nebenfachpraktikum. *PhyDid B – Didaktik der Physik: Beiträge zur DPG-Frühjahrstagung*, Berlin.

¹ grau: Ohne Bezug zum Promotionsthema

Lebenslauf

Daniel Rehfeldt, geb. 1987 in Berlin.

2007–2010 Bachelor-Studium der Physik/Mathematik mit Lehramtsoption an der Humboldt Universität zu Berlin

Abschlussarbeit: *Eine dynamische Fresnel'sche Zonenplatte: Optische Abbildungen mit einem LCD* (Note 1,3)

Abschluss: Bachelor of Science (Note 1,3)

2010–2012 Master-Studium der Physik/Mathematik mit Lehramtsoption an der Humboldt Universität zu Berlin

Abschlussarbeit: *Assoziationen zur Physik – Die Erprobung impliziter Testverfahren als Instrumente in der physikdidaktischen Forschung* (Note 1,0)

Abschluss: Master of Education (Note 1,2)

2010–2012 Studentischer Mitarbeiter in der AG Didaktik der Physik (Prof. Schön).

2012–2016 Wissenschaftlicher Mitarbeiter und Promotionsstudent in der AG Didaktik der Physik (Prof. Nordmeier) an der Freien Universität Berlin im Projekt SUPPORT: Technology SUPPORTed Labs.

2012–2016 Lehrbeauftragter an der Freien Universität Berlin, Didaktik der Physik

Master-Seminare: Methoden fachdidaktischer Bildungsforschung, Hauptseminar Physikdidaktik, Lehr-Lern-Labor-Seminar Klimawandel

Bachelor-Seminare: Wissenschaftliches Schreiben, Einführung in die Fachdidaktik Physik, Schulorientiertes Experimentieren, Freies Experimentieren

Seit 2016 Projektkoordinator K2teach Teilprojekt 3: Erprobung von Handlungsstrategien in Lehr-Lern-Laboren.

Forschungsinteressen: Hochschul-Experimentalpraktika, Lehr-Lern-Labor-Seminare, Studienerfolg, Lehrkräftebildung, Quantitative Forschung.

Kurzfassung

Theoretischer Hintergrund: In naturwissenschaftlichen Experimentalpraktika an deutschen Hochschulen wurden bisher teils mit geringem Theoriebezug und wenig generalisierbaren Methoden Befunde erarbeitet, die ein problematisches Verhältnis der intendierten Ziele von Praktika und der tatsächlichen Erreichung derselben betreffen. Dies gilt recht unabhängig von Fachrichtung und Studierendenschaft. Zugleich demonstriert die in Teilen sehr elaborierte Lehrevaluationslandschaft, wie sowohl theoriegeleitet, als auch generalisierbar Befunde zur Lehrqualität von Vorlesungen und Seminaren gewonnen werden können. Praktika wurden indes nicht berücksichtigt. Aus den Erkenntnissen der Praktikumsforschung und den Anforderungen guter Lehrevaluation erwuchs so das in dieser Arbeit vorgestellte theoretische, fächerübergreifende *3L-Modell* der Praktikumsqualität, das zur Konstruktion des Praktikumsqualitätsfragebogens *PraQ* führte. Dieser quantifiziert die Qualität von Praktika in den Dimensionen *Lernzuwachs*, *Lehrkompetenz* und *Lernumgebung* mit je zugeordneten Qualitätskonstrukten.

Methode: Der PraQ-Fragebogen wurde als Rating-Evaluations-Fragebogen für Studierende in sechs Testevaluationsstudien in insgesamt 18 Praktika im deutschsprachigen Raum erprobt ($N = 1314$). Die Studien umfassten Expert*inneninterviews, explorative und konfirmatorische Faktorenanalysen, konvergente und diskriminante Vergleiche mit etablierten Instrumenten *experimenteller Kompetenzdiagnostik* und mit dem *Sozialklima* im Praktikum. Zusätzlich wurde das Antwortverhalten von Studierenden mit Expert*innenantworten verglichen. Die umfangreiche Stichprobe erlaubte Forschungsfragen zu einem deutschlandweiten Qualitätsbenchmarking von Praktika und zu Unterschieden zwischen Chemie- und Physik-, sowie Block- und Semesterpraktika.

Ergebnis: Die Testevaluationsstudien zeigten empirische Evidenz für eine gelungene inhaltliche Breite der Konstrukte des PraQ. Strukturelle Analysen zeigten nur marginal von der Theorie abweichende Faktorstrukturen und solide statistische Modelle. Konvergente Korrelationen der PraQ-Skalen mit einem Kompetenztest fielen nichtsignifikant, mit *Selbstwirksamkeitserwartungsskalen* mittelstark aus. diskriminante Vergleiche mit dem *Praktikumsklima* zeigten meist schwache Korrelationen. Multiperspektivische Abgleiche von Studierenden- und Fachdidaktiker*innenbewertungen zur *Skriptqualität* gelangen gut. Die Analyse inhaltlicher Forschungsfragen brachte hervor, dass ein Qualitätsbenchmarking für Praktika mittels PraQ gelingt, das *Fachwissen* wird am stärksten, *Experimentelle Kompetenz* mittelmäßig gefördert, das *Interesse* vernachlässigt. Physikpraktika wurden in einigen Bereichen der *Experimentellen Kompetenz* und der *Lehrkompetenz* der Betreuenden besser bewertet als Chemiepraktika. Semesterpraktika wiesen zwar eine deutlich bessere Betreuungsqualität auf, allerdings bewirkten sie teils geringere Kompetenzzuwächse, z. B. im *Versuchsanordnung aufbauen* und *Zeitmanagement*.

Diskussion: Der PraQ-Fragebogen schließt die Forschungslücke eines theoriebasierten und interpretationsvaliden Evaluationsinstruments. Diverse Interpretationsrichtungen sind nunmehr zulässig und Studien im deutschsprachigen Raum konnten den gewinnbringenden Einsatz des Instruments untermauern. Befunde zum Fokus auf *Fachwissen* bei gleichzeitig geringem *Interessenzuwachs* der Studierenden erscheinen ungünstig für die Ziele von Praktika, das Kerngeschäft des Experimentierens wird nur mittelmäßig gefördert.

Die jeweiligen Stärken der Praktikumsformen können in der Zukunft genauer beobachtet und zielgerichtete Interventionen für schwächere Praktika daraus abgeleitet werden.

Abstract

Background: Studies concerning scientific hands-on laboratory courses in German universities have thus far been carried out with little theoretical fundament and few generalizable methods. Independently from the field of study and the student corps, this leads to a problematic relationship between the intended aims of the coursework and the actual achievement of these aims.

At the same time professional evaluation of different university teaching formats demonstrates how the quality of teaching can be quantified on a theoretical as well as generalizable basis, yet laboratory course work has not been included.

Based on recent findings about the quality of hands-on-trainings and the requirements for good evaluation, the interdisciplinary 3L model for quality of laboratory courses was established in this thesis.

Furthermore, a questionnaire called PraQ was developed to quantify quality of hands-on-trainings based on the dimensions *gain of knowledge*, *teaching skills* and *learning environment*.

Method: The PraQ was field tested as a rating-questionnaire for students in six evaluation studies at 18 independent laboratory courses within Germany and Austria (N = 1314). Studies included expert interviews, exploratory/confirmatory factor analysis and convergent as well as discriminant comparisons with established instruments for experimental competence and social climate.

Furthermore, the answers of examined students were compared to those of experts. The extensive sample size made it possible to examine both a quality benchmarking for German laboratory courses and to shed light on the differences between chemistry and physics labs as well as between block and semester courses.

Results: Evaluation studies showed empirical evidence for sufficient content validity of the PraQ's constructs. Structural analysis showed factor structures marginally different from theory and solid statistical models. Convergent correlations of PraQ's scales with tests of experimental competence were not significant, correlations with scales of self-efficacy were intermediate. Discriminant comparisons with social climate of laboratory courses showed mostly weak correlations. Multiperspective alignments of student's and teacher's rating for the script's quality were excellent. A quality benchmarking via PraQ showed a main focus on theoretical knowledge in laboratory courses while experimental competence is only moderately enhanced and interests of students are neglected. Physics Laboratory courses were rated better than chemistry labs regarding parts of experimental competence as well as competence of assistance. Semester courses showed an increased quality of assistance but a decreased gain of competence, e.g. in preparing experimental settings and time management.

Conclusion: The PraQ questionnaire fills the academic void for a theoretically based evaluation instrument valid for interpretation. Multiple interpretations are valid now and studies within Germany and Austria could show the benefits of this new instrument. Laboratories' focus on theoretical knowledge and critically low interests of students seem contrary to laboratories main goals, since even experimental skills have been shown to be supported only on an intermediate level. Strengths and weaknesses of the different forms of laboratory courses can be examined in more detail using the PraQ questionnaire allowing for interventions to be tailored more specifically for the weaknesses of the correspondent courses.

Inhaltsverzeichnis

I. Theoretischer Hintergrund	13
1. Einleitung	15
2. Lehre an Hochschulen	19
2.1. Hochschullehre nach Bologna	19
2.2. Qualität der Lehre	20
3. Lehrformen in naturwissenschaftlichen Studiengängen	25
3.1. Vorlesung	25
3.2. Seminar	26
3.3. Übung	27
3.4. Praktikum	28
3.4.1. Praktika gemäß Bologna-Reform und Studienordnungen . .	29
3.4.2. Ziele und Legitimation von Praktika	30
3.4.3. Typische Gestaltung und Eigenschaften von Praktika . . .	33
3.4.4. Typen von Praktika	35
4. Lehrevaluation an Hochschulen	39
4.1. Ziele von Lehrevaluationen	39
4.2. Methodik(en) der Lehrevaluation	41
4.3. Strukturelle Verankerung von Lehrevaluationen	41
4.4. Anforderungen an Evaluationsinstrumente und -auswertungen . .	43
4.5. Standard-Evaluationsinstrumente für die Hochschullehre	48
4.5.1. Evaluation des Lehrens (SEEQ und LeKo)	49
4.5.2. Evaluation des Lernens (BEvaKomp)	51
4.5.3. Evaluation der Struktur einer Lehrveranstaltung (CEQ) . .	52
4.5.4. Evaluationsinstrument für Praktika (SLEI)	53
5. Forschungsstand: Lehrformat Naturwissenschaftliche Praktika	55
5.1. Charakteristika von Praktika	56
5.2. Probleme von Praktika	57
5.2.1. Phasenübergreifende Probleme	57
5.2.2. Probleme in der Vorbereitung	59

5.2.3.	Probleme in der Durchführung	59
5.2.4.	Probleme in der Nachbereitung	60
5.2.5.	Exkurs: Nebenfachpraktika	60
5.3.	Empirische Interventions-Studien zu Praktika (Auswahl)	61
5.4.	Zusammenfassung	65
6.	Forschungsstand: Theoretische Modelle zu Praktika	67
6.1.	Wirksamkeitsmodell nach PSILLOS & NIEDDERER (2002)	67
6.2.	Praktikums-Lehrqualität nach LAGOWSKI (2002)	69
6.3.	Kompetenzmodell nach KREITEN (2012)	73
6.4.	Zusammenfassung	73
6.5.	Forschungslücke: Lehrevaluation von Praktika	77
7.	Strukturierung von Praktikumsqualität	81
7.1.	Vorgehen	83
7.2.	Lernzuwachs der Studierenden	83
7.3.	Lehrkompetenz bei der Betreuung	88
7.4.	Lernumgebungsqualität	93
7.5.	Zusammenfassung: 3L-Modell der Praktikumsqualität	95
8.	Ziele und Forschungsfragen	99
8.1.	Ziele zum PraQ-Instrument	99
8.2.	Forschungsfragen	100
II.	Empirischer Teil	103
9.	Einleitung	105
10.	Testkonstruktion	107
10.1.	Auswahl der Erhebungsmethode	107
10.2.	Zielgruppe	109
10.3.	Erstellung des Itempools	110
10.3.1.	Fächerübergreifende Formulierung	110
10.3.2.	Adaption etablierter Skalen	111
10.3.3.	Itemzuordnung je Konstrukt	113
10.3.3.1.	Dimension Lernzuwachs	113
10.3.3.2.	Dimension Lehrkompetenz	115
10.3.3.3.	Dimension Lernumgebung	118

10.4. Vorbereitung der Durchführung	119
11. Objektivität	121
12. Testevaluation (Validität und Reliabilität)	123
12.1. Forschungsfragen zur Validierung	123
12.2. Validierungsdesign	126
12.3. Studie 1: Inhaltlicher Umfang des PraQ (»Inhaltsvalidität«)	127
12.3.1. Studie 1.1: Relevanz und Vollständigkeit: PraQ	128
12.3.1.1. Einleitung	128
12.3.1.2. Methode	128
12.3.1.3. Ergebnisse	130
12.3.1.4. Diskussion	133
12.3.2. Studie 1.2 Relevanz und Vollständigkeit: Skriptqualität . .	135
12.3.2.1. Einleitung	135
12.3.2.2. Methode	135
12.3.2.3. Ergebnisse	136
12.3.2.4. Diskussion	137
12.4. Studie 2: Strukturelle Eigenschaften (»faktorielle Validität«) . . .	137
12.4.1. Studie 2.1: Explorative Analyse	138
12.4.1.1. Einleitung	138
12.4.1.2. Methode	139
12.4.1.3. Ergebnisse	143
12.4.1.4. Diskussion	149
12.4.2. Studie 2.2: Konfirmatorische Analyse	153
12.4.2.1. Einleitung	153
12.4.2.2. Methode	154
12.4.2.3. Ergebnisse	156
12.4.2.4. Diskussion	168
12.5. Studie 3: Zusammenhänge mit etablierten Maßen	172
12.5.1. Studie 3.1: Leistungstest und Selbstwirksamkeitserwartung (»konvergent«)	172
12.5.1.1. Einleitung	172
12.5.1.2. Methode	174
12.5.1.3. Ergebnisse	176
12.5.1.4. Diskussion	178

12.5.2. Studie 3.2: Praktikumssozialklima (»diskriminant«)	179
12.5.2.1. Einleitung	179
12.5.2.2. Methode	180
12.5.2.3. Ergebnisse	181
12.5.2.4. Diskussion	181
12.6. Studie 4: Zusammenhang mit Expert*innenratings (»Multiperspektivität«)	183
12.6.1. Einleitung	183
12.6.2. Methode	184
12.6.3. Ergebnisse	185
12.6.4. Diskussion	185
12.7. Zusammenfassung: Testevaluation	187
12.7.1. Evidenz zur Validität und Reliabilität	187
12.7.2. Methodenkritik	188
12.7.3. Generalisierbarkeit der Testevaluation	190
12.7.4. Ausblick: Weitere Testevaluationsschritte	191
13. Studien: Einsatz von PraQ im deutschsprachigen Raum	193
13.1. Studie 5: Stärken und Schwächen von Praktika in Deutschland	194
13.1.1. Einleitung	194
13.1.2. Methode	197
13.1.3. Ergebnisse	199
13.1.4. Diskussion	205
13.2. Studie 6: Unterschiede zwischen Praktika der Chemie und der Physik	207
13.2.1. Einleitung	207
13.2.2. Methode	208
13.2.3. Ergebnisse	211
13.2.4. Diskussion	217
13.3. Studie 7: Unterschiede zwischen Block- und Semesterpraktika	223
13.3.1. Einleitung	223
13.3.2. Methode	225
13.3.3. Ergebnisse	226
13.3.4. Diskussion	232
14. Zusammenfassung und Ausblick	239
14.1. Evaluation von Praktika	239

14.2. Erforschung von Lernzuwachs, Lehrkompetenz und Lernumgebungs- qualität	240
14.3. Erkenntnisse zu Praktika im deutschsprachigen Raum	241
14.4. Ausblick	242
14.4.1. Gewichtung der Konstrukte	244
14.4.2. Handlungsempfehlungen für Praktika	244
14.4.3. Demonstrationspraktika (Lehramt)	246
14.4.4. Fortgeschrittenen-Praktika	247
14.4.5. Projektlabore (offenere Praktikumsformen)	247
Literaturverzeichnis	247
Anhang	267
A. Konstrukte und Items des PraQ – Finale Version	269
A.1. PraQ-A: Lernzuwachs	270
A.2. PraQ-B: Lehrkompetenz des/der Betreuenden & Lernumgebung	275
B. Skalenscores PraQ-A Gesamtstichprobe	279
C. Skalenscores PraQ-B Gesamtstichprobe	281
D. Leitfaden für die kognitiven Expert*inneninterviews der Studie 1.1	283
E. Manuale für Erhebungen mittels PraQ außerhalb Berlins	285
F. Praktikumsskript der Humboldt-Universität zu Berlin: Auszug für Stu- die 4	289
G. Mindestkriterien CFA	291
H. Beispielitem Ko-WADiS-Test zu Auswertung und Interpretation (Strau- be, 2016)	293

I.

Theoretischer Hintergrund

Kapitel 1

Einleitung

Universitäre Lehre steht in seiner Konzeption immer mehr im Zeichen der sogenannten Bologna-Reform (BOSBACH et al., 2007). Für den Lernprozess selbst bedeutet das, dass Lehre an Hochschulen »outputorientiert« sein soll. Lehrende sollen vernetzte Kompetenzen der Studierenden fördern und prüfen, ob diese erreicht wurden, anstatt sich »inputorientiert« nur an ihren scheinbar gänzlich abzuleistenden Themen zu orientieren (BLK, 2002, S. 4).

Das Bild von Lehrenden, die »ihren Stoff durchkriegen« wollen, ist gemäß Reform-Befürwortern nicht mehr aktuell (BOSBACH et al., 2007, S. 240), auch wenn die Lehrrealität an Universitäten vielerorts noch anders aussehen mag.¹ Lehrevaluationen sind daher inzwischen kompetenzorientiert angelegt (LeKo nach THIEL et al., 2012; BeVaKomp nach BRAUN et al., 2008). Dennoch wird speziell die experimentelle Ausbildung in den Naturwissenschaften im Rahmen naturwissenschaftlicher Hochschulpraktika diesem Anspruch oft nicht in Gänze gerecht (SACHER et al., 2015).

Praktika sind seit jeher Bestandteil der naturwissenschaftlichen Ausbildung an deutschen Hochschulen (NEUMANN, 2004, S. 7–9). Sie gelten u. a. als Vermittler zwischen der »(...) 'Welt der Theorien und Modelle' [(Physik)] (...) und der 'Welt der Dinge' [(Technik)]« (HUCKE, 1999, S. 28). Doch gerade Praktika des Grundstudiums scheinen den Studierenden wenig Freiräume zu bieten, ihre Fähigkeiten auszutesten und selbsttätig und -denkend zugleich zu sein (HOFSTEIN & LUNETTA, 2004, S. 32; HUCKE, 1999, S. 83) – die Wirkungen bleiben eher klein.² Praktika zeigen sich darüber hinaus als äußerst innovationsresistent, viele laufen noch wie vor 50 Jahren ab (SACHER et al., 2015, S. 106).

Erkenntnisse über Praktika im deutschsprachigen Raum stammen fast ausschließlich aus Forschungsarbeiten Doktorierender. Die Forschungsansätze sind dabei so vielfältig und divers wie die Praktika selbst. Sie fokussieren zudem fast immer auf

¹Im Gegensatz zur Forschung spielt die Lehre ohnehin oftmals kritischerweise eine eher untergeordnete Rolle (SCHIMANK, 2010, S. 52).

²So ergab eine der wenigen standortübergreifenden Umfragen zu Praktika, dass diese weder ausreichend auf Folgepraktika, noch auf eine experimentelle Abschlussarbeit vorbereiten (RUICKOLDT, 1996, S. 1023)

1. Einleitung

das lokal an der Universität der Forschenden ansässige Praktikum und gehen selten explizit theoriebasiert vor.

REINHOLD (2004, S. 118) hebt aber in einer Ausführung über die Qualität der Lehrkräftebildung an der Hochschule hervor:

Zur Klärung der Fragen gehören normativ geprägte oder theoriegeleitete Entwicklungen und auf Rechtfertigung anlegte Evaluationen ebenso wie qualitative oder quantitative Wirkungsforschung.

REINHOLD (2004, S. 118)

Ziel dieser Arbeit ist es, die Forschungslücken der Theorien für Praktikumsqualität und standardisierter summativer Evaluationsinstrumente für Praktika zu schließen. Dies liefert nicht nur ein Benchmarking (summativ) und eine Entwicklungsgrundlage für Praktika, sondern auch neue Theorien und Instrumente für die physikdidaktische Forschung, etwa neue Skalen zur Selbsteinschätzung experimenteller Kompetenzen.

Der Schwerpunkt der Arbeit liegt also auf der Theorie- und Instrumententwicklung. Zusätzlich erwachsen aus den multiplen Querschnitts-Erhebungen in insgesamt 18 Praktika im deutschsprachigen Raum auch inhaltliche Erkenntnisse über Lernprozesse und -ergebnisse in Praktika. Zur Orientierung in dieser Arbeit dient eine »Leitgrafik«, die sich im Verlauf der Abschnitte 2, 6 und 9 entwickeln wird (Abb. 1.1).

In der Leitgrafik für die Abschnitte 2 bis 5 (Abb. 1.2) wird deutlich, dass zur Erarbeitung eines Qualitätsverständnisses der Lehre in Praktika zunächst ein Blick in die (Qualität von) Lehre an Hochschulen im Allgemeinen und dann im Speziellen für Praktika lohnt. Es wird in Abschnitt 4 zur Vorbereitung der Evaluation von Praktika auf bisherige Lehrevaluation an Hochschulen nebst Qualitätsansprüchen an selbige eingegangen.

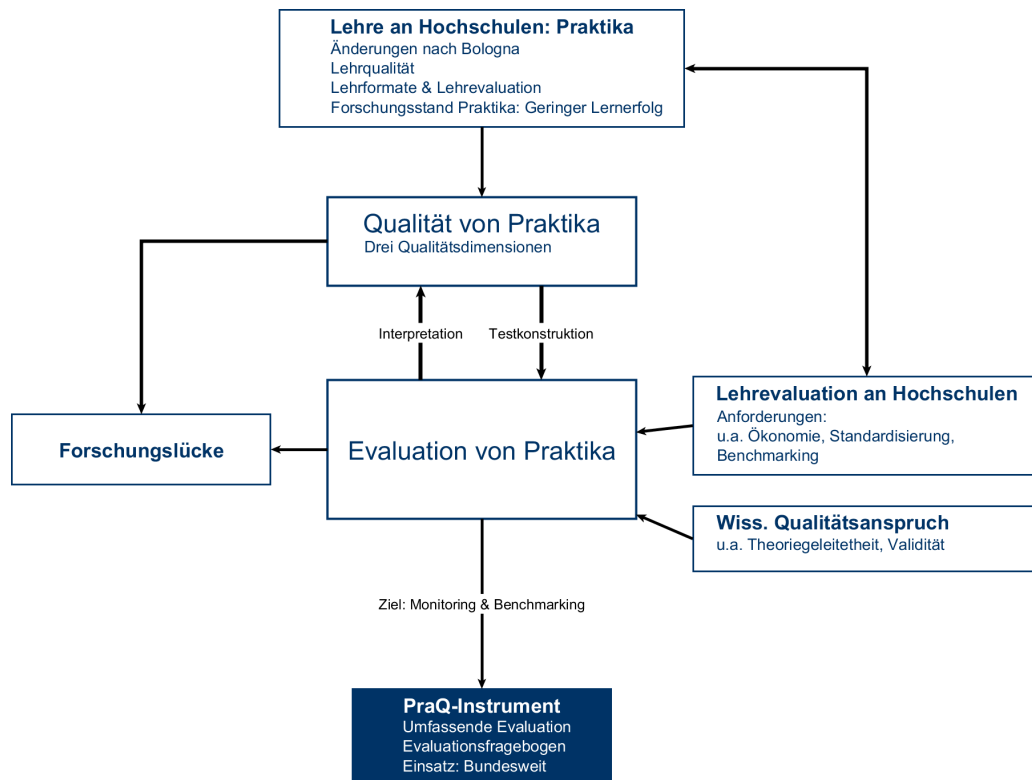


Abbildung 1.1.: Leitgrafik für diese Arbeit: Die Zusammenhänge der einzelnen Kapitel und Abschnitte werden noch stärker zu Beginn eines jeden Teils verdeutlicht. Die oberen beiden Bereiche »Lehre an Hochschulen: Praktika« und »Qualität von Praktika« bilden hauptsächlich den Theorieteil ab, in dem die Herleitung eines theoretischen Modells der Qualität von Praktika vorgenommen wird. In der Mitte der Grafik wird über Bezüge zu Standards der »Lehrevaluation an Hochschulen« und des »wissenschaftlichen Qualitätsanspruchs« an Messungen die »Forschungslücke« der »Evaluation von Praktika« verdeutlicht. Für den unteren Bereich des »PraQ-Instruments« werden hauptsächlich Instrumententwicklung und Testevaluation (»Validierung«), nebst inhaltlicher Forschungsergebnisse zum Lernen in Praktika.

1. Einleitung

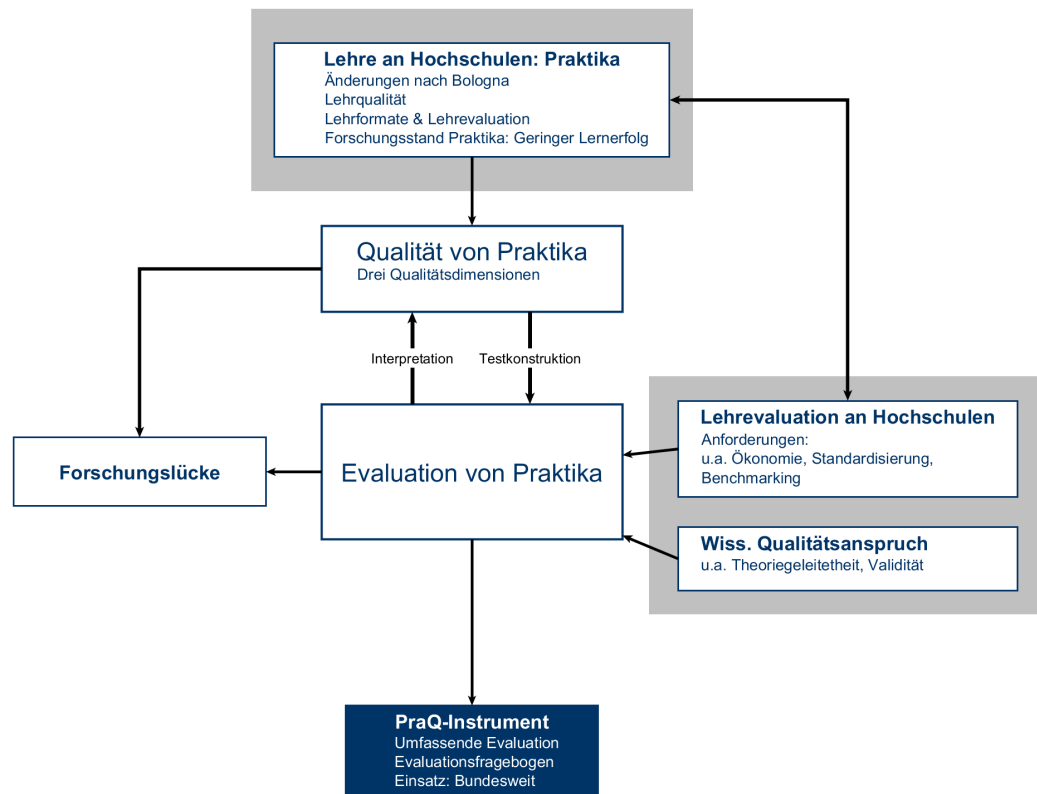


Abbildung 1.2.: Leitgrafik für die Abschnitte 2 bis 6: Die (bisherige) theoretische und gänzliche strukturelle Rahmung der Arbeit geschieht in diesen Abschnitten. Vor allem soll das Wechselspiel der »Lehre an Hochschulen im Bereich Praktika« und der »Lehrevaluation an Hochschulen« herausgestellt werden, nebst einem kritischen Blick auf die »wissenschaftlichen Qualitätsansprüche« an Evaluationserhebungen und -auswertungen.

Kapitel 2

Lehre an Hochschulen

2.1. Hochschullehre nach Bologna

Mit der Bologna-Reform wurden einige tiefgreifende Veränderungen in der Struktur und auch in den Programmen von europäischen Studiengängen erwirkt. Zentrales Ziel war ein gemeinsamer europäischer Hochschulraum, mit vergleichbaren Abschlüssen, einem einheitlichen Studiensystem und damit einer Verbesserung im globalen Wettbewerb inklusive der Erhöhung von Studienflexibilität (BOSBACH et al., 2007, S. 166) und Erhöhung der Berufsrelevanz der Studiengänge für die Studierenden (HOPBACH, 2005b, S. 15).

Dazu gehörte die flächendeckende Einführung eines (modularisierten) zweigliedrigen Studienabschlusssystems (Bachelor- und Masterstudiengänge), die sowohl eine Flexibilisierung in Bezug auf die gewünschte Studiendauer (drei bis vier bzw. fünf Jahre), als auch einen einheitlichen Qualitätsrahmen umfasste (ENQA, 2009). Hinzu kommt die übergreifende Regelung, studentischen Workload in ECTS-Punkten auszudrücken (BOSBACH et al., 2007, S. 169) und diesen für die Bachelor- und Masterstudiengänge zu vereinheitlichen. Über diese so genannte Makromodularisierung (BOSBACH et al., 2007, S. 166) in Bachelor und Masterprogramme hinaus wird die Mesomodularisierung (»Fächer«) im Wesentlichen beibehalten.

Bei Neu- oder Umstrukturierungen von Studiengängen soll aber der Lernprozess der Studierenden im Fokus stehen und nicht traditionelle bestimmte Fachinhalte. Hieraus bereits in der Mesoebene erkennbar ist die Output-Orientierung der Lehre an Hochschulen, die statt der bisherig oftmals vorherrschenden content- (BOSBACH et al., 2007, S. 172) oder Input-Orientierung etabliert werden soll. Dies führt dazu, dass heute die Kompetenzziele bereits in einzelnen Modulbeschreibungen festgelegt werden und damit das gesamte Studium kompetenzorientiert strukturiert wird (ebd., S. 169). Dies stellt neben den Fachinhalten vor allem die Vernetzung von methodischem Repertoire des Faches, Anwendungsbeispielen (problemorientiert, transferfähig) und dem Selbstmanagement in den Vordergrund (ebd., S. 171–172), denn nur im Zusammenspiel dieser wird erfolgreiches Handeln möglich.

Es wird an exemplarischen, besonders gut zur Verzahnung dieser Bereiche geeigneten Fachinhalten gelehrt, die Reflexion des Lernprozesses wird betont (ebd., S. 174). Dies stellt ein zentrales und innovatives Element der Bologna-Reform dar, die

2. Lehre an Hochschulen

sich vor allem in der Mikroebene, den Studienmodulen, zeigt. Diese bestehen aus verschiedenen Lehr- und Lernkomponenten, die sowohl inhaltlich, als auch in Bezug auf die von den Studierenden zu erwerbenden Kompetenzen eine Einheit bilden (ebd., S. 167) und aufeinander aufbauende Bestandteile enthalten (BLÜTHMANN, 2012, S. 13). Da diese Module über ein Semester, aber auch über mehrere Semester verlaufen können, offenbart sich hier ein »Experimentierraum« für Reformen in der Hochschullehre. Dieser ist vor allem am Kompetenzzuwachs der Studierenden orientiert, also denjenigen Fähigkeiten und Fertigkeiten, die zu Ende eines Moduls erlangt worden sein sollen (für eine genauere Definition des Kompetenzbegriffs, siehe z. B. WEINERT, 2001, S. 27).

Allerdings praktizieren die Universitäten in Deutschland eine weitgehend selbstständige Implementierung z. B. jener Output-Orientierung. Daher ist diese Orientierung nicht überall in gleichem Maße vertreten, wie sich schon an der Beschreibung aus den Studien- und Prüfungsordnungen verschiedener Hochschulstandorte zeigt (vgl. Abschnitt »Lehrformate in den naturwissenschaftlichen Studiengängen«). Um den Qualitätsansprüchen dennoch genügen zu können, wurde daher die Pflicht zur internen und externen Qualitätssicherung eingeführt (orientiert an ENQA, 2009). Ulrich Welbers spricht sich im Rahmen der Modularisierung für die Etablierung von »Lerndramaturgien« aus (BOSBACH et al., 2007, S. 167), die der bloßen Addition bereits vorhandener Lehrveranstaltungen in Form von Mikromodulen entgegen wirken sollen und damit letztlich einen Beitrag zur Outputorientierung leisten können, da die Lehre im Zuge dessen aus neuer, output-orientierter Perspektive bedacht werden muss.

Lerndramaturgie meint einen Aufbau von Modulen nach hochschuldidaktischen Prinzipien, z. B. den didaktisch sinnvoll zu vollziehenden Lernschritten der Studierenden. Solche Lerndramaturgien können etwa »Einführung – Anwendung – Transfer« oder »Anwendung – Reflexion – Vertiefung« sein (ebd., S. 168). Darüber hinaus sollte zur Verzahnung der Modulbestandteile zu einer guten Lerndramaturgie hochschuldidaktisch sinnvoll sowohl inter- als auch intramodular verknüpft werden (ebd., S. 168; vgl. hierzu auch PraQ-Konstrukt »Integration«). Zentrales Instrument, um diese weitreichenden Struktur- und Studienprogrammanpassungen zu begutachten ist, neben der Studiengangsasskreditierung, die interne Evaluation (vgl. Abschnitt »Lehrevaluationen an Hochschulen«).

2.2. Qualität der Lehre

Die Qualität von Lehre an Hochschulen ist ein multidimensionales Konstrukt. Klassischerweise wird die Qualität von Lehre wortgemäß als die Qualität der lehrenden

Darbietung verstanden (KOCH, 2004). Doch gerade im Zuge von Strukturakkreditierungen und auch der Bologna-Reform rücken andere Blickwinkel in den Vordergrund zur Definition einer Qualität von Hochschullehre.

Eine Unterscheidung der Perspektiven bieten beispielsweise HARVEY & GREEN (2000, S. 26, 31), die die Bildungssysteme aus der Tradition der Qualitätssicherung in der Ökonomie betrachten. Diese Perspektive umfasst die Qualität als »Maß der Erfüllung der Ziele der Institution«. Darunter verstehen HARVEY & GREEN (2000), dass Studien- und Prüfungsordnungen diejenigen Bestandteile enthalten, die der Studienabschluss (das »Ziel«) umfassen soll oder dass intermodular vernetzt wird (BOSBACH et al., 2007, S. 168). Dies wird in Form von Akkreditierungsverfahren an Universitäten geprüft.

Eine zweite Perspektive nach HARVEY & GREEN (2000), die deutlich tiefer in die Lehre selbst hineinblickt, ist die Qualität »als Maß für Transformationsprozesse« der Studierenden (»Konsumenten«). Es geht in dieser lernerorientierten Perspektive um die Veränderung (»value added«) des Wissens, Verhaltens und der Einstellungen des Lernenden in der Universität, was im Sinne von Bologna bereits einer Outputorientierung entspricht.

Die Qualität von Lehre an Hochschulen lässt sich nach BLÜTHMANN (2012, S. 18ff), die wiederum auf DONABENDIAN (1966) verweist, in drei verschiedene Dimensionen unterteilen:

(1) Bei der strukturellen Dimension (*Lernumgebung*) rücken organisatorische und materielle Sachverhalte in den Fokus der Qualitätsbetrachtung von Lehre. Sind genügend materielle Ressourcen wie Räume, Verbrauchsmaterialien und technische Geräte vorhanden? Stimmt die Qualität begleitender Lehrwerke und Skripten? Ist das Verhältnis der Anzahl Lehrender zur Anzahl Studierender ausreichend? Stimmen sich die Lehrveranstaltungen organisatorisch miteinander ab? Dies weist Übereinstimmungen zum »Maß der Erfüllung der Ziele der Institution« nach Harvey auf.

(2) Bei der prozessorientierten Dimension (*Lehrkompetenz*) der Lehrqualität stehen die zahlreichen Interaktionen zwischen Lehrenden und Studierenden im Mittelpunkt der Betrachtung. Es kommt hier vermehrt auf die fachlichen, pädagogischen, didaktischen und methodischen Fähigkeiten der Lehrenden an (ebd., S. 18ff). Bezogen auf den Darbietungsstil beim Lehrendenvortrag wäre dies z. B. die Aufmerksamkeit der Studierenden zu wecken, anschauliche Beispiele im Sinne einer Elaboration zu bringen, allgemeine Verständlichkeit zu erreichen, kritisch Vor- und Nachteile verschiedener theoretischer Konzeptionen herauszustellen (KOCH, 2004, S. 65–69), auch Enthusiasmus und Engagement auszustrahlen (ebd., S. 215ff). Organisatorisch-

2. Lehre an Hochschulen

methodisch fiele darunter die strukturierte Planung der Lehrveranstaltung als Ganzes mit inhaltlich sinnvoll aufeinander aufbauenden Bestandteilen, eine Variabilität des Lehrstils mittels Darbietungshilfen und das Stellen effizienter Aufgaben wechselnder Formulierung für die Studierenden im Sinne regelmäßiger Wiederholung. Didaktisch-pädagogisch wiederum wird von den Dozierenden erwartet, dass sie die Studierenden zur eigenen Wissensüberprüfung und zum Nachdenken anregen, Überforderungen der Studierenden vorausschauend diagnostizieren können, Eigenaktivität ermöglichen und fördern, sowie Arbeitsstrategien und Verbesserungsmöglichkeiten aufzeigen (ebd., 2004, S. 65-69, 206ff). Manche Autoren sehen diese didaktisch-pädagogischen Aspekte als zentral für die Lehrqualität an:

Es ist jedoch andererseits davon auszugehen, daß [sic] der Kern von Bildungsarbeit in der Beziehung zwischen Lehrenden und Lernenden liegt, die im Lehr-Lern-Prozeß aufgebaut wird.

HARVEY & GREEN (2000, S. 26)

Folgt man den Grundsätzen der Bologna-Reform, so ist zusätzlich eine dritte Dimension (3) von zentraler Bedeutung.

(3) Die Ergebnis- oder Outputorientierung (BerlHG, §3, Abs. 1) bildet die dritte Dimension einer möglichen Strukturierung der Qualität von Lehre an Hochschulen. Sehr analog zur zweiten Perspektive nach HARVEY & GREEN (2000) ist man hier an den Lernfortschritten der Studierenden (Lernzuwachs) interessiert. Es geht also, anders als bei der Betrachtung der strukturellen und prozessorientierten Dimensionen, hier nicht mehr um manifest beobachtbare, sondern latente Merkmale der Lehre. Die Qualität einer einzelnen Lehrveranstaltung bemisst sich danach, zu welchen studien- oder berufsrelevanten Handlungen die Studierenden nach dem Besuch der Lehrveranstaltung fähig sind (»Kompetenzorientierung«).

Diese Handlungen wiederum müssen zu Beginn der Lehrveranstaltung von den Dozierenden festgelegt und am Ende der Lehrveranstaltung auf ihre Erreichung geprüft werden, zumindest in Form von der Erreichung handlungsrelevanter Voraussetzungen (»Kompetenzen«). Dies kann in Form von Bildungsstandards, also gemeinsamen, hochschulübergreifenden Festlegungen geschehen, wie beispielsweise dem Deutschen Qualifikationsrahmen (HRK, 2011), der Wissensverbreiterung, Wissensvertiefung, Instrumentale Kompetenz (z.B. Problemlösen), Systemische Kompetenz (z.B. Recherche) und Kommunikative Kompetenz als Kompetenzen definiert. Folgt man der Beschreibung von Kompetenzen der Rahmenlehrpläne aus vier

deutschen Bundesländern (Landesinstitute der Länder Berlin, Brandenburg, Bremen sowie Mecklenburg-Vorpommern, 2004), so wären dies Fachkompetenz, Methodenkompetenz (Arbeitstechniken), Sozialkompetenz (Kommunikation, Kooperation) und Personalkompetenz (produktive Selbsteinstellung).

Durch den lehrveranstaltungs- und fächerübergreifenden Charakter dieser dreidimensionalen Betrachtung von Qualität der Lehre ist eine Zuordnung jedweder Qualitätsaspekte aus der Literatur zu einer dieser drei Dimensionen möglich.

Kapitel 3

Lehrformen in naturwissenschaftlichen Studiengängen

Im Fokus dieser Arbeit steht mit dem »Naturwissenschaftlichen Praktikum« eine spezielle Lehrform naturwissenschaftlicher Studiengänge. Im Folgenden werden daher die gängigen Lehrveranstaltungsformen (naturwissenschaftlicher) Studiengänge überblicksartig dargestellt und das Praktikum in diesen Kanon eingeordnet, um wesentliche organisatorische Unterschiede herauszustellen.

Dies dient dann in Abschnitt 6.2 der Argumentation, dass Standard-Lehrveranstaltungsevaluationen der besonderen Struktur und Qualitätsansprüchen von Praktika nicht gerecht werden. Es wird vom Allgemeinen zum Speziellen vorgegangen: Es erfolgt zunächst je Lehrveranstaltungsform eine Einordnung in nationale Vorgaben (auch abgeleitet aus der Bologna-Reform), gefolgt von einer Betrachtung der fächerübergreifenden Regularien der Freien Universität Berlin bis hin zur physikspezifischen, aber universitätsübergreifenden Perspektive. Für die Lehrveranstaltungsform Praktikum wird darüber hinaus insbesondere auf den aktuellen Forschungsstand bezüglich der Lernwirksamkeit und übergreifender Probleme im Lehrveranstaltungsbetrieb eingegangen.

3.1. Vorlesung

Naturwissenschaftliche Studiengänge enthalten als Lehrformat, wie nahezu alle anderen Studiengänge auch, die Vorlesung. Allgemein wird diese definiert als »große Frontal-Lehrveranstaltung« vor einem größeren Auditorium. Bei Vorlesungen mit unmittelbarer studienbegleitender Prüfung wird allerdings eine Maximalgröße von 100 Teilnehmer*innen vorgeschlagen (BOSBACH et al., 2007, S. 30). Der/die Dozent*in wird hierbei als aktiver Part bezeichnet, die Teilnehmer*innen sollen vornehmlich durch die Rezeption lernen. Bis auf kurze Rückfragen oder »u.U. kurze Übungsteile« sind die Interaktionsmöglichkeiten stark beschränkt (ebd., S. 30).

Fächerübergreifend, aber universitätsspezifisch (hier am Beispiel der Freien Universität Berlin; FUB, 2012a, S. 4ff) wird die Vorlesung auch als »Frontallehrveranstaltung mit kurzen Interaktionen« betitelt. Blickt man nun konkreter in die naturwissenschaftlichen (hier speziell physikalischen) Studien- und Prüfungsordnungen (SPO), so werden diese Eigenschaften fortgeführt. Es wurden vor allem diejenigen

3. Lehrformen in naturwissenschaftlichen Studiengängen

SPO betrachtet, deren Universitäten direkten Anteil an den Studien dieser Arbeit hatten (Abschnitte 12 & 13). Im Curriculum der Universität Wien heißt es beispielsweise »Darstellung von Themen, Gegenständen und Methoden (...) unter kritischer Berücksichtigung verschiedener Lehrmeinungen« (UNIVERSITÄT-WIEN, 2011, S. 17), was den methodischen Part betont. Hier können beispielsweise Demonstrationsexperimente während der Vorlesung gemeint sein, die einen theoretisch dargestellten Sachverhalt veranschaulichen sollen. Ein Blick in die Bachelor-SPO Physik der Freien Universität zeigt ebenfalls eine Berücksichtigung von »methodischen/-theoretischen Grundlagen« (FUB, 2012b, S. 1179). Darüber hinaus wird hier auch der »Vortrag der jeweiligen Lehrkraft« betont (ebd., S. 1179).

Die Universität Tübingen spart es indes in der SPO Physik des Bachelors gänzlich aus, den Begriff der Vorlesung noch genauer zu definieren. Hier heißt es lediglich in der bereits veralteten SPO von 2009 im §5 »Arten von Lehrveranstaltungen«: »(...) werden regelmäßig Vorlesungen zur theoretischen Physik und zur Experimentalphysik angeboten« (UNIVERSITÄT-TÜBINGEN, 2009), die aktuelle SPO von 2013 enthält keine solchen Informationen (UNIVERSITÄT-TÜBINGEN, 2013). Die Bergische Universität Wuppertal und die Universität zu Köln dagegen definieren weder das Lehrformat Vorlesung, noch Seminare, Übungen oder Praktika (UNIVERSITÄT-KÖLN, 2010; BUW, 2013).

Zusammenfassend lässt sich feststellen, dass Vorlesungen im Naturwissenschaftsstudium den eher rezeptiven Part der Veranstaltungsformen ausmachen. Über die klassische und fächerübergreifende Funktion als »große Frontallehrveranstaltung« hinaus dient sie allerdings speziell in den Naturwissenschaften auch der Darstellung experimenteller Methoden, was eine inhaltliche Vernetzung zu stattfindenden Praktika nahelegen kann (vgl. Qualitätsdimension »Integration« im Modell der Praktikumsqualität in Kapitel 7).

3.2. Seminar

Das Seminar wird im Zuge des Bologna-Prozesses als »kleine Lehrveranstaltung mit signifikantem, unterschiedlich aktivem Anteil der Teilnehmer« (BOSBACH et al., 2007, S. 31) verstanden. Die Interaktion mit dem oder der Dozierenden nimmt hier also gerade im Vergleich zur Vorlesung oder zur Übung einen hohen Anteil ein. Dies geht soweit, dass Teilnehmer*innen sich gegenseitig präsentieren oder ihre Lernumgebung selbst gestalten können. Die Gruppengröße einer solchen, methodisch aufwändigeren Lehrform wird bestimmt durch die Intensität der Aktivitäten (ebd., S. 31).

Universitätsweit wird für die Freie Universität Berlin zwischen klassischen Seminaren (LV-Typ B, FUB, 2012a, S. 5; KAPVO, 2012, S. 83) und eher forschungsorientierten Projektseminaren (LV-Typ C, FUB, 2012a, S. 6; KAPVO, 2012, S. 83) unterschieden. Diese inhaltliche Unterscheidung geht auch einher mit dem Betreuungsschlüssel, der von 1 zu 30 in Seminaren zu 1 zu 15 in Projektseminaren reicht (KAPVO, 2012, S. 83). Erstere werden hier breit definiert als »Veranstaltungen mit aktiven Beiträgen der Studierenden« und umfassen z. B. auch die »Übung in den Naturwissenschaften« (vgl. Abschnitt »Übung«). Letztere dagegen sind charakterisiert durch eine hohe Interaktionsrate beim Lehr-Lern-Prozess, sowie das Befassen mit »vorwiegend neue[n] Problemstellungen« (FUB, 2012a, S. 6). Im Gegensatz zur Vorlesung wird auf Basis der hier betrachteten Stichprobe seltener auf die Spezifika von Seminaren in Physik-SPO eingegangen. Weder die Universität Wien, noch Tübingen, Wuppertal oder Köln gehen hierauf gesondert ein.

Lediglich die SPO Physik im Bachelor der Freien Universität Berlin geht auf die methodische Vielfalt ein: »Die vorrangigen Arbeitsformen sind Seminargespräche auf der Grundlage von Unterrichtsmitteln, von vorzubereitender Lektüre (Fachliteratur und Quellen), von Arbeitsaufträgen sowie die Gruppenarbeit.« (FUB, 2012b, §5, S. 1179). Zusammenfassend bildet das Seminar im Physikstudium ein sehr schulähnliches und gemäß konstruktivistischer Lerntheorien als deutlich wirksamer anzunehmendes Lehrformat als etwa die Vorlesung (BERENDT, 1994, S. 40).

3.3. Übung

Die »Übung« wird bei fächerübergreifender Betrachtung als »kleine Frontal-Lehrveranstaltung« (BOSBACH et al., 2007, S. 31) verstanden. Dies bedeutet, dass ähnlich wie bei der Vorlesung die Hauptaktivität dem oder der Dozierenden obliegt. Ein dialogischer Austausch mit den Studierenden ist hier etwas mehr erwünscht als in Vorlesungen. Sehr intensiv wird dieser aber i. d. R. nicht. Allerdings wird auch seminaristischer Unterricht mit in diese Definition aufgenommen, wodurch eine gewisse Unterschiedlichkeit in der Auslegung je Fach entsteht (ebd., S. 31).

Fächerübergreifend gibt es an der Freien Universität Berlin die Übung als spezielle Definition im Lehrveranstaltungstyp B nur für die naturwissenschaftlichen Fächer. Hier soll die »Übung in den Naturwissenschaften« dafür sorgen »noch unsichere erste Lernstrukturen, Inhalte oder Kompetenzen durch mehrfache Wiederholung« zu stabilisieren (FUB, 2012a, S. 5). Es geht also um ein »Einüben« von z. B. analytischen Verfahren und Algorithmen, das dann wiederum »in neuen Situationen angewendet werden kann« (ebd., S. 5).

3. Lehrformen in naturwissenschaftlichen Studiengängen

Dies ist eine direkte Anknüpfung an die Forderungen gemäß Bologna-Reform nach mehr anwendungsrelevantem Wissen (vgl. Abschnitt 3.4.1).

Konträr zur Darstellung im Bologna-Kontext, aber konvergent zur Darstellung der Freien Universität Berlin besagen auch die SPO des Bachelors der Physik dieser und anderer Universitäten, dass die Übung als eine aktivierende und für die Studierenden herausfordernde Lehrform verstanden wird. In dieser werden methodische Kenntnisse weiterentwickelt und Problemlöseaufgaben bewältigt (vgl. FUB, 2012b, §5; UNIVERSITÄT-TÜBINGEN, 2009, §5; UNIVERSITÄT-WIEN, 2011, §7). Die Rolle des/der Dozierenden wird hier sogar mitunter als lediglich »kontrollierend« bezeichnet (FUB, 2012b, §5, S. 1179).

Zusammenfassend bildet die Übung im Physikstudium ein nah an der Kompetenzorientierung zu verortendes Beispiel für Lehrformen. Da diese Lehrform zudem in den SPO deutlich häufiger vorliegt als z. B. Seminare, lässt sich hieran ein deutlicher Anteil an studierendenaktivierender Lehre ausmachen.

3.4. Praktikum

Unter einem »Praktikum« soll im Folgenden eine Lehrform (an der Universität) bezeichnet werden, die manchmal auch als »Naturwissenschaftliches Praktikum«, »Grundpraktikum« oder »Laborpraktikum« benannt wird. Dies geschieht in Abgrenzung zu beispielsweise Betriebspraktika oder Schulpraktika, sowie Praxisseminaren, wie sie z. B. im Lehramtsstudium üblich sind.

»Unter Praktikum ist dabei eine Form von praktischer Arbeit zu verstehen, bei der Studierende [...] Experimente in einer speziell dafür vorbereiteten Lernumgebung durchführen, sich mit Geräten und Materialien beschäftigen, naturwissenschaftliche Phänomene beobachten und verstehen lernen und naturwissenschaftliche Arbeitsweisen erlernen können.«

NAGEL (2009, S. 4)

Obiger Definition wird im Rahmen dieser Arbeit gefolgt. Allerdings beziehen sich die Ausführungen primär auf die ersten Praktika des Naturwissenschaftsstudiums (Kriterium: spätestens 4. Semester), oft als »Anfängerpraktika« bezeichnet. Gemäß dem Schwerpunkt der Arbeit soll die Vorstellung dieser Lehrform in einem deutlich größeren, differenzierterem Ausmaß geschehen, um mögliche Qualitätsindikatoren für Praktika zu identifizieren, die in ein theoretisches Modell einfließen können. Dafür wird zunächst analog zum bisherigen Vorgehen auf die Definition des Lehrformats gemäß Literatur der Bologna-Reform im Allgemeinen, fächerübergreifend an

der Freien Universität und nach den exemplarischen SPO der Physik im Speziellen eingegangen, um die offizielle Sichtweise zu erhalten.

Es folgen Ausführungen zur (fachdidaktischen) Legitimation und zu Zielen dieser Lehrform. Auf den extrem verbreiteten und vielfach kritisierten (vgl. Abschnitt 5) Typus des »traditionellen« bzw. »expository-style« Praktikums wird dann aus der Perspektive von Praktikern eingegangen, gefolgt von weiteren Typisierungen dieser Lehrveranstaltungsform.

3.4.1. Praktika gemäß Bologna-Reform und Studienordnungen

Der fächerübergreifende Standpunkt der Bologna-Reformen wird in Ihrer Definition eines »Praktikums« deutlich. Es wird hierunter ein Lehrformat verstanden, bei der Teilnehmer*innen in kleineren Gruppen konkrete Aufgabenstellungen über einen längeren Zeitraum hinweg weitgehend selbstständig erarbeiten (BOSBACH et al., 2007, S. 31), im Wesentlichen also Projektarbeit betreiben. Für die Rolle des oder der Dozierenden wird hier angeführt, dass eine dauerhafte Anwesenheit gar nicht notwendig sei, wohl aber die Besprechung des Projektfortschritts in regelmäßigen Abständen. Als Übungsmedium für die Studierenden werden Protokolle vorgesehen. Die Berliner Verordnung über die Kapazitätsermittlung, die Curricularnormwerte und die Festsetzung von Zulassungszahlen beschreibt hierzu:

Veranstaltung mit weitgehend selbständiger (Gruppen-)Arbeit der Studierenden, Erwerb und Vertiefung von Kenntnissen durch Bearbeitung wissenschaftlicher, praktischer oder experimenteller Aufgaben; Lehrender leitet die Studierenden an; Studierende führen Beobachtungen, Arbeiten und Versuche durch, wenden ihre Kenntnisse an, ziehen wissenschaftliche Schlussfolgerungen

KAPVO (2012, S. 83)

Im Rahmen der Modularisierungsfestlegungen der Freien Universität Berlin (FUB, 2012a, S. 7) wird das »interne Praktikum in Naturwissenschaften« unter den Lehrveranstaltungstyp D gefasst, in dem im Wesentlichen Formate angeleiteter Projektarbeit wie »Projektmodul« oder »Problemorientiertes Lernen« angesiedelt sind. Diesen gemein ist die »0.5« im Betreuungsschlüssel, also die nicht dauerhafte Anwesenheitsnotwendigkeit des/r Dozierenden, da es sich um »weitgehend selbstständige (Gruppen-)Arbeit« handelt (KAPVO, 2012, S. 84). Hierbei macht das »interne Praktikum in Naturwissenschaften« aus, dass es die Theorie aus Vorlesung oder Übung mit experimentellen Erfahrungen verknüpfen möchte. Die Erarbeitung der

3. Lehrformen in naturwissenschaftlichen Studiengängen

Fragestellungen und Lösungen soll selbstständig geschehen, praktische und analytische Fähigkeiten sollen erworben werden.

Eine exemplarische Begutachtung der Studien- und Prüfungsordnungen für Physik im Bachelor der Universitäten Köln, Tübingen, Wien und Wuppertal ergab nur für die Universität Tübingen nennenswerte Ausführungen zur Lehrform Praktikum (UNIVERSITÄT-KÖLN, 2010; UNIVERSITÄT-TÜBINGEN, 2009; UNIVERSITÄT-WIEN, 2011; BUW, 2013). In den Definitionen der Lehrformen und in den Modulbeschreibungen ist der Erwerb »methodischer Fähigkeiten der experimentellen Physik« zentral (UNIVERSITÄT-TÜBINGEN, 2009, §5; FUB, 2012b, §5). Im Rahmen der betrachteten SPO wird ausschließlich in der SPO der FUB (§5) auf zwei weitere Schwerpunkte der Praktika eingegangen. Hier wird sowohl die »Anwendung der erworbenen fachwissenschaftlichen Kenntnisse« als auch das Testen der Eignung für bestimmte Berufsfelder angeführt. Aggregiert man die Informationen der betrachteten SPO und wagt sich an eine Verallgemeinerung, so erscheinen Praktika demnach in einer Doppelrolle: Einerseits dienen sie ähnlich wie etwa Übungen oder Seminare dazu, das Fachwissen zu vertiefen, andererseits zeichnen sie sich aber auch als Lehrveranstaltung für eine direkte oder indirekte Vorbereitung möglicher späterer Berufstätigkeit aus.

Zusammenfassend sind damit auf theoretischer, rechtlicher und struktureller Ebene die Praktika als mit vielfältigen theoretischen wie praktischen Erwartungen belegte Lehrveranstaltungen zu sehen.

3.4.2. Ziele und Legitimation von Praktika

Betrachtet man konkreter die tatsächliche Umsetzung bzw. befragt man Praktikumsleitende, so äußern diese vielfältige Ziele und damit Legitimationsgründe von Praktika. In 1999 wurden Lehrende in Praktika befragt, welche Ziele sie mit ihren Praktika verfolgen (HALLER, 1999). Die Angaben können teils als implizite Ziele verstanden werden, da diese nicht zwangsläufig institutionell standardisiert und damit kommuniziert sind. Es finden sich drei zentrale Ziele, die mit der experimentellen Ausbildung an Hochschulen erreicht werden sollen. Die Ziele bestätigten sich fächerübergreifend und in diversen Studien:

(1) Theorie und Praxis verbinden, (2) experimentelle Fähigkeiten erwerben und (3) Methoden wissenschaftlichen Denkens kennenlernen. (BORAWSKI & HEINKE, 2005; HALLER, 1999; NEUMANN, 2004; THEYSEN, 2000; WELZEL et al., 1998).

Eine europaweite Befragung zu der allgemeineren Fragestellung der Ziele, die Lehrende mit dem Experimentieren verfolgen, kam zu analogen Ergebnissen (WELZEL et al., 1998).

Unter (1) wird subsumiert, Gesetze zu verifizieren, experimentelle Methoden naturwissenschaftlichen Inhalten zuzuordnen oder technische Anwendungen erläutern zu können. Dies bildet den eher inhaltlichen Rahmen der Ziele und ist ein Stück weit mit der heutigen Vorstellung des Kompetenzbereichs Fachwissen verwandt. Gemäß Strukturierung der Lehrqualität in die Dimensionen *Lernzuwachs*, *Lehrkompetenz* und *Lernumgebung* (vgl. Abschnitt 2.2) kann (1) also klar dem *Lernzuwachs* zugeordnet werden.

Punkt (2) beinhaltet vor allem Standardtechniken, wie das Protokollieren, Beobachten und/oder Messfehlerbetrachtung. Dies weist Ähnlichkeit mit dem Kompetenzbereich Erkenntnisgewinnung auf und gehört ebenso in die Dimension *Lernzuwachs*. HOFSTEIN (2004) nennt sogar konkret den Erwerb von Erkenntnisgewinnungskompetenz als Eigenschaft von Praktika.

(3) schließlich betrifft vor allem die Metaebene des Experimentierens mit dem Planen und Durchführen von Untersuchungen, der kritischen Interpretation von Daten, und dem Wissen um Erkenntnismethoden. Hier kann nach heutiger Modellierung von einer Überschneidung zwischen Erkenntnisgewinnung und Bewertung ausgegangen werden. Eine erneute Zuordnung zur outputorientierten Dimension *Lernzuwachs* kann hier erfolgen.

Nach HALLER (1999) geht es bei Praktika auch um persönliche und soziale Aspekte wie Interesse, Kommunikation und Teamfähigkeit (Weiterentwicklung der sozialen Kompetenz, Persönlichkeit und Motivation), weswegen auch Aspekte der Kommunikationskompetenz (auch bei HOFSTEIN, 2004; ENGELS & HÜLSBUSCH, 2008, S. 59) und fächerübergreifender Metakompetenzen wie Selbstmanagement oder Kooperationsfähigkeit (HOFSTEIN, 2004) angesprochen werden. Diese lassen sich erneut der Dimension *Lernzuwachs* zuweisen. Ähnliche Ziele formulieren für Fortgeschrittenen-Praktika ZWICKL et al. (2013) und für Grundpraktika die AAPT (1998). HUCKE (1999) ordnete diese Zielstellungen noch einmal mit WELLINGTON (1998) in Bereiche der »Fertigkeiten«, sowie in »affektiv« und »kognitiv« bestimmte Bereiche. Unter Fertigkeiten wird u. a. das Messen lernen (Hauptziel nach DIEMER et al., 1998, S. 11), experimentelle Fähigkeiten und Methoden wissenschaftlichen Denkens verstanden, was Schnittmengen zu den Zielen 2 und 3 nach HALLER (1999) aufweist und damit erneut der *Lernzuwachsdimension* zuzuschreiben ist.

Der affektive Bereich wird wie bei ebd. durch motivationale und einstellungsbezogene Merkmale (vgl. HOFSTEIN, 2004) bestimmt, erneut eher studierendenorientiert und daher der Dimension *Lernzuwachs* zuzuordnen.

3. Lehrformen in naturwissenschaftlichen Studiengängen

Kognitiv werden schließlich das Anwenden naturwissenschaftlicher Konzepte, der Erwerb theoretischen Wissens¹ und die Vertiefung des Wissens durch Praxis verstanden (Hauptziel nach DIEMER et al., 1998, S. 11), was nah am Ziel (1) nach HALLER ist und der Dimension *Lernzuwachs* zuzuordnen ist. Auf der anderen Seite stehen die expliziten, kommunizierten Ziele von Praktika, die sich in veröffentlichten Zielempfehlungen und Studienordnungen wiederfinden. Die Konferenz der Fachbereiche Physik (KFP, 2010) beispielsweise definiert für die »Gruppe 3: Experimentiertechniken« im Bachelorstudium viele »Lernergebnisse« (ebd., S. 8), die sich auf den Experimentierzyklus beziehen (Messgeräte, Messmethoden, Messdatenerfassung, Messdateninterpretation, Messfehler etc.), aber auch auf schriftliche Kommunikation wie die saubere Protokollführung. Hier wird also u.a. explizit dem Ziel (2) nach HALLER (1999) entsprochen, aber auch erneut die Kommunikationskompetenz angesprochen. Beides ist erneut dem Lernerfolg der Studierenden zuzuordnen.

Die KAPVO (2010) konstatiert, dass interne Praktika in den Naturwissenschaften dazu dienen, den in der Vorlesung bzw. Übung behandelten Stoff durch Anwendung von Methoden an einem realen experimentellen Aufbau zu erproben (ebd., S. 83). Dabei sollen selbstständig Fragestellungen erarbeitet, Methoden und Kenntnisse angewandt und Problemlösungen gefunden werden, allerdings unter Anleitung. Auch HOFSTEIN (2004) sieht das Problemlösen als Bestandteil der Praktika. Mit kompetenzorientiertem Blick geht es dabei explizit um das Erlernen praktischer und analytischer Fähigkeiten, was erneut der Qualität von Lehre in der Dimension *Lernzuwachs* zuzurechnen ist.

Hinzu kommt in der SPO Physik im Bachelor der Freien Universität Berlin die Formulierung: »In Veranstaltungen, die Teil eines Praktikums sein können, soll besonders auf Lehrinhalte in den Praktika eingegangen, eventuelle Unklarheiten beseitigt und Erfahrungen aus der Praxis reflektiert werden« (FUB, 2012b, S. 1179) Dies spricht zum ersten Mal strukturelle Anforderungen (Dimension *Lernumgebung*) an Praktika an, da die Inhalte des Praktikums mit anderen oder integrierten Lehrveranstaltungen inhaltlich verzahnt werden sollen.

Es kann insgesamt festgehalten werden, dass sehr viele unterschiedliche Ziele mit Praktika verfolgt werden, dass sich aber auch je nach Quelle die Ziele unterschiedlich gewichten lassen. Übergreifender Schwerpunkt ist eine Bologna-konforme Formulierung der Erkenntnisgewinnung bzw. der experimentellen Kompetenz. Die Zielkategorien nach HALLER (1999) bieten einen guten Referenzpunkt, da sie alle wesentlichen Zielsetzungen umfassen und eine der wenigen fächerübergreifenden Unter-

¹HOFSTEIN (2004) sieht konstruktivistisches inhaltliches Lernen als Stärke von Praktika an.

suchungen darstellen. Ergänzt werden sollte dieser Katalog gemäß den offiziellen Zielempfehlungen und Studienordnungen um kommunikative, kooperative Kompetenzen, Motivation und Selbstorganisation. Bezüglich der *Lernumgebung* sollte eine Integration anderer Lehrveranstaltungen stattfinden (vgl. 2.2, BOSBACH et al., 2007, S. 168).

Auf die Zielsetzungen von Praktika wird auch im Abschnitt zur Modellierung der Qualität von Praktika (Abschnitt 7) eingegangen werden. Insgesamt werden viele Stärken von Praktika in der Qualitätsdimension *Lernzuwachs* angenommen, allerdings sind die meisten davon empirisch nicht abgesichert, sondern bleiben spekulativ. Eine empirische Untermauerung einiger dieser Annahmen soll in Abschnitt 13 dieser Arbeit geliefert werden.

3.4.3. Typische Gestaltung und Eigenschaften von Praktika

Praktika finden unabhängig vom naturwissenschaftlichen Studiengang meist innerhalb der ersten Studiensemester statt (ENGELS & HÜLSBUSCH, 2008; HUCKE, 1999; KREITEN, 2012; MEYER-ARNDT, 2000; NAGEL, 2009) und sind große Veranstaltungen mit bis zu 1000 Studierenden pro Semester, was einen entsprechenden Organisationsaufwand erfordert (ENGELS & HÜLSBUSCH, 2008, S. 58; DIEMER et al., 1998, S. 3). Für die Lehrveranstaltung ist meist ein/e Hochschullehrende/r organisatorisch verantwortlich.

Die eigentliche Betreuung und technische/inhaltliche Verantwortung erfolgt i. d. R. aber durch studentische oder wissenschaftlich mitarbeitende Betreuer*innen (ENGELS & HÜLSBUSCH, 2008, S. 58; HUCKE, 1999, S. 35ff.). Pro Versuch sind meist ein bis zwei dieser Betreuer*innen anwesend und für eine Gruppe von üblicherweise vier bis fünfzehn Studierenden verantwortlich (ebd.). Es gibt in der Lehrveranstaltung in physikalisch oder technisch orientierten Praktika meist je Woche einen Präsenztermin (Ausnahme: Blockpraktika) mit üblicherweise drei bis vier Stunden Präsenzzeit (ENGELS & HÜLSBUSCH, 2008, S. 58; KISSMANN et al., 2010; KREITEN, 2012; NAGEL, 2009; PLOMER, 2011; THEYSSEN, 2000), in der Chemie sind die Präsenzzeiten manchmal offener und dadurch auch länger gehalten (z. B. ARENZ & KNOLL, 2011, S. 5).

In jeder dieser Präsenzzeiten wird ein Versuch zu einem bestimmten Thema bearbeitet, gegliedert in meist mehrere Unterexperimente. Die Durchführung geschieht durch die Studierenden meist in Zweiergruppen (HUCKE, 1999; KREITEN, 2012; MEYER-ARNDT, 2000; NAGEL, 2009; PLOMER, 2011; THEYSSEN, 2000). Dabei arbeiten in der Regel mehrere dieser Gruppen zeitgleich in einem Praktikumsraum an identischen Versuchsaufbauten. Der Aufbau ist hierbei in physikalisch-technischen

3. Lehrformen in naturwissenschaftlichen Studiengängen

Praktika meist zu einem Großteil vorbereitet, in Chemiepraktika ist fachbedingt oft mehr Aufbauarbeit zu leisten.²

Die Abfolge der Experimente ist meist in einer Versuchsanleitung festgelegt (DIEMER et al., 1998, S. 3). Mit der Versuchsanleitung – oft auch als Skript oder Versuchsskript bezeichnet – wird auch die Vorbereitung der Studierenden maßgeblich gesteuert (ABRAHAM et al., 1997; HUCKE, 1999, S. 149). Das Skript stellt hiermit einen wesentlichen Bestandteil der Qualitätsdimension *Lernumgebung* dar (vgl. Abschnitt 2.2). Bei einer Betrachtung verschiedener Skripte zu Experimentalpraktika in der Physik erkennt man sehr schnell strukturelle Gemeinsamkeiten und einen einheitlichen Aufbau (BRÜGGEMANN, 2015). Üblicherweise kann man die Skripte zunächst in zwei Teile gliedern: Als erstes werden alle organisatorischen und strukturellen Fragen zum Praktikum behandelt. Das umfasst ein Vorwort sowie Inhaltsverzeichnis und eine kurze Zusammenfassung der Praktikumsordnung, der Zielstellung und des Ablaufs der Lehrveranstaltung. Zudem werden häufig weitere Unterkapitel für experimentübergreifende Themen verfasst. Dazu gehören der zeitliche Ablauf einzelner Experimentiertermine, Methoden der Messdatenerhebung, Verfahren der Fehlerrechnung, Anmerkungen zur korrekten Protokollführung und die formalen Gestaltung des Versuchsberichtes. Danach folgen die Versuchsanleitungen, bei denen jedes einzelne Experiment sein eigenes Unterkapitel erhält. Sinn dieses Fachtextes ist es, den Studierenden einen Überblick über die Thematik zu geben und einen Ausgangspunkt für die theoretische Vorbereitung (und damit auch für die Nachbereitung) des Versuches zu geben. Es sind hier noch keine Hinweise oder Angaben zur Durchführung des Experimentes enthalten. Der Umfang dieses Textteils variiert je nach Skript und Komplexität des Versuches, bewegt sich aber in einem allgemeinen Rahmen von einer bis zehn (im Durchschnitt sechs) Seiten bei den untersuchten Skripten von BRÜGGEMANN (2015).³

Danach folgt i. d. R. ein Abschnitt, der sich mit der konkreten Versuchsdurchführung befasst. Er enthält die durchzuführenden Aufgaben sowie eine Versuchsanleitung und ergänzende Fragen, die auf eine Erweiterung des Experiments oder eine vertiefende Nachbereitung abzielen. Die Versuchsanleitung selbst ist typischerweise in einem Fließtext verfasst. Sie folgt dabei der zeitlichen Abfolge der Versuchshandlungen und beschreibt den Versuchsaufbau sowie die zu verwendenden Geräte und

²Dies kommt daher, als das Lösungen und v. a. gefährliche Chemikalien erst unmittelbar zum Experimentbeginn umgefüllt werden können.

³»Die hier betrachteten Beispiele wurden aus online verfügbaren Skripten zu physikalischen Grundpraktika ausgewählt. Es handelt sich dabei um die Skripte der TU Darmstadt, FAU Erlangen-Nürnberg, Martin-Luther-Universität Halle-Wittenberg, Leibniz Universität Hannover und Universität Duisburg-Essen« (BRÜGGEMANN, 2015, S. 8).

deren Besonderheiten. Zudem werden in diesem Abschnitt die für den Versuch notwendigen Angaben zu den technischen Daten der Geräte am Versuchsplatz gemacht und Sicherheitshinweise gegeben – sofern erforderlich. Diese Angaben erfolgen selbst innerhalb der meisten Skripte nicht einheitlich (BRÜGGEMANN, 2015).

Die Vorbereitung findet plangemäß einige Tage vor dem Präsenztermin statt. Die Studierenden bereiten sich hierbei selbstständig auf die Theorie des Versuchs vor. Das Skript liefert hierfür wichtige Informationen und beinhaltet sowohl physikalische Theorie, als auch experimentelle Aufgaben für den Präsenztermin. Manchmal finden sich im Skript auch Skizzen zum Versuchsaufbau oder von Geräten (HUCKE, 1999, S. 35ff.). Die erfolgreiche Vorbereitung wird zu Beginn der Präsenzzeit meist in einem mündlichen Kolloquium durch den Betreuenden in der Gruppe geprüft (ENGELS & HÜLSBUSCH, 2008, S. 59). Während der Durchführung der Experimente erhalten die Studierenden Hilfestellungen durch die Betreuenden. Die am Versuchsort vorzufindenden Platzskripte⁴ enthalten spezifische Zusatzinformationen. Die Protokollerstellung geschieht nach der Durchführung meist zu Hause, allein oder mit dem Partner gemeinsam. Das Protokoll umfasst mehrere Seiten und beinhaltet Theorie, Messung und Auswertung, wird i. d. R. benotet und in einem mündlichen Abschlussgespräch besprochen. In Chemiepraktika werden gern auch theoretische Begleitseminare angeboten (FRESE, 2001, S. 30).

3.4.4. Typen von Praktika

Obschon die meisten Beschreibungen von Praktika im deutschsprachigen Raum »traditionellen« Praktika entsprechen (HUCKE, 1999; SANDER, 2000; KREITEN, 2012; ZASTROW, 2001; NAGEL, 2009; NEUMANN, 2004), lohnt ein Blick auf eine mögliche Kategorisierung anders gearteter Praktika. DOMIN (1999, 543) schlägt hierzu folgende Klassifizierung vor (Tab. 3.1):

Das »traditionelle Praktikum« wird im internationalen Raum als »expository style« bzw. »verification«, also Verifizierungs-Praktikum bezeichnet. Der Name ist

Tabelle 3.1.: Klassifizierung von Praktika nach DOMIN (1999)

Style	Descriptor		
	Outcome	Approach	Procedure
Expository	Predetermined	Deductive	Given
Inquiry	Undetermined	Inductive	Student generated
Discovery	Predetermined	Inductive	Given
Problem-based	Predetermined	Deductive	Student generated

⁴Dies sind in die Tischoberfläche eingepflegte Kurzanweisungen.

3. Lehrformen in naturwissenschaftlichen Studiengängen

insofern Programm, als dass es implizites Ziel ist, bereits bekannte Gesetze zu verifizieren, in der Literatur wird hier verächtlich von »nachkochen« gesprochen (HENSE, 2012, 79). Das Ergebnis des Versuchs steht also bereits zu Beginn fest, auch das naturwissenschaftliche Prinzip, das erarbeitet werden soll, steht deduktiv bereits zu Beginn zur Verfügung, sei es explizit im Skript oder beim Gespräch mit dem/r Betreuer*in. Zudem steht auch die Prozedur, um das gegebene Gesetz zu verifizieren, von vornherein fest und bildet einen klaren Anwendungsfall, der meist kleinschrittig abgearbeitet werden kann (z. B. zu sehen in der Auswahl an Skripten bei BRÜGGE-MANN, 2015).

Die hohe Verbreitung dieses Praktikumstyps ist allerdings eher durch eine zweckmäßige Organisationsstruktur, denn durch didaktische Begründungen zu erklären (LAGOWSKI, 2002, S. 3), ermöglicht er doch das simultane Arbeiten sehr vieler Studierender mit wenig benötigter Betreuung, da ja alle Modalitäten des Versuchs deduktiv gegeben sind. Er ist dadurch auch kostengünstig und logistisch am einfachsten zu realisieren. Von der Lernwirksamkeit her gesehen scheint es allerdings nur geringe Erfolge bei den Studierenden zu geben (vgl. Abschnitt 5). Studien zeigen, dass sogar die um eine geringe Lerndienlichkeit ihrer Experimentierumgebung wissenden Lehrenden den Praktikumstypus beibehalten (Montes und Rockly, 2002, S. 247). So genannte »inquiry style« Praktika (vgl. Tab. 3.1) haben einen gegensätzlichen Ansatz. Das Ergebnis ist offen, das zu erarbeitende Prinzip wird erst durch mehrere einzelne Phänomene deutlich, die Verfahren müssen von den Studierenden erdacht werden. Solche Praktika existieren auch an deutschsprachigen Hochschulen, allerdings meist in höheren Fachsemester-Modulen (Projektpraktika) oder in der Fachdidaktik (Demonstrations- und freie Praktika).

Da hier die Verantwortlichkeit für den Großteil des Erkenntnisgewinnungsprozess bei den Studierenden liegt, überrascht es vor einer konstruktivistischen Perspektive (HOFSTEIN & LUNETTA, 2004, S. 40) nicht, dass Forschungsergebnisse den inquiry-Praktika die Verbesserung von Einstellungen zum Experimentieren attestieren, sofern sie entsprechend lernfreundlich gestaltet werden (MERRITT et al., 1993). Der »discovery style« (vgl. Tab. 3.1) ist indes eine Art gelenktes inquiry. Die Studierenden verfolgen zwar einen induktiven Ansatz zur Ergründung ihres Lerngegenstands, das Ergebnis ist allerdings dem/der Betreuer*in bekannt und auf dieses kann hin gelenkt werden. Ein »problem based« Ansatz (vgl. Tab. 3.1) schließlich nutzt wieder einen deduktiven Zugang, diesmal mit offener Methode. Die Problemstellung das antreibende Element, indem ein neues Problem aufgeworfen wird, das unter Anwendung bekannter Prinzipien lösbar ist.

Das Ziel steht, im starken Unterschied zum inquiry-Ansatz, von vornherein fest, aber der Weg dahin kann vielfältig sein. So klar diese Typen unterschieden werden können, so wenig Klarheit herrscht selbst international über deren Unterschiede in der Lernwirksamkeit, wie LAGOWSKI (2002, S. 4) feststellt. Der Fokus dieser Arbeit liegt auf den »expository style« Praktika, da diese zum Großteil in den ersten Semestern der naturwissenschaftlichen Studiengänge vorkommen.

Kapitel 4

Lehrevaluation an Hochschulen

4.1. Ziele von Lehrevaluationen

Die Evaluation wird als das adäquate Verfahren zur Verbesserung der eigenen Qualität angesehen, da es den Anforderungen einer eigenverantwortlichen Qualitätsentwicklung entspricht.

HOPBACH (2005a, S. 16)

Für die Lehrevaluation an Hochschulen werden im Folgenden häufig genannte Ziele dargelegt, die schließlich auch die Methodik und die Anforderungen an Evaluationsinstrumente steuern. HOPBACH (2005) spricht dabei bereits das Wesentliche aus: Das Verfahren der Evaluation ist eng mit dem Anspruch an Qualität und Qualitätsentwicklung verknüpft. Im Falle der Lehrevaluation an Hochschulen geht es um die Qualität der Lehre. Diese besteht, wie in Abschnitt 2.2 auf Seite 20 dargestellt, aus einem umfassenden Konglomerat von Konstrukten, über die auch nur teilweise eine Einigkeit besteht, ob sie denn wirklich für die Qualität von Lehre von Bedeutung sind.

Viele Konstrukte lassen sich einer der drei Dimensionen *Lernzuwachs*, *Lehrkompetenz* oder *Lernumgebung* zuordnen (vgl. Abschnitt 2.2). Die Frage nach den Qualitätskonstrukten ist also die erste in der Lehrevaluation. Die zweite ist die Frage der Perspektive. Je nachdem, welche Akteure des Hochschulsystems man befragt, werden unterschiedliche Konstrukte als wertvoll genannt. So unterscheidet etwa HOPBACH (ebd., S. 15) vier verschiedene Perspektiven, deren Qualitätsansprüche berücksichtigt werden müssten.

(1) Zum Ersten stellen die betroffenen Studierenden eine Zielgruppe dar, die nach HOPBACH (ebd., S. 15) vor allem auf ihre spätere berufliche Tätigkeit blickt und dafür ideal ausgebildet das Hochschulsystem verlassen möchte.¹ RICHARDSON (2005, S. 388) ergänzt hierzu noch die mittelfristige Perspektive, dass Studierende auch

¹Dies deckt sich allerdings nur bedingt mit den Erkenntnissen zur Studienmotivation Physik-Studierender an der Freien Universität Berlin. ALBRECHT (2011) fand hier vor allem intrinsisch geprägte motivationale Orientierungen. Für die Physik im Lehramt allerdings wurden im Einklang mit HOPBACHS Aussage stärker extrinsisch berufsbezogene motivationale Orientierungen festgestellt.

4. Lehrevaluation an Hochschulen

ihre Kurswahlen von Ergebnissen der Qualitätsmessung abhängig machen können, also gewillt sind, diejenigen Kurse vorzugsweise zu besuchen, die auch eine bessere Lehrqualität versprechen.

(2) Die entgegengesetzte Perspektive stellt diejenige der Lehrenden dar. Diese besitzen an Hochschulen einerseits oft die Perspektive der aktuellen Forschung und messen daher die Qualität ihrer Lehrveranstaltung auch an dem Grad der Bezüge zu aktuellen Erkenntnisfortschritten (HOPBACH, 2005, S. 15). Auf der anderen Seite sind gerade an Lehrverbesserung interessierte Lehrende stark interessiert an Feedback bezüglich ihrer Lehrveranstaltung seitens der Studierenden. Dies kann ihnen nämlich wertvolle Erkenntnisse liefern, wie ihre Lehrveranstaltung zielgruppenspezifisch verbessert werden könnte (RICHARDSON, 2005, S. 388). Ein starker Bezug zur Qualitätsdimension *Lehrkompetenz* wird ersichtlich.

(3) HOPBACH (2005, S. 15ff) differenziert außerdem in eine staatliche Ebene, die vor allem auf die Vergabe finanzieller oder personeller Ressourcen abzielt. Hier würde dann vor allem die Effektivität und Effizienz des Ressourceneinsatzes von Relevanz für die Qualität der Lehre sein. RICHARDSON (2005, S. 388) wiederum ergänzt dies mit dem Blickwinkel der Evidenzgewinnung, um administrative Entscheidungen treffen zu können. Die Qualitätsmanagementsysteme von Universitäten nehmen meist diesen dritten Blickwinkel ein (MERKT, 2016, S. 8). Es handelt sich hier vor allem um strukturelle Maße, die die Qualitätsdimension *Lernumgebung* ansprechen.

(4) Zu guter Letzt führt HOPBACH (2005, S. 15) einen Blickwinkel der Ministerien an, die vor allem auf Abschlüsse, Vorgaben oder Standards fokussieren, die ein gewisses Abstraktionslevel weg vom eigentlichen Lehrprozess erreicht haben.

(5) RICHARDSON (2005, S. 388) wiederum ergänzt einen fünften Blickwinkel: Auch die Forschung hat ein Interesse an der Qualität im Bildungs-, hier im Hochschulwesen. Bezogen auf die Lehrqualität sind hier vor allem Bildungswissenschaftler*innen und Didaktiker*innen Teil dieser Perspektivlage. So kann für diese Berufsgruppen der outcome von Lehrveranstaltungen ein Forschungsgegenstand sein, etwa zum Vergleich verschiedenartiger Lehrveranstaltungen in Bezug auf ihre Lernwirksamkeit. Damit sind sowohl eine erste Auswahl an Qualitätsindikatoren (vgl. Abschnitt 2.2), als auch die Ziele von Lehrevaluationen aus unterschiedlichen Perspektiven verdeutlicht worden. Wenn die Ziele und Qualitätsdimensionen feststehen, stellt sich als nächstes die Frage nach der Umsetzung, hier also der Portierung in ein Instrument, das zur Erreichung der Evaluationsziele geeignet ist, d. h. passende Evidenz für die unterschiedlichen Perspektiven und Qualitätsaspekte liefert. Der folgende Abschnitt geht daher näher auf die Methodik von Lehrevaluationen ein.

4.2. Methodik(en) der Lehrevaluation

Lehrevaluationen an Hochschulen folgen einer langen Tradition. Eine auch international seit über 30 Jahren verbreitete und etablierte Methode hierzu ist die Evaluation von Lehrveranstaltungen durch Einschätzungen der Studierenden (RICHARDSON, 2005; SELDIN, 1999), meist in Form von quantitativ ausgerichteten, schriftlichen Befragungen. Trotz des Primats dieser Methode sind noch viele weitere Methoden denkbar und auch eingesetzt worden (Tab. 4.1).

Hier wird meist per rating der Befragten ihre Zufriedenheit und/oder Selbsteinschätzungen zu Lehrenden oder dem ganzen Kurs forciert (RICHARDSON, 2005, S. 388). Technisch realisiert wird dies über einen meist universitäts- und lehrveranstaltungsübergreifenden Standard-Fragebogen, der automatisiert eingescannt und der/dem Lehrenden bzw. dem Institut in einer überblicksartigen, deskriptiven Zusammenfassung zur Verfügung gestellt wird (ebd., S. 388). Insgesamt wirkt dies wie ein einfacher und für Lehrende, Administrative und Studierende zufriedenstellender Prozess, der mittlerweile Routine geworden ist (ebd., S. 388, 392). Ein Grund für die enorme Beliebtheit und auch das Diktat seitens der Universitäten bezüglich des Einsatzes von schriftlichen, quantitativen Befragungen wird hierbei in den von der Teilnehmer*innenschaft her immer größer werdenden Kursen gesehen (ebd., S. 402).

So kann der großen Anzahl an Teilnehmenden insofern gerecht werden, als das jede Meinung in das Evaluationsergebnis einfließen kann. Andererseits sind quantitative Messungen i. d. R. ökonomischer auszuwerten als qualitative (RICHARDSON, 2005, S. 388). Aus Perspektive der Lehr- und Lernforschenden ergeben sich allerdings begründete Zweifel an dieser Praxis: Die skizzierten Instrumente für die Lehrevaluation entstehen oft durch universitätsinterne Konstruktionen (RICHARDSON, 2005, S. 388) und sind traditionell meist einseitig input- oder Lehrenden- und nur selten kompetenzorientiert (BRAUN et al., 2008, S. 30). Es werden also nicht alle Perspektiven berücksichtigt, v. a. der Lernerfolg der Studierenden ist hier zu nennen. Durch die universitätsinterne Konstruktion und Auswertung der Befragungen, die nicht immer wissenschaftlichen Standards genügen (POHLENZ, 2008, S. 67–69; KROMREY, 2000, S. 247; SCHMIDT, 2009, S. 164ff) sind die vorhandenen Interpretationen der Messergebnisse zudem in ihrer Gültigkeit anzuzweifeln (vgl. Abschnitt 4.4).

4.3. Strukturelle Verankerung von Lehrevaluationen

Der Akkreditierungsrat regelt das Akkreditierungssystem in Deutschland und liefert daher auch Richtlinien für die Qualität von Studium und Lehre an Universitäten (AKKREDITIERUNGSRAT, 2007). Ländergemeinsame Strukturvorgaben der Kul-

Tabelle 4.1.: Auswahl: Typen von Evaluationsmethoden nach Berk (2005, S. 49). Hervorgehoben sind die für diese Arbeit besonders relevanten Methoden.

Source of Evidence	Type of Measure(s)	Who Provides Evidence	Who Uses Evidence	Decision ¹
Student Ratings	Rating Scale	Students	Instructors	F/S/P
Peer Ratings	Rating Scale	Peers	Instructors	F/S
Self-Evaluation	Rating Scale	Instructors	Instructors	F/S
Videos	Rating Scale	Instructors/Peers	Instructors/Peers	F/S
Student Interviews	Questionnaires	Students	Instructors	F/S
Alumni Ratings	Rating Scale	Graduates	Instructors	F/S/P
Employer Ratings	Rating Scale	Employers	Instructors	P
Administrator Ratings	Rating Scale	Administrators	Administrators	S
Teaching Scholarship	Judgemental Review	Instructors	Administrators	S
Teaching Awards	Judgemental Review	Instructors	Faculty Committees	S
Learning outcomes	Tests, Simulations	Students	Instructors	F/P
Teaching Portfolio	Most of the above	Instructors, Students, Peers	Promotions Committees	S

¹F = formative, S = summative, P = program

4.4. Anforderungen an Evaluationsinstrumente und -auswertungen

tusministerkonferenz ergänzen dies um die strukturelle Ebene mit dem Fokus auf Gleichwertigkeit einander entsprechender Studien- und Prüfungsleistungen (KMK, 2010), was eine direkte Verknüpfung zur Lehre und damit zu ihrer Qualität beinhaltet. Darin ist die regelmäßige Lehrevaluation als Rückmeldungsinstrument festgelegt. Diese ist auch im Berliner Hochschulgesetz rechtlich verankert:

Wesentlicher Bestandteil des hochschulinternen Qualitätssicherungssystems ist die regelmäßige Durchführung von Evaluationen, insbesondere im Bereich der Lehre. Die Studenten und Studentinnen und die Absolventen und Absolventinnen sind bei der Evaluation an der Lehre zu beteiligen.

BERLHG (2011, §8a (1))

Evaluationen sind fast immer institutionell organisiert und diktiert (RICHARDSON, 2005, S. 409), was die Objektivität negativ beeinflussen kann (vgl. kommender Abschnitt). Einer empirischen Befragung zufolge betrieben schon vor mehr als einem Jahrzehnt 90% der Hochschulen regelmäßig interne Evaluationen (SCHWARZ-HAHN & REHBURG, 2003, S. 89), weshalb diese als Standardverfahren in Universitäten gewertet werden kann.

4.4. Anforderungen an Evaluationsinstrumente und -auswertungen

Auch wenn oft nicht von Hochschuleseite eingehalten (vgl. Abschnitt 4.2), so gibt es doch zahlreiche Anforderungen, die an ein Instrument zur Lehrevaluation an Hochschulen gestellt werden können. Vor der Neukonzeption eines solchen Instruments für Praktika (vgl. Abschnitt 10 auf Seite 107) sollten diese Anforderungen bedacht und implementiert werden, teilweise ergeben sich auch Notwendigkeiten für Testevaluationsstudien (»Validierung«) aus den Anforderungen.

Hierbei wird multiperspektivisch (RICHARDSON, 2005, S. 389; JUSSIM, 2012, S. 202; HOPBACH, 2005, S. 16) aus Sicht der Akteure der Hochschulbildung auf die Evaluation geblickt, wie es bereits im Abschnitt 2 auf Seite 19 geschehen ist. So ist der/die Studierende vielleicht eher an einem Benchmarking der Lehrveranstaltungen interessiert, um seine/ihre Kurswahlen zu optimieren, aus Universitätsperspektive ist ein Monitoring der Lehrqualität u. a. für Akkreditierungsprozesse wichtig und Dozierende wollen vielleicht eine Rückmeldung zu ihrer eigenen Lehrleistung haben (s. o.).

Anforderung 1: Theoriegeleitetheit der Konstruktion und Interpretation

Für viele Evaluationsinstrumente an Hochschulen mangelt es an einer lerntheoretischen Fundierung für die zu erfassenden lehr-, struktur- oder lernrelevanten Konstrukte (RICHARDSON, 2005, S. 388), die aber essentiell für eine geleitete Interpretation der Ergebnisse ist (THIEL et al., 2012, S. 1, 4). Es gibt daher auch Schwierigkeiten mit der Passung solcher meist sehr universell eingesetzter Evaluationsinstrumente und in Folge dessen auch mit der Akzeptanz derselben (CSONKA, 2014). Zur Theoriegeleitetheit wird hier auch hinzugezählt, dass der Evaluationsgegenstand und sein Kontext genau beschrieben werden mögen (DEGEVAL, 2008, S. 10). Dies ist in unserem Fall für Praktika bereits in den vorherigen Abschnitten geschehen.

Anforderung 2: Validität der Interpretation der Messung Nicht nur die theoretische Einbettung allein, sondern auch die Validität (Gültigkeit) von Evaluationsinterpretationen spielt eine entscheidende Rolle zu einer gezielten Diagnose der Qualität von Lehrveranstaltungen, prinzipiell sollte sich an den Gütekriterien der empirischen Sozialforschung orientiert werden (ebd., S. 11).

[...] adopting a broad construct-validation approach, recognizing [...] effective teaching [...] (is needed).

MARSH & ROCHE (1997, S. 1)

Für Evaluationsinstrumente gilt im Besonderen die Notwendigkeit valider Interpretierbarkeit der Ergebnisse, da diese häufig direkte Konsequenzen auf den Evaluationsgegenstand haben können, wenn etwa Interventionen aus den Evaluationsergebnissen abgeleitet werden. Dabei muss je gewünschter Interpretationsrichtung eigene Evidenz angeführt werden, um die anvisierte Interpretation zu rechtfertigen. Es ist also nicht ein Instrument selbst valide, sondern höchstens die anvisierten Interpretationsrichtungen, welche es vorher in einer Testevaluation zu prüfen gilt (AERA et al., 2014, S. 11).

Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests. [...] It is the interpretations of test scores for proposed uses that are evaluated, not the test itself.

AERA et al. (2014, S. 11)

Evidenz für eine solche valide Interpretierbarkeit kann aus theoretischen Überlegungen stammen (vgl. Anforderung 1) und sollte darüber hinaus aus empirischen Forschungsarbeiten im Rahmen von Testevaluationsstudien gewonnen werden (RICHARDSON, 2005, S. 404). Zu testende Interpretationsrichtungen ergeben sich dann aus dem gewünschten späteren Gebrauch des Instruments. Will man etwa einen bestimmten Aspekt der Lehre lückenlos und klar interpretieren können, so muss das betreffende Konstrukt im Rahmen seiner Operationalisierung im Instrument auf »Construct irrelevant variance« (Messick, 1995), also konstrukt fremde Einflüsse und »construct underrepresentation« (ebd.), das »Vergessen« von wesentlichen Aspekten, hin untersucht werden.

Möchte man die verschiedenen Aspekte der Lehre getrennt voneinander interpretieren können, so gilt es, diese Trennbarkeit theoretisch plausibel zu machen und empirisch abzusichern. Weitere Interpretationsrichtungen samt entsprechender empirischer Studien finden sich auch im Abschnitt 12 auf Seite 123, eine komplette Übersicht liefern AERA et al. (2014).

Anforderung 3: Ökonomie Ein wesentlicher Punkt betrifft die Ökonomie der Befragung selbst. In der Regel sind Standard-Evaluationsformen ökonomisch einsetzbar (RICHARDSON, 2005, S. 388) und somit eine repräsentative Befragung der teilnehmenden Studierenden möglich. Dies sollte auch bei theoriebasierten und validierten Evaluationsinstrumenten erhalten bleiben. Einen Schritt weiter geht die Deutsche Gesellschaft für Evaluation: Sie betont in ihrem Standard »Durchführbarkeit« die Einhaltung des Aufwand-Nutzen-Prinzips, also dass der Zeitaufwand und die Belastung der Evaluationsbeteiligten in einem angemessenen Verhältnis zur Menge der zu erhaltenen Information stehe (DEGEVAL, 2008, S. 9).

Anforderung 4: Standardisierung Standardisierung ist in der Regel eine Stärke bestehender Evaluationsinstrumente (»formal instruments«, RICHARDSON, 2005, S. 401). So wird die Evaluation verschiedenster Kurse und verschiedenster Fächer mitunter mit nur einem einzigen Lehrevaluationsinstrument vorgenommen (ebd., S. 404), was universitätsweit einer maximalen Standardisierung und damit Vergleichbarkeit der Lehrevaluation gleichkommt. In den USA beispielsweise geht dies tatsächlich so weit, dass universitätsübergreifend dieselben Evaluationsinstrumente eingesetzt werden (ebd.). Die Passung der Evaluationsinstrumentes zur jeweiligen Lehrveranstaltung sollte dabei aber erhalten bleiben, um aussagekräftige Ergebnisse zu erhalten. Die Standardisierung bietet eine Möglichkeit der Erhebung der gesamten interessierenden Population und damit eine systematische Dokumentation der Ergebnisse, sofern sie mehrheitlich quantitativ vorgenommen wird.

4. *Lehrevaluation an Hochschulen*

Qualitative Methoden (etwa Studierenden-Interviews oder Fragebögen mit offenem Antwortformat) gehören indes eher nicht zum Evaluationsstandard, vielleicht da sie einen Mehraufwand in der Auswertung bedeuten:

... the burden of analyzing open-ended responses and other qualitative data is immense, even with only a relatively modest sample.

RICHARDSON (2005, S. 401)

Dies bedeutet im Umkehrschluss allerdings nicht, dass qualitative Methoden für den Bereich der Lehrevaluation an Hochschulen gänzlich ungeeignet sind. Qualitative Daten eignen sich besonders gut zur tiefenanalytischen Erfassung von Gegebenheiten in einem bestimmten, nicht verallgemeinerbaren Kontext (BORTZ & DÖRING, 2006, S. 301–302). Für eine inhaltliche Analyse von Lehrveranstaltungen eignen sich also vor allem qualitative Daten.

Will man auch die Vorteile quantitativer Erhebungen (Ökonomie, Standardisierung, Repräsentativität, Verallgemeinerbarkeit, etc.) nutzen, so überführt man anschließend klassischerweise die Erkenntnisse der qualitativen Forschung als Hypothesen in ein quantitatives Messinstrument (»Vorstudienmodell«, MAYRING, 2001). Auch RICHARDSON (2005, S. 402) betont, dass qualitative Erhebungen die inhaltliche Basis für quantitative Lehrevaluationen liefern können. Andersherum können qualitative Methoden auch dazu genutzt werden, eine quantitativ gewonnene Erkenntnis im Hinblick auf den dahinterliegenden Prozess zu analysieren (»Vertiefungsmodell«, MAYRING, 2001).

Anforderung 5: Externe Evaluation und Anleitung für Lehrende Bei aller Messgenauigkeit und Gültigkeit der Interpretation von Evaluationsmessungen darf nicht außer Acht gelassen werden, dass diese Ergebnisse auch adäquat an die Adressaten weitergetragen werden müssen, um die Ergebnisse der Evaluation auch in zukünftigen Durchführungen der Lehrveranstaltung berücksichtigen zu können und zur Reflexion anzuregen. Dass dies teilweise im aktuellen Evaluationsverlauf nicht ausreichend zu geschehen scheint, zeigen u. a. auch Studierenden Sorgen darüber, ob die Evaluation überhaupt von den Dozierenden ernst genommen wird (RICHARDSON, 2005, S. 407).

Auf Lehrendenseite dagegen werden Evaluationsergebnisse meist als nützliche Information gewertet (SCHMELKIN et al., 1997). Die Ergebnisse sollten zudem für die Lehrenden aufbereitet werden, da sonst mögliche Veränderungen der Lehrrealität meist ausbleiben und auch eine Weitergabe an institutionelle Mitglieder ausbleibt

(NASSER & FRESKO, 2002). NEUMANN (2000) spricht sich daher für Unterstützung für die Lehrenden aus, wie Ergebnisse interpretiert werden müssten.

Dies spricht einerseits für eine institutionell gesteuerte Evaluation durch Externe, andererseits für eine angeleitete Interpretation der Ergebnisse für Lehrende. Es müssen zudem, um aus Evaluationsergebnissen auch Interventionen zu generieren, die von den Lehrenden selbst kommen, Anreize geschaffen werden, eben jenes zu tun (KEMBER et al., 2002).

Anforderung 6: Veröffentlichung der Ergebnisse für Studierende An Evaluationserhebungen teilnehmende Studierende investieren Zeit in die Bearbeitung und erwarten daher, dass ihre Evaluationen auch gesehen und ernst genommen werden. Um dies zu gewährleisten, müssen Evaluationsergebnisse transparent gemacht werden (RICHARDSON, 2005, S. 408).

Anforderung 7: Benchmarking Langfristig kann durch standardisierte Evaluationsbefragungen ein Benchmarking, also ein zielgerichteter Vergleich von Lehrqualität vorgenommen werden. So können interuniversitär Lehrqualitäten verschiedener Fachbereiche miteinander verglichen werden, etwa um best-practice-Beispiele herauszustellen oder innerhalb eines Fachbereichs verschiedene Dozierende im Rahmen ihrer Lehre zu vergleichen. Dies könnte einen gegenseitigen Austausch gezielter initiieren. Aber auch extrauniversitär kann ein Wert im Benchmarking liegen, etwa wenn die Lehrqualität eines bestimmten Standorts als herausragend erkannt und von diesem profitiert wird.

Selbst das einfache Einordnen der eigenen Lehrqualität in einen Referenzdurchschnitt kann gewinnbringende Erkenntnisse liefern. So »haben die Evaluationsverfahren auch eine stark nach außen gewandte Funktion. Sie werden zunehmend in koordinierten und tendenziell standardisierten Benchmarkingverfahren durchgeführt« (HOPBACH, 2005, S. 16).

Anforderung 8: Reliabilität der Messung Für die Interpretation der Konstrukte der Evaluationserhebung sollte dem/der Interpretierenden bewusst sein, welche Messgenauigkeit hinter dem zu interpretierenden Zahlenwert steht (DEGEVAL, 2008, S. 11). Wird das Ergebnis einer Konstruktmessung mit besonders niedriger Messgenauigkeit interpretiert, so ist Vorsicht und Zurückhaltung vor großen Schritten geboten. Kritisch an den meisten Reliabilitätsmaßen ist, dass sie mit der inhaltlichen Itemähnlichkeit zunehmen (Homogenität).

So kann ein facettenreich operationalisiertes Konstrukt oftmals eine niedrigere Reliabilität aufweisen als dasselbe Konstrukt ohne Facettenreichtum. Ersteres ist aber deutlich plausibler inhaltlich interpretierbar, da es das beschriebene Konstrukt

auch tatsächlich abzubilden versucht (vgl. Anforderung 2). Dies ist als »Reliabilitäts-Validitäts- Dilemma« bekannt (ROST, 1996, S. 392ff).

Anforderung 9: Objektivität der Messung Für die Objektivität wird zwischen Durchführungs-, Auswertungs- und Interpretationsobjektivität unterschieden (BORTZ & DÖRING, 2006, S. 195). Die Durchführungsobjektivität betrifft die Erhebung selbst und ob der Erhebende oder begleitende Umstände die Messung stören. So sollten sich Erhebende so gut wie möglich neutral verhalten und Erhebende an unterschiedlichen Orten ein konsistentes Verhalten zeigen. Erreicht werden kann dies über Erhebungsmanuale, die Einleitungsworte, Reaktionen auf mögliche Nachfragen und Hinweise zum Datenschutz beinhalten können. Die Auswertungsobjektivität betrifft die Datenaufbereitung und die Wahl der Auswertungsmethoden. Auch diese sollten bei einer Lehrevaluation standardisiert werden.

Schließlich betrifft die Interpretationsobjektivität den Umgang mit den Ergebnissen in Bezug auf die Schlussfolgerungen und Interventionen, die daraus abgeleitet werden, was es ebenfalls zu standardisieren gilt.

Anforderung 10: Nützlichkeit Gute Lehrevaluationen implizieren den Gewinn an Informationen, die am Bedarf der Betroffenen orientiert sind (DEGEVAL, 2008, S. 8). Dazu müssen zunächst die Betroffenen identifiziert und die Evaluationszwecke geklärt werden. In Übereinstimmung mit den Anforderungen 5 und 6 sollte nicht nur der Nutzen, sondern auch die Nutzung der Informationen gewährleistet sein ebd., S. 8–9).

Eine weitere Anforderung der DEGEVAL, die Fairness (ebd., S. 10) im Umgang mit von der Evaluation betroffenen Personen, wird hier nicht weiter ausgeführt, da sie sich von selbst verstehen sollte.

4.5. Standard-Evaluationsinstrumente für die Hochschullehre

Die folgende Auswahl an standardisierten und teils auch theoriegeleitet entwickelten Evaluationsinstrumenten für Hochschulen soll exemplarisch demonstrieren, wie viel bereits in diesem Bereich an Vorarbeit geleistet wurde.² Die Perspektiven der einzelnen Instrumente unterschieden sich hierbei in eine traditionelle Sicht auf die lehrende Person (Qualitätsdimension *Lehrkompetenz*), eine auf die Lernumgebung (Qualitätsdimension *Lernumgebung* bzw. *Strukturdaten*) und eine auf den Lernzuwachs der Studierenden (Qualitätsdimension *Lernzuwachs*, vgl. Abschnitt 2.2 auf Seite 20). Im Rahmen der Kompetenzorientierung gewinnt letztere Perspektive an Bedeutung, da output-orientiert im Sinne von Bologna evaluiert wird. Der Blick soll

²Eine ausführliche Übersicht zu Lehrevaluationsinstrumenten im deutschsprachigen Raum liefern BRAUN & GUSY (2006, AS. 154–155).

hierbei sowohl auf nationale wie auch internationale Instrumente gerichtet werden, obwohl die Ausbildungssysteme nicht unmittelbar vergleichbar sind (HEINE et al., 2006). Abschließend wird das einzig in der Literatur verfügbare internationale Evaluationsinstrument für Praktika vorgestellt und in die drei Qualitätsdimensionen eingeordnet.

4.5.1. Evaluation des Lehrens (SEEQ und LeKo)

Prozessdaten beschreiben den Verlauf einer Lehrveranstaltung, d. h. ganz wesentlich die Gestaltung der Lehrveranstaltung durch die jeweilige Dozentin

BRAUN et al. (2008, S. 32).

International eingesetzt (z. B. in Nordamerika, Australien und Spanien) und gut erforscht ist der SEEQ-Fragebogen (Student Evaluation of Educational Quality Instrument) nach MARSH (1982). Er besteht aus 35 Studierenden-Rating-Items und ist faktoriell validiert und reliabel (MARSH, 1982, 1987; MARSH & BAILEY, 1993; MARSH & ROCHE, 1997). Ersteres bedeutet hierbei, dass Testevaluationsstudien durchgeführt wurden, in denen die Struktur der Items des Instruments untersucht wurde. Eingeteilt in sogenannte Faktoren, die Konstrukte widerspiegeln sollen, ist eine getrennte Interpretation der Konstrukte des Fragebogens demnach zulässig. Inhaltlich umfasst der SEEQ neun Aspekte erfolgreicher Lehre, darunter *Lernen*³, *Enthusiasmus des/der Lehrenden*, *Organisation*⁴, *Gruppeninteraktion*⁵, *persönliche Beurteilung*⁶, *Inhaltliche Breite des Lehrspektrums*, *Benotung*⁷, *Aufgaben*⁸, sowie *Workload & Schwierigkeit* (Marsh, 1982).

Die Forschungsergebnisse zum SEEQ legen nahe, dass »erfolgreiche Lehre«, so wie sie durch das Instrument selbst definiert wird, vor allem von dem/der Dozierenden abhängt und weniger von der entsprechenden Kurseinheit bzw. dem Modul (RICHARDSON, 2005, S. 389). Das gesamte Instrument fokussiert also mehr oder minder auf die Qualitätsdimension *Lehrkompetenz* (vgl. Abschnitt 2.2 auf Seite 20), wobei der Faktor *Lernen* eher der Dimension *Lernzuwachs* und der Faktor *Aufgaben* auch der *Lernumgebungs*-Dimension zugeordnet werden kann. Es korrelieren beim SEEQ die Selbsteinschätzungen der eigenen Lehrqualität von Dozierenden mit

³Dem Gelernten einen Wert zuweisen.

⁴Strukturiertheit des/der Lehrenden

⁵Lehrende/r regt zu Diskussionen und Kooperation an.

⁶Lehrende/r bewirkt eine angenehme Lernatmosphäre.

⁷Konstruktives Feedback und Fairness der Benotung

⁸(Haus-)Aufgaben trugen zum Verständnis bei.

4. Lehrevaluation an Hochschulen

Fremdeinschätzungen der zugehörigen Studierenden, wonach bereits zwei Perspektiven in der Interpretation zulässig sind (Multiperspektivität, RICHARDSON, 2005, S. 389). Bislang wurde trotz der Realisierung des Instruments als Rating-Skala nur von geringen Verzerrungseffekten berichtet (RICHARDSON, 2005, S. 390).

Der SEEQ weist eine explorativ gewachsene theoretische Grundlage auf (»logical analysis«, MARSH, 1987, S. 264), die über Rückmeldungen der Evaluationsbeteiligten immer weiter optimiert wurde und daher als zielgruppenspezifisch, aber wenig theoriegeleitet im Sinne einer theoretischen Modellierung bezeichnet werden kann (MARSH, 1987, S. 264–265). Initiativ wurden durch Review erhaltene neunzehn Dimensionen guter Dozierendenleistung angesetzt, dann mehrfach empirisch u. a. faktorenanalytisch erforscht, teils fusioniert und ergänzt. Schließlich ergaben sich die neun oben genannten Aspekte. Durch die vielen Vergleichsstudien, die mit dem SEEQ durchgeführt wurden, kann sein Testevaluationsstatus und damit die Validität der Interpretation seiner Testergebnisse (Anforderung 2 gemäß 4.4) als sehr gut bezeichnet werden (für Einzelheiten hierzu vgl. MARSH, 1987, S. 264). Ökonomie und Standardisierung sind gut, da es sich um einen klassischen geschlossenen Rating-Fragebogen handelt. Alle wesentlichen Anforderungen an Evaluationsinstrumente werden erfüllt, wenngleich nur eine von drei möglichen Qualitätsdimensionen guter Lehre umfassend abgedeckt wird.

Ein anderes im deutschsprachigen Raum aktuell verbreitetes Lehrevaluationsinstrument ist der Fragebogen zur Erfassung der Lehrkompetenz (LeKo, THIEL et al., 2012). Dieser basiert auf einem durch Review erarbeiteten Modell der Unterrichtsqualität bezogen auf die Kompetenzen des/r Lehrenden, die besonders im Hochschulbereich bei Vorlesungen und Seminaren (ebd., S. 10) gefordert sind. Es wird also bewusst etwa das selbstgesteuerte Lernen in Gruppen oder problemorientierte Lernarrangements ausgeklammert (ebd., S. 10–11). Auf Basis der theoretischen Herleitung ergeben sich drei trennscharfe Lehrkompetenz-Dimensionen: (A) *Vermittlung von Wissen und Unterstützung von Verstehen (Unterstützung der Wissenskonstruktion)*, (B) *Motivieren und lerndienliche Atmosphäre herstellen (Motivierung)* und (C) *Steuerung des Lernprozesses in der Lerngruppe (Steuerung der Interaktion)*. Jede dieser Dimensionen umfasst wiederum Subdimensionen, die sich auch empirisch als Faktoren ergaben.⁹ Insgesamt ergeben sich so 15 Konstrukte der *Lehrkompetenz*. Die betreffenden Skalen wiesen darüber hinaus eine gute bis sehr gute Reliabilität von Cronbachs alpha zwischen .72 und .91 auf (ebd., S. 13). Analysen zur Eignung des Instruments über Fächergruppen hinweg (Messinvarianz) und zur Konstrukt-

⁹z. B. zu (A) *klar und strukturiert darstellen*, zu (B) *gute Lernatmosphäre herstellen* und zu (C) *Zeitmanagement*

validität stehen indes noch aus (ebd., S. 22). Zudem wird einschränkend darauf hingewiesen, dass die Inhalte des Fragebogens eine Schnittmenge an Anforderungen für das Lehren in Vorlesungen (vgl. 3.1) und Seminaren (vgl. 3.2) bilden, die Konstrukte also für anders geartete, kooperativere Lernarrangements zu ergänzen und zu verändern sind (ebd., S. 22).

Die Anforderungen an gute Evaluationsinstrumente erfüllt der LeKo gut¹⁰ (vgl. Abschnitt 4.4). Lediglich bei der Validität (2) der Interpretationen sind weitere Testevaluationsstudien von Nöten, wodurch auch ein Benchmarking (7) bisher fragwürdig erscheint.

Vergleichend mit dem SEEQ ist der LeKo als umfangreicher und in seiner Konstruktion theorieorientierter zu beschreiben. Allerdings weist der SEEQ auf Grund seines jahrzehntelangen Einsatzes und diverser Vergleiche mit lehrrelevanten Konstrukten einen hervorragenden Stand der Testevaluation und damit Interpretierbarkeit auf, den der LeKo noch nicht erreicht hat.

4.5.2. Evaluation des Lernens (BEvaKomp)

Da die Output-Orientierung der Lehre nach Bologna eine neuere Entwicklung darstellt, ist es nicht verwunderlich, dass erst wenige Lehrevaluationsinstrumente diese Perspektive einnehmen und nach dem Erwerb von Kompetenzen fragen. Eine Ausnahme bildet das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp) nach BRAUN & GUSY (2006). Der BEvaKomp orientiert sich explizit am Kompetenzbegriff.

Das Modell definiert die Kompetenzbereiche *Fachkompetenz*, *Methodenkompetenz* (*Vorgehensweisen und Verfahren*), *Sozialkompetenz* (*Kommunikation und Kooperation*) und *Personalkompetenz* (*Kritikfähigkeit, Kreativität, Selbstorganisation, Motivation und Selbstwirksamkeit*) (BRAUN & GUSY, 2006, S. 162). Im Instrument BEvaKomp selbst finden sich allerdings als Ergebnis von Testevaluationsstudien und resultierenden Itemausschlüssen deutlich weniger Konstrukte, nämlich *Fachkompetenz*, *Methodenkompetenz*, *Präsentationskompetenz*, *Kommunikationskompetenz*, *Kooperationskompetenz* und *Personalkompetenz* (letzteres meint *Interesse*). Diese werden mittels Selbsteinschätzung des eigenen Kompetenzzuwachses in der betreffenden Veranstaltung durch die Studierenden erfasst. Empirische Evidenz für die Validität von Testinterpretationen des BEvaKomp lieferten Studien zur Relia-

¹⁰Die Theoriegeleitetheit ist mit dem LeKo-Modell erfüllt, die Ökonomie mit der Realisierung als Rating-Fragebogen ebenso. Eine Standardisierung und ein Benchmarking liegen nun vor (BLÜTHMANN et al., 2017). Die Ergebnisse werden indes leider nicht veröffentlicht, sondern den Dozierenden zur Verfügung gestellt. Reliabilität und Objektivität sind gut, die Nützlichkeit als Rückmeldeinstrument für Dozierende in Vorlesungen und Seminaren gegeben.

bilitätsprüfung, Konstruktvalidität und diskriminanter Validität. Die Testung der »Konstruktvalidität« geschah mittels konfirmatorischer Faktorenanalyse der Testergebnisse bei Studierenden unterschiedlicher Lehrveranstaltungen. Die Korrelationen zwischen den Faktoren fielen hierbei im Bereich .19 bis .76 aus, was empirisch für eine trennbare Interpretation spricht.

Zudem wurde eine Vergleichsmessung mit der Studienzufriedenheit vorgenommen (»diskriminante Validierung«), die messbare, aber moderate Zusammenhänge mit den Kompetenzselbsteinschätzungsskalen aufwies. Die Testergebnisse lassen sich demnach über bloße Zufriedenheit hinaus interpretieren (BRAUN et al., 2008, S. 39). Vergleiche mit anderen Kriterien, wie Studienleistungen, stehen noch aus. Eine Interpretation als Kompetenzzuwachs im Sinne einer zu Studienleistungen befähigenden Voraussetzung ist also noch nicht abgesichert. Die Reliabilitäten der Skalen fallen gut aus (ebd., S. 36). Zusammenfassend lässt sich festhalten, dass der BEvaKomp wie auch der LeKo die meisten Anforderungen an Evaluationsinstrumente erfüllen (vgl. Abschnitt 4.4). Lediglich in den Dimensionen (2) »Validität der Messung« und (7) »Benchmarking« besteht noch Entwicklungsbedarf.

4.5.3. Evaluation der Struktur einer Lehrveranstaltung (CEQ)

Instrumente, die eine Lernumgebung ausschließlich nach ihrer Struktur evaluieren, ließen sich nicht finden. Allerdings gibt es mit dem »in Course Experience Questionnaire-Instrument« (CEQ, RAMSDEN, 1991) eines, das aus Institutionsperspektive entwickelt wurde und daher eine stark strukturell geprägte Sicht beinhaltet, wenn auch nicht ausschließlich (vgl. Abschnitte 2.2 und 4.3).

Seine inhaltlichen Ursprünge verdankt das Instrument wie viele andere auch einer Expert*innengruppe, die aus Vertreter*innen der Wirtschaft und Lehre zusammengesetzt war und auf Indikatoren guter Lehre an Hochschulen definierte. Das Instrument wurde vornehmlich in Australien eingesetzt und bereits mit über 80.000 Studierenden erprobt (RICHARDSON, 2005, S. 394). Seine 30 Items in fünf Skalen erfassen ökonomisch Aspekte »guter Hochschullehre« mit befriedigender Reliabilität (ebd., S. 396). Konstrukte sind auf Seite der Lehrenden: *Gute Lehre* im Sinne hilfreichen Feedbacks und die Betonung von Unabhängigkeit. Die strukturelle Prägung findet man bei Skalen wie *Klare Ziele und Standards*, *angemessener Workload* und *angemessene Prüfungen* (ebd., S. 393).

Diese sind faktoriell untersucht und RASCH-analysiert worden, mit geringen Nebenladungen und einer insgesamt zufriedenstellenden Struktur (ebd., S. 396). Nach klassischer Testevaluation wurde damit die »Konstruktvalidität« als zufriedenstellend beschrieben (ebd., S. 396). Die signifikanten Korrelationen zwischen den Ska-

lenausprägungen und der Zufriedenheit mit der Lehrveranstaltung insgesamt wurden erstaunlicherweise als Evidenz für »kriteriale Validität« gewertet. Es kann aber auch als Evidenz für meist ungewünschte affektive Einflüsse des Subjekts auf die Skalenausprägungen gesehen werden, was gegen eine klare Interpretierbarkeit der Skalenwerte in Bezug auf ihren Inhaltsbereich spricht.

Für die Untersuchung der »diskriminanten Validität« ergaben sich höhere CEQ-Scores bei denjenigen Fachbereichen, die eher auf problembasiertes Lernen setzten, was als Evidenz für die Interpretierbarkeit der hohen CEQ-Scores als »Gute Lehre« dienen kann. So ergibt sich insgesamt für den CEQ ein eher mäßiges Bild, um es als Vorbild für ein Evaluationsinstrument für Praktika zu verwenden.

4.5.4. Evaluationsinstrument für Praktika (SLEI)

Der SLEI (Science Laboratory Environment Inventory)-Fragebogen (FRASER & McROBBIE, 1995) ist das einzige internationale Instrument zur Evaluation von Praktika.¹¹ Die Skalen des SLEI sind *Studierenden-Zusammenhalt*, *Offenheit*, *Integration*, *Regelklarheit* und *materielle Umgebung*.

Dabei gingen FRASER & McROBBIE (1995) in der Auswahl der Konstrukte von einem Literatur-Review von vier Quellen zu Praktika aus den Jahren 1990–1991 aus. Der *Studierenden-Zusammenhalt* beschreibt das Ausmaß, zu dem sich die Studierenden gegenseitig helfen, ist also verwandt mit der *Kooperationskompetenz* und damit in die Ziele von Praktika klar einzuordnen und als Konstrukt des *Lernzuwachses* zu verstehen. *Offenheit* meint ganz im Sinne der Typisierung von Praktika nach DOMIN (1999, vgl. 3.1 auf Seite 35) die Offenheit der Durchführung, also »student generated« beim Deskriptor »procedure« und verfolgt damit das Ziel der Messung, ob es sich um ein problemorientiertes Praktikum handelt (DOMIN, 1999, Typ »problem-based«), ist also ein strukturelles Maß der *Lernumgebung*. Stellt man sich das Praktikum nicht als isolierte Lehrveranstaltung vor, sondern als eingebunden in ein Studium und damit in wechselseitiger Beziehung zu anderen Lehrveranstaltungen, so stellt die *Integration* den Grad dar, zu dem sich Vorlesung und Praktikum inhaltlich abstimmen, ist also ein an Organisation interessiertes Maß, daher den Strukturdaten bzw. der *Lernumgebung* zuzuordnen.

Regelklarheit meint das Ausmaß, zu dem die Studierenden eine Klarheit der Verhaltensregeln im Praktikum wahrnehmen. Dies ist auch ein eher strukturell-

¹¹Der FEPRA nach STAUFENBIEL (2000) ist das einzige nationale Pendant. Seine Skalen wirken allerdings zumeist einem allgemeinen Evaluationsinstrument für Vorlesungen entnommen. Tatsächlich finden sich beträchtliche Überschneidungen zum FEVOR-Fragebogen für Vorlesungen (ebd.).

4. Lehrevaluation an Hochschulen

organisatorisches Konstrukt, auch wenn die Betreuenden für dieses Maß eine exekutive Rolle einnehmen, daher auch die *Lehrkompetenz* involviert ist.

Mit *materielle Umgebung* schließlich ist gemeint, ob die räumlichen und geräte-technischen Voraussetzungen für das Absolvieren des Versuchs gegeben sind, also klar der *Lernumgebungsdimension* zuzuordnen. Die Items und Skalen wurden mehrschrittig einer Testevaluation unterzogen und zum in 6 Ländern ($N > 5000$) eingesetzt (FRASER & MCROBBIE, 1995, S. 294). Bezüglich der Theoriegeleitetheit (Anforderung 1, vgl. Abschnitt 4.4) der Konstruktion weist der SLEI eine Literaturliteraturbasis mit vier Quellen auf, die heute bereits über 25 Jahre alt sind. Die nur fünf Qualitätskonstrukte erscheinen hinsichtlich der Komplexität von Praktika (vgl. Abschnitt 3.4) nach Ansicht des Autors zu wenig für eine adäquate Evaluation. Bezüglich der Validität, Ökonomie, Standardisierung (Anforderungen 2–4), des Benchmarkings, der Reliabilität und Objektivität (Anforderungen 7–9) besticht der SLEI mit vielen Testevaluations-Studien und großen Proband*innenzahlen.

Auf Basis der veralteten Basisquellen und einer deutlich vielfältigeren Praktikumskultur im angloamerikanischen Raum (LAGOWSKI, 2002) erscheint allerdings die Nützlichkeit (Anforderung 10) im Sinne einer Bedarfsorientierung an den Praktika im deutschsprachigen Raum nicht erfüllt, zumal das Konstrukt Offenheit in Anfängerpraktika sehr sicher lediglich einen Bodeneffekt feststellen würde (»expository-style«, vgl. Abschnitt 3.4.4).

Kapitel 5

Forschungsstand: Lehrformat Naturwissenschaftliche Praktika

Nachdem ein Einblick in die Lehre und die Lehrevaluation an Hochschulen gegeben wurde und bereits erste Aspekte der Umsetzung des Lehrformats »naturwissenschaftliches Praktikum« anklagen, wird nun auf den Forschungsstand zu eben jenem Lehrformat eingegangen. Gemäß der Ausrichtung dieser Arbeit wird dies vorrangig aus der Perspektive der Physikdidaktik geschehen, aus der Chemiedidaktik lassen sich aber Großteils übertragbare Schlüsse ziehen (vgl. ELLIOTT et al., 2008; FRESE, 2001; ENGELS & HÜLSBUSCH, 2008; ROLLNICK et al., 2002; BOND-ROBINSON & RODRIQUES, 2006).

Ziel dieses Abschnitts ist es auch, mögliche Qualitätsindikatoren für Praktika aus dem bisherigen Forschungsstand abzuleiten. Dazu werden zunächst empirische Erkenntnisse zu Charakteristika von Praktika vorgestellt. Es werden darüber hinaus die erforschten Probleme berichtet. Diese bieten einen Kontrast zu den bereits dargestellten Zielen und Legitimationsgründen (vgl. Abschnitt 3.4.2), da sich meistens auch Interventionen anschlossen, die wiederum in ihrer Wirksamkeit in Praktika erforscht wurden. Um an die bestehenden Forschungsarbeiten anzuknüpfen, werden zudem ausblicksartig Schlussfolgerungen für mögliche zukünftige Interventionen in Praktika resümiert. Insgesamt wird hierbei nur wenig auf internationale Publikationen Bezug genommen, da die unterschiedlichen Bildungssysteme eine schlechte Übertragbarkeit der Erkenntnisse zeigen (HEINE et al., 2006).

Dem einfach gegliederten deutschen Hochschulsystem mit überwiegend staatlichen Einrichtungen steht ein hochdifferenziertes US-amerikanisches System gegenüber, welches neben den staatlichen auch eine Reihe von privaten und kommerziellen Instituten aufweist. Die sogenannten *research universities* entsprechen in etwa der deutschen »vollakademischen« Universität, welche Forschung und Lehre integriert. Nur ca. 15 % der US-amerikanischen Studierenden sind an einer solchen Universität eingeschrieben.

FREYER (2013, S. 33)

5.1. Charakteristika von Praktika

Wurde im Abschnitt 3.4 lediglich auf die zwar literatur-, aber dort nur erfahrungsbasiert geschilderte Praktikumssichtstruktur eingegangen, so werden hier die wenigen empirischen und theoretisch abgeleiteten Erkenntnisse zu Charakteristika und Mechanismen in Praktika vorgestellt, ohne zu stark auf deren lerntheoretische Konsequenzen einzugehen.¹ Allerdings ist der Forschungsstand recht überschaubar, so dass auch diese Ergebnisse nicht als gesichert angenommen werden, sondern lediglich Hinweise bezüglich des Praktikumsgeschehens und damit einer theoretischen Modellierung von Praktikumsqualität (siehe Abschnitt 7) geben können.

Gemäß dem vorherrschenden traditionellen Praktikumstyp »expository stlye« (vgl. Abschnitt 3.4.4) verwundert es wenig, dass vor allem regulierende und stark anleitende Charakteristika die Vorbereitungsphase ausmachen. So wird in dieser Phase dem Skript eine große Rolle zuteil, da es meist die einzige Vorbereitungsquelle für die Studierenden darstellt (ENGELS & HÜLSBUSCH, 2008, S. 61) und deren Lernen reguliert. Es konnten diesbezüglich in einer Videostudie in Physikpraktika sogar mehr Wissenszuwachs für die Beschäftigung mit dem Skript festgestellt werden als beim Absolvieren der Versuchsdurchführung selbst (HUCKE, 1999, S. 126, 132, 149). Das Skript wird dabei nicht nur als lernregulierend, sondern auch während der Durchführungsphase als handlungsleitend erlebt (ROLLNICK et al., 2002, S. 1060; HUCKE, 1999, S. 149; SANDER, 2000), da es in einer Doppelrolle als Vorbereitungswerk mit kompakten theoretischen Hintergründen zum Versuch, sowie als Versuchsanleitung im Sinne schrittweiser Arbeitsanweisungen fungiert. Die *Skriptqualität* lässt sich der Lehrqualitätsdimension *Lernumgebung* zuordnen.

Die Motivation für die Vorbereitung kann indes nur theoretisch abgeschätzt werden, so vermutet NAGEL (2009, S. 16), dass durch die Eingangsprüfungen vor den Versuchen die Beschäftigung mit den Praktikumsinhalten vor allem extrinsisch motiviert ist. Eine Studie mit Lehramtsstudierenden bestätigte diese Vermutung, wenngleich auch intrinsisch-motivationale Aspekte vertreten waren (ebd., S. 84). Auch NEUMANN (2004, S. 164) führte hierzu bereits eine Debatte um das Für und Wider des häufigen Prüfens in Praktika mit Bezug auf bildungswissenschaftliche Theorien zum Lern- und Leistungsraum. HUCKE (1999, S. 150) macht sich ebenfalls für eine Erforschung affektiver Variablen in Praktika stark, da in seiner Studie kein Zusammenhang von Handlung und Lernen messbar war und er daher affektive Variablen als entscheidend vermutet.

¹Dies geschieht im Abschnitt 14.4.2 auf Seite 244 genauer.

Die Durchführung ist empirisch stärker beforscht als die Vor- und Nachbereitungsphasen. Dies lässt sich dadurch begründen, dass Forschende deutlich einfacheren Zugang zur Präsenzzeit erhalten, als zu Zeiten, wo die Studierenden sich im Selbststudium befinden. HUCKE (1999) untersuchte in seiner Dissertation im Rahmen von Videoanalysen des Praktikumsgeschehens in Physikpraktika, welche Anteile die jeweiligen Tätigkeiten und Gedanken der Studierenden am gesamten Versuchsgeschehen einnehmen. Die Gedanken wurden je aus den Sprechanteilen rekonstruiert (HUCKE, 1999, S. 145).

Er kam zu dem Schluss, dass zu über 80% der Praktikumspräsenzzeit während des lauten Denkens nicht über physikalische Konzepte gesprochen wurde.² Stattdessen stand das eher mechanistisch geprägte Aufnehmen von Messdaten im Vordergrund (ebd., S. 148), was sich erneut über extrinsische Anreize erklären lässt.³ Auch TIBERGIEN et al. (1998, S. 495) befinden empirisch, dass Messen und Rechnen die Haupttätigkeiten im Praktikum darstellen. Gespräche laufen daher fast ausschließlich auf der deskriptiven (Beschreibungen, Geräte) statt der abstrakten (Relationen, Theoriebezug) Ebene ab (FISCHER, 1994), es kommt zu wenig Handlungsspielraum und vielen Routinetätigkeiten (HUCKE, 1999, S. 148). Eine aktive Hypothesenbildung, wie sie in der Theorie-Praxis-Verknüpfung und im Erkenntnisgewinnungsprozess als unabdingbar angesehen wird (SANDER, 2000, S. 32), kommt selten vor (TIBERGIEN et al., 1998, S. 496). Die Betreuendenrolle wird als Hauptinitiator verstanden (TIBERGIEN et al., 1998, S. 500) und Betreuer*innen-Trainings für notwendig befunden (SERE, 2002), die konkreten Anforderungen an Betreuende sind allerdings empirisch noch nicht untersucht.

5.2. Probleme von Praktika

Um nun auf lerntheoretische Konsequenzen der Gestaltung von und Geschehnisse in Praktika einzugehen, wird, wie in der Literatur üblich (NEUMANN, 2004; SCHREIBER, 2012), in die Phasen Vorbereitung, Durchführung und Nachbereitung für die Problemdarstellung unterschieden, nebst phasenübergreifender Probleme. Für einen Überblick über typische Tätigkeiten innerhalb dieser Phasen sei auf den Abschnitt 3.4 auf Seite 28 und den vorherigen Abschnitt verwiesen.

5.2.1. Phasenübergreifende Probleme

Sie [die Praktika] zeigen sich in ihrer Ausgestaltung gegenüber neueren didaktischen Erkenntnissen jedoch erstaunlich änderungsresistent.

²Hucke bezeichnet dies als »handlungsleitende Kognitionen« (ebd., S. 148).

³Das ordentlich geführte Messdatenprotokoll gilt als implizites Ziel der Präsenzzeit, da es meist auch Bewertungsgrundlage darstellt.

DIEMER et al. (1998, S. 3).

Mit der Darstellung der phasenübergreifenden Probleme wird ein Zustand der Praktika beschrieben, der trotz teils älterer Quellen wohl so heute noch bestand hat (vgl. Zitat DIEMER et al., 1998).

Probleme in physikalischen Grundpraktika werden, vor allem von Lehrenden, häufig durch mangelnde Voraussetzungen der Studierenden beschrieben, auf die eine didaktisch ausgerichtete Lehrveranstaltung natürlich Rücksicht nehmen muss. Die Studierenden kommen demnach mit zu geringen experimentellen Grundkenntnissen in das Praktikum und das unabhängig vom Studiengang (KREITEN, 2012, S. 42–43; NEUMANN, 2004, S. 11; PLOMER, 2011, S. 4–6). Auch die theoretischen Vorkenntnisse werden als zu gering eingestuft, vor allem für die Nicht-Physikstudierenden in Physikpraktika (KISSMANN et al., 2010; KREITEN, 2012, S. 41–43; THEYSSEN, 2000, S. 5). Vor allem in der Medizin und Biologie wird eine zu geringe Anpassung an die Heterogenität der Studiengänge kritisiert, nebst geringer Vernetzung der fachphysikalischen Inhalte mit dem jeweiligen Studienfach (BORAWSKI & HEINKE, 2005; KISSMANN et al., 2010; PLOMER, 2011, S. 62–64; THEYSSEN, 2000, S. 4–5). Letzteres ist vor allem aus Sicht der Theorie-Praxis-Verknüpfung anderer Studiengänge als Physik problematisch, da sich die Praxis auch in der Übertragung der physikalischen Inhalte auf den Hauptstudiengang manifestiert.

Nicht überraschend bringen daher Nicht-Physiker auch ein geringes Interesse am Physik-Praktikum mit (KISSMANN et al., 2010; THEYSSEN, 2000, S. 5). Grundsätzliches Problem der Nicht-Anpassung der Lehrveranstaltung auf Nebenfächler*innen können zudem erhöhte Arbeitsbelastung und Erfolgsdruck sein, wenn innerfachliche theoretische Inhalte nicht elementarisiert werden. Dies begründen Befunde, dass erhöhte Arbeitsbelastung und Erfolgsdruck hauptsächlich bei Nicht-Physikstudierenden in Physikpraktika auftreten (HAAKE & MÜLLER, 2011, S. 542; KREITEN, 2012, S. 41–42; THEYSSEN, 2000, S. 5). Dies kann sich schlecht auf die intrinsische Motivation (MANDERLINK & HARACKIEWICZ, 1984, S. 918, 926) und damit den Lernerfolg (DECI, 1993, S. 233) auswirken.

Konsistent dazu ist der Lernerfolg von physikalischen und auch chemischen Grundpraktika für verschiedene Studiengänge gemäß Quellenlage gering (ELLIOTT et al., 2008, S. 147; ENGELS & HÜLSBUSCH, 2008, S. 59; HUCKE, 1999, S. 134; KREITEN, 2012, S. 14; PLOMER, 2011, S. 62–64; THEYSSEN, 2000, S. 5). Im Bereich der Nebenfachpraktika, also Physikpraktika, die von Nicht-Physikstudierenden besucht werden, mangelt es speziell den Studiengängen Chemie, Biochemie und Geowissenschaften an Studien, die sich explizit auf diese Zielgruppe beziehen.

5.2.2. Probleme in der Vorbereitung

Die Vorbereitung der Studierenden auf den Praktikumstermin wird wie angesprochen stark von Vorbereitungsmaterialien reguliert (HUCKE, 1999, S. 10, 99). Vor allem das Versuchsskript spielt eine entscheidende Rolle für die Vorbereitung der Theorie. Es hat großen Anteil am Wissenszuwachs der Studierenden (ebd., S. 126, 132).

Kritisch sind daher Befunde zu schlechter Skriptqualität zu sehen (KREITEN, 2012, S. 183–185; NAGEL, 2009, S. 41–42). Die Skripte der Grundpraktika sind oft veraltet (ABRAHAM et al., 1997) und beinhalten zu enge Vorschriften, die zu geringem Handlungsspielraum in Vorbereitung und späterer Durchführung führen (HUCKE, 1999, S. 93, 99). Der Beitrag zur Vorabbeschäftigung mit praktischen Inhalten des Experiments ist zudem selten gegeben (KREITEN, 2012, S. 183–185; NAGEL, 2009, S. 41; THEYSEN, 2000, S. 63, 69). Insgesamt lässt sich hier also keine gute Theorie-Praxis-Verknüpfung vermuten, da die Vorbereitung eher theoretischer Natur ist und wenig Einblick in das Experiment gibt.

Bezogen auf Methoden wissenschaftlichen Denkens ist festzuhalten, dass eine eigene Hypothesenbildung zum Experiment durch die engen Vorschriften des Skripts erschwert, wenn nicht verhindert wird (vgl. Abschnitt 3.4.4). Es fehlt hier somit ein wichtiger Schritt des Erkenntnisgewinnungsprozesses.

5.2.3. Probleme in der Durchführung

Anknüpfend an die stark durchstrukturierte Vorbereitung wird auch in der Durchführung wenig Handlungsspielraum gewährt, da das Skript experimentelle Teilschritte meist vorgibt. Häufig wird in diesem Zusammenhang von einer »rezeptartigen« Struktur gesprochen, also dem eher unreflektierten Abarbeiten von Teilschritten (HUCKE, 1999, S. 10f; KREITEN, 2012, S. 13; RUICKOLDT, 1996, S. 1024).

Da die Durchführung die einzige Phase ist, in der regulär experimentiert wird, kann dieser auch der Erwerb experimenteller Kompetenzen am stärksten zugeordnet werden, auch wenn andere Bestandteile experimenteller Kompetenz (z. B. Auswertung und Interpretation) erst in der Nachbereitung geschult werden. HUCKE (1999, S. 10f) fasst in Anlehnung an Vorarbeiten zusammen, dass der Erwerb experimenteller Fähigkeiten selten gefördert wird. Diesbezüglich kritisieren z. B. FRÜHWEIN & HEINKE (2007), dass die Lehramtsstudierenden das Praktikum mit geringen experimentellen Fähigkeiten verlassen.

Für die Physikstudierenden stellt NEUMANN (2004, S. 11) Ähnliches fest. Auch ZASTROW (2001) machte Erfahrungen mit Misserfolgen von Studierenden mit wenig vertrauten Messgeräten zum Ausgangspunkt ihrer Dissertation. Die im Studium zu

absolvierenden Grundpraktika befähigten gemäß einer Umfrage schon vor etwa 20 Jahren nicht dazu, eigenständig eine experimentelle Abschlussarbeit anzufertigen (RUICKOLDT, 1996, S. 1023). Bezüglich des Ziels, experimentelle Fähigkeiten zu erwerben, sieht die Ausgangslage vieler Praktika also verbesserungswürdig aus.

5.2.4. Probleme in der Nachbereitung

Die Phase der Nachbereitung ist vom Auswerten und Schreiben des Protokolls geprägt. In der Nachbereitung findet sowohl die theoretische als auch die praktische Auseinandersetzung mit dem Versuch statt. Es kann daher der Fokus auf Reflexion und Theorie-Praxis-Verknüpfung gelegt werden, sowie auf die Phasen der Auswertung und Interpretation im Rahmen der experimentellen Kompetenz und das innerfachliche Bewerten der Ergebnisse auf ihre Plausibilität hin. Zudem spielt die schriftliche Kommunikation im Rahmen der Protokollanfertigung eine Rolle.

Bezüglich der Intensität der Reflexionen wurde keine Studie gefunden. Es kann aber vermutet werden, dass bedingt durch die eher mathematiklastige Auswertung in Physikpraktika und die stark vorgegebene Protokollstruktur keine intensive Reflexion des Versuchs stattfindet. Bezüglich der Verzahnung von theoretischen und praktischen Aspekten kritisiert HALLER (1999, S. 85, 187), dass Durchführungs- und Nachbereitungsphase zeitlich stark getrennt geplant seien. So finde eine Entkopplung dieser Phasen statt und damit eine Entzweiung von Phasen des Erkenntnisgewinnungsprozesses. Die Zielerreichung der Verknüpfung von Theorie und Praxis in Grundpraktika ist damit anzuzweifeln.

Bezogen auf den Erwerb schriftlicher, kommunikativer Fähigkeiten im Rahmen der Protokollanfertigung berichten LAMMERTZ & HEINKE (2014) von mangelhafter Protokollqualität. Es ist daher anzunehmen, dass auch dieser Kompetenzbereich eher mäßig angesprochen wird bzw. schlichtweg die Betreuung für das Anfertigen der Protokolle fehlt.

5.2.5. Exkurs: Nebenfachpraktika

Physikpraktika, die von Nicht-Physikstudierenden absolviert werden, werden als Nebenfachpraktika bezeichnet. Bezüglich Forschungsergebnissen zu physikalischen Nebenfachpraktika fällt auf, dass vor allem bei Praktika für Medizinstudierende viele Erkenntnisse aus Studien vorliegen. Darin wird wiederholt als Problem genannt, dass zu viele quantitative Versuche durchgeführt würden, obwohl Expert*innen der Medizin ein mehr qualitatives Verständnis fordern (vgl. THEYSSEN, 2000, S. 45, PLOMER, 2011, S. 8).

5.3. Empirische Interventions-Studien zu Praktika (Auswahl)

Zudem fehle fast immer der notwendige Bezug zur Medizin (vgl. THEYSSEN, 2000, S. 4–5, PLOMER, 2011, S. 62–64) und somit auch das Interesse am Praktikum (vgl. THEYSSEN, 2000, S. 5). Auch das geringe bzw. stark heterogene Vorwissen in Physik wurde wiederholt festgestellt (vgl. THEYSSEN, 2000, S. 5, 51, PLOMER, 2011, S. 62–64). Zusätzlich dazu ist der Anteil der Beschäftigung mit Fragen zur Technik oder Geräten so hoch, dass nur geringe Lerneffekte und kaum Medizin-Physik-Transfer stattfinden (vgl. THEYSSEN, 2000, S. 5, 68, PLOMER, 2011, S. 4, 6, 62).

Ähnliche Befunde ergaben sich für Biologiestudierende. Die Befunde von HAAKE & MÜLLER (2011, S. 541) indizieren geringe Bezüge zum eigenen Studienfach und geringes Praktikumsinteresse. KISSMANN et al. (2010) zeigen zudem Evidenzen bezüglich zu geringen Vorwissens der Studierenden (vgl. Abschnitt 5.2.1).

Problemanalysen bezüglich der Studiengänge Chemie, Biochemie oder Geowissenschaften liegen nur eingeschränkt vor. Es wird auch hier davon ausgegangen, dass das physikalische Vorwissen unzureichend ist (vgl. FRICKE & SCHECKER, 2011, S. 1) und eine allgemeine Unzufriedenheit der TeilnehmerInnen geäußert wird (vgl. BORAWSKI & HEINKE, 2005, S. 1). Insgesamt sind aber keine empirisch gesicherten Befunde für diese Zielgruppe vorhanden.

5.3. Empirische Interventions-Studien zu Praktika (Auswahl)

In der Tradition der deutschsprachigen Praktikums-Forschungsliteratur, vor allem in den fachdidaktischen Dissertationen, folgt fast immer auf eine mehr oder weniger theoriegeleitete und empirische Analyse der Forschungslage und der Missstände im Praktikum eine Intervention. Um die vermuteten Mechanismen hinter den Problemen in Praktika besser zu verstehen und auf eine vorgenommene Praktikumsevaluation mit Interventionen reagieren zu können, ist es hilfreich, eine Auswahl an bereits durchgeführten und teils auch evaluierten Interventionen zu betrachten.

Um die Jahrtausendwende erforschten HUCKE (1999) und SANDER (2000) gemeinsam ihr physikalisches Anfängerpraktikum für Physikstudierende an der Universität Dortmund (heute: Technische Universität Dortmund) mit Studierenden des meist dritten Semesters. Eine Bedarfsanalyse des Praktikumsgeschehens wurde nicht vorgenommen, stattdessen die lernrelevanten Ziele »Anwendung physikalischer Konzepte« und »Aktiver Erwerb physikalisch-theoretischen Wissens beim Experimentieren« für die Studierenden festgesetzt und deren Erreichung untersucht (HUCKE, 1999). Darüber hinaus wurden Modifikationen des regulären Praktikums in Form zweier Interventionsgruppen vorgenommen. Die eine Gruppe beschäftigte sich zusätzlich zum Versuch mit Modellbildungssystemen (MBS), die andere nutzte zusätzlich elektronische Messwerterfassung (MBL). Die Kontrollgruppe (TRAD) ab-

solvierte das reguläre Praktikum (»expository-style«, vgl. Abschnitt 3.4.4). Ein kurzer Blick auf die Wirksamkeit der Interventionen zeigt, dass die MBL-Gruppe keine häufigere Beschäftigung mit fachlicher Theorie auf der abstrakten Ebene auswies, also die Kommunikation der Studierenden untereinander vor allem deskriptiv über Bestandteile des Versuchsaufbaus und technische Details verlief. Hücke vermutete hier, dass MBL einen größeren Effekt bei offeneren Lernumgebungen erwirken würde (ebd., S. 148f). Das TRAD-Praktikum wurde für den intendierten Lernerfolg als ungeeignet befunden, da viele Routinetätigkeiten und eine häufige Kommunikation auf der deskriptiven Ebene vorlagen. Die Intervention MBS konnte eine deutliche Steigerung der Diskussionsanteile auf der abstrakten Ebene erzielen und damit auch zu stärker theoriegeleitetem Handeln im Praktikum im Gegensatz zu deskriptivem, rezeptartigem Abarbeiten verleiten.

In einem vergleichbaren Zeitraum untersuchte und intervenierte THEYSSEN (2000) im Physikpraktikum für Medizinstudierende an der Universität Düsseldorf im Rahmen der Methode der Didaktischen Rekonstruktion (KATTMANN et al., 1997). In Vorerhebungen mit Studierenden, Betreuenden und der Praktikumsleitung wurde Kritik an den Inhalten, Kontexten, der Motivation und dem Lernerfolg im Praktikum geäußert. Das Vorwissen der Studierenden wurde als überraschend niedrig, aber heterogen verteilt gemessen. Die Ziele einer notwendigen Intervention waren eine theoriegeleitete Entwicklung eines neuen Praktikumskonzeptes, dass medizinische Bezüge transparent herausstellt und damit die Motivation der Studierenden steigert. Die Aufgaben für die Studierenden sollten vereinfacht werden, die Versuchsanleitungen unter Berücksichtigung des äußerst geringen Vorwissens sehr detailliert und kleinschrittig verfasst werden. Im Praktikum sollten die Studierenden den neuen Kontextbezug immer wieder selbst erneut herstellen und die kleinschrittige Strukturierung nutzen. Die teils komplexe Auswertung der Messungen wurde in die Präsenzzeit integriert, um adäquate Betreuung zu gewährleisten. Die Auswertungen selbst wurden ebenfalls vereinfacht und eher qualitativ ausgerichtet, wie dies von Expert*innen der Medizin gefordert wurde. Größte Veränderung zum Praktikum vorher stellte der Umgang mit fachlicher Theorie dar, da diese nunmehr erst nach dem Experimentieren bzw. währenddessen gegeben wurde (THEYSSEN, 2000, S. 127). Dies ist also ein induktiver statt ein deduktiver Zugang (vgl. Abschnitt 3.4.4).

Die Evaluation dieser Intervention mittels Studierendenbefragungen und Videoanalysen lieferte eine messbare Steigerung von Motivation und Lernerfolg der Studierenden, eine Erhöhung der wahrgenommenen Bedeutung des Praktikums für den Beruf, eine Reduzierung des Anweisungs-Zeitaufwands für die Betreuenden und ei-

ne stärkere Verzahnung der Auswertephasen. Bezüglich des Aufbaus und Messens im Praktikum wurden Wortäußerungen auf niedriger Komplexitätsebene festgestellt (vgl. HUCKE, 1999), die Auswertephasen fielen dagegen auf höherem Niveau sowohl für die physikalischen, als auch für die medizinischen Inhaltsbereiche aus. Diese Ergebnisse, obwohl für ein spezifisches Mediziner*innenpraktikum erforscht, sieht THEYSSEN (2000) als übertragbar auch auf andere Nebenfächler-Praktika an, was angesichts ähnlicher Probleme (vgl. Abschnitt 5.2.5).

Vier Jahre später veröffentlichte NEUMANN (2004) seine Ergebnisse zur Überarbeitung des Physikalischen Praktikums für Physiker, ebenfalls an der Universität Düsseldorf. Methodisch ging er dabei analog zu THEYSSEN im Rahmen der Didaktischen Rekonstruktion vor. In Expert*innengesprächen und offenen Befragungen der Studierendenschaft erforschte er hierfür zunächst die Sicht der Lehrenden und Lernenden im Praktikum, wie es Teil der Didaktischen Rekonstruktion ist. Zielsetzung war eine Betonung der beiden Kernziele nach HALLER (1999), experimentelle Fähigkeiten zu erwerben und Methoden wissenschaftlichen Denkens kennenzulernen. Diese sollten in einer Neukonzeption und Evaluation eines neuen Praktikumskonzepts münden. Die Neukonzeption baute auf einem konsequent konstruktivistischen Lernmodell auf, kognitive Aktivität und Konzeptwechselentwicklungen wurden angestrebt. Schließlich entstand ein Praktikumskonzept, das aus drei Teilen aufgebaut ist. Im *Gerätekunde*-Teil wird der Umgang mit Geräten und das manuelle Bedienen geschult. Der Teil zu *wissenschaftlichen Methoden* behandelt Messmethoden der Physik. Schließlich können dann im *Projektpraktikum*steil die erworbenen Fähigkeiten in einer offeneren Lernumgebung an Experimenten angewendet werden. Wirkungen der Intervention waren deutliche Zuwächse der Leistung im Wissenstest zu Geräten und Methoden, sowie die Steigerung der Sicherheit im Umgang mit Experimenten im Experimentaltest.

In Wien wurden bis zum Jahr 2009 das Physikalische Anfängerpraktikum für Lehramtsstudierende der Physik evaluiert und Interventionen durchgeführt (NAGEL, 2009). Zur Bestimmung der Ausgangslage wurden narrative Interviews und Gruppengespräche mit Studierenden geführt und ausgewertet. Diese offenbarten abermals geringes und stark heterogenes Vorwissen der Studierenden, auf die es in der Intervention einzugehen galt. Nachdem die qualitativ gewonnenen Erkenntnisse auch quantitativ geprüft wurden, wurde abgezielt auf ein besseres, neues eLearning-System, das formativ evaluiert wurde und die Verbesserung der kognitiven und prozessorientierten Fertigkeiten der Studierenden im Sinne selbstbestimmten Lernens verfolgte. Dieses umfasste Vorbereitungsmaterial für 90 Experimente zu 42 Themen in Form einer eLearning-Umgebung innerhalb der bestehenden Praktikums-

Homepage. Es wurden eSkripte, beschriftete Abbildungen der Versuchsaufbauten, open-source-Webtools, Videos und Denkaufgaben implementiert. In einer anschließenden Längsschnittstudie über zwei Semester konnten Leistungsanstiege im Wissenstest und eine höhere Selbsteinschätzung kognitiver und praktischer Fähigkeiten festgestellt werden. Die Etablierung selbstgesteuerten Lernens konnte festgestellt werden, 75% der Studierenden äußerten eine Verbesserung ihres Vorwissens.

Ebenso auf das Lehramt spezialisiert veröffentlichte KREITEN (2012) Ergebnisse ihrer Interventionsforschung zu den Experimentellen Übungen für Lehramtsstudierende der Physik, die ein zeitreduziertes Anfängerpraktikum mit zwei Wochenstunden Präsenzzeit umfassen. Evaluationen der Veranstaltung ergaben für die Studierenden unklare Anforderungen, fachliche Probleme, geringes Vorwissen und zu wenig Lehramtsbezug. Ziele der Neugestaltung war dementsprechend die Verbesserung der fachlichen Lernwirksamkeit durch eine Balance aus konstruktiven und instruktiven Anteilen des Praktikumsgeschehens im Rahmen eines eigens konzipierten Kompetenzmodells für Praktika (vgl. Abschnitt 6) und des Modells der vollständigen Handlung.

Die Intervention bestand aus der Etablierung elektronischer, textbegleitender Aufgaben, die kognitiv aktivierend, die Problemlösekompetenz schulend, Transparenz und Feedback enthaltend, Vorwissen aktivierend und Self-Assessment-ähnlich gestaltet sein sollten. Es wurden exemplarisch einzelne Versuche überarbeitet, zu elektrischen Stromkreisen und zum schiefen Wurf. Befunde waren, dass Studierende mit textbegleitenden Aufgaben deutlich besser vorbereitet waren, der neue Versuch »elektrische Stromkreise« zu einem höheren Lernzuwachs führte und die erstellte Simulation zum schiefen Wurf in Kombination mit dem Realexperiment eine deutliche Verbesserung des konzeptionellen Verständnisses bewirkte.

Das Nebenfächlerpraktikum der Physik für Geologie- und Chemiestudierende der Universität Bremen erforschte FRICKE (2012). Im Rahmen einer Intervention mit eSkript und Selbsttestaufgaben wurde untersucht, inwieweit die Studierenden sich mehr mit den Praktikumsthemen auseinandersetzen, ob das Aufbauen von Versuchen schneller und fehlerloser erfolgt und wie die Intervention von den Studierenden angenommen wird. Das eSkript (»hypermedia-Skript«) enthielt die regulären Inhalte des bisherigen Papierskripts im Hauptstrang und Vertiefungsthemen in Nebensträngen. Diese waren etwa Fehlerrechnung, interaktive Bildschirmexperimente oder auch die Selbsttestaufgaben zu den physikalischen Hintergründen des betreffenden Versuchs und betrafen die Vorbereitungsphase im Praktikum. Problematisch war das geringe Nutzungsverhalten der Studierenden in der ersten Kohorte: Nur neun von 53 Teilnehmenden nutzten das Angebot wie erwartet. Zukünftige Kohor-

ten erhielten dann Aufwandsentschädigungen für die Teilnahme. Die Wirkungen wurden mit Videos der realen Experimentiersituation im Praktikum, halbstrukturierten Leitfadeninterviews, Hypermedia-Logfiles, Fragebögen und Fachwissenstest untersucht.

In der Selbsteinschätzung der Studierenden verbesserte sich das Gefühl, ausreichend auf den Versuch vorbereitet zu sein, eine klarere Vorstellung vom Versuchsaufbau erhalten zu haben und die Teilschritte des Versuchs besser verstehen zu können. Erwartungsgemäß fiel die Vorbereitungsdauer mit dem Hypermedia-Skript allerdings deutlich länger aus, als mit dem Papierskript, und die Selbsttestaufgaben wurden als sehr unbeliebt evaluiert.

Zusammengefasst lässt sich sagen, dass naturwissenschaftliche Praktika je nach Fachgebiet und Zielgruppe unterschiedliche Bedarfe an Interventionen aufweisen. In jedem Fall kann von bereits existierenden best-practice-Beispielen (s.o.) gesprochen werden. Allerdings haben diese keinen Einzug in die deutschlandweite Praktikumskultur gefunden, sondern stellen lokal vereinzelte und wenig generalisierbare Evaluations- und Interventionsansätze dar, die als Übersicht für mögliche Interventionen für eine Überarbeitung der Lehrveranstaltungsform Praktikum nützlich sein können. Was indes fehlt, ist eine praktikumsübergreifende Erforschung, Evaluation und systemstische Intervention in vielen Praktika.

5.4. Zusammenfassung

Die Kernziele von Praktika sind das Verknüpfen von Theorie und Praxis (Fachwissen und Erkenntnisgewinnung), das Erlernen experimenteller Fähigkeiten (Erkenntnisgewinnung) und das Kennenlernen von Methoden wissenschaftlichen Denkens (Erkenntnisgewinnung und Bewertung), sowie die adäquate Kommunikation derselben in einer in Teilen selbstorganisierten Lehrveranstaltung. Die Ausgestaltung der Lehrveranstaltungsform Praktikum steht dem allerdings oft im Weg.

So wird vermutlich aus organisatorisch-ökonomischen Gründen fast der gesamte Erkenntnisgewinnungsprozess schriftlich im Skript vorgegeben, die Beschäftigung mit nicht-kontextbezogener Theorie über- und die mit Praxis in der Vorbereitung unterbetont. Das Reproduzieren gegebener Sachverhalte ist das implizite Ziel des Versuchs, was einem konstruktivistischen Lernprozess und dem Erlernen experimenteller Kompetenz als Ganzes (inkl. Hypothesenbildung, Explorieren, Rückinterpretation etc.) widerspricht. Die Beschäftigung mit Theorie wird zudem oft zu stark geprüft, mit Folgen für den Prüfungsdruck und damit die Motivation. Die Anforderungen der Lehrveranstaltung werden zu wenig auf das Niveau der Studierenden hin angepasst, was Untersuchungen zu schlechter Skriptqualität und dem geringem

Vorwissen der Studierenden aufzeigen. Die positive Entwicklung des Fachwissens ist damit auch fraglich. Das Skript stellt für die Theorie und die Durchführung das zentrale Medium im Praktikum dar. In der Durchführung hat die Theorie aber wenig Platz, fast die gesamte Zeit geht es um das Manipulieren von Geräten nach einer rezeptartigen Anleitung mit dem Ziel eines adäquaten Messprotokolls anstatt eines naturwissenschaftlichen Ergebnisses im Rahmen eines vollständigen Erkenntnisgewinnungsprozesses.

Mit dem Schreiben eines Versuchsprotokolls jede Woche wird die ohnehin schon lang ausfallende Präsenzzeit um weitere Stunden in Heimarbeit verlängert, die zeitliche Arbeitsbelastung ist oft hoch (METZGER & SCHULMEISTER, 2011) oder es werden Protokolle abgeschrieben. Die Protokolle selbst werden ebenfalls bemängelt, der Erwerb schriftlicher Kommunikationskompetenz ist anzuzweifeln. Insgesamt berichten Studien von geringen Lernerfolgen der Studierenden während der Präsenzzeit, sowohl im Fachwissen (HUCKE, 1999, S. 126, 132, 149) als auch in der experimentellen Kompetenz.⁴ Best-practice-Interventionen zeigen aber auch, dass der Erwerb von Fachwissen (NAGEL, 2009; NEUMANN, 2000) und experimenteller Fähigkeiten (NEUMANN, 2004) in Praktika erreicht werden kann, sofern eine Umstrukturierung des Praktikums und/oder des Skriptmaterials erfolgt.

Insgesamt fällt bezüglich des Forschungsstands einerseits auf, dass die wenigen empirischen Studien ausschließlich ihr lokales Praktikum analysierten, was zu wenig verallgemeinerbaren Resultaten führt. Andererseits sind die theoretische Bezüge der Studien, ihre Zielgruppen, ihre häufig mit der Studie einhergehenden Interventionen und ihre eingesetzten Methoden so divers, dass eine Vergleichbarkeit der Studien untereinander schwierig erscheint. An dieser Stelle wird bereits deutlich, dass eine gemeinsame theoretische Modellierung und eine Methode zum Vergleich dieser Praktika nützlich wäre, um die Erkenntnisse besser zusammenfassen zu können.

⁴Es scheint kaum das Erlernen von Hypothesenbildung möglich (KREITEN, 2012, S. 183–185; NAGEL, 2009, S. 41; THEYSEN, 2000, S. 63, 69). Auch experimentelle Fähigkeiten werden wenig gefördert (FRÜHWEIN & HEINKE, 2007; HUCKE, 1999, S. 10f; RUICKOLDT, 1996, S. 1023).

Kapitel 6

Forschungsstand: Theoretische Modelle zu Praktika

Aufgrund der Diversität und der teils mangelhaften Qualität der theoretischen Bezüge bisheriger Forschung zu Praktika und dem oft nur implizit gedachten Verständnis von einem »guten Praktikum« sollen nun zunächst die wenigen, bisherigen Ansätze der theoretischen Modellierung von Praktikumsqualität beschrieben werden.

Daraus folgt dann die Forschungslücke im Bereich der theoretischen Modellierung und Evaluation der Qualität von Praktika (Abschnitt 6). Dem Desiderat folgend wird alsdann eine Strukturierung von Praktikumsqualität vorgenommen (Abschnitt 7). Auf Basis dessen werden die Ziele und Forschungsfragen des empirischen Teils formuliert (Abschnitt 8). Dies fasst erneut die Leitgrafik in Abb. 6.1 zusammen.

Die in diesem Abschnitt vorzustellenden bisherigen Modelle zu Praktika sind das Wirksamkeitsmodell für Praktika nach PSILLOS & NIEDDERER (2002) und TIBERGHIE et al. (1998), die Praktikums-Lehrqualität nach LAGOWSKI (2002), und das Kompetenzmodell für die experimentellen Übungen nach KREITEN (2012). Die Praktikumsqualität nach FRASER & McROBBIE (1995) wurde bereits in Abschnitt 4.5.4 auf Seite 53 vorgestellt und wird zu den übrigen Modellen in Beziehung gesetzt werden.

6.1. Wirksamkeitsmodell nach PSILLOS & NIEDDERER (2002) und TIBERGHIE et al. (1998)

Begonnen wird mit dem Modell der zweifachen Wirksamkeit von Praktika aus dem europäischen Projekt »Labwork in Science Education« (1996–1998), welches als Meta-Modell betrachtet werden kann. Zwei Formen der Wirksamkeit werden definiert, die den Erfolg eines Praktikums ausmachen:

Die Wirksamkeit 1 definiert sich hierbei als das Maß, zu dem die Studierenden im Praktikum tatsächlich die Tätigkeiten ausführen, die von den Praktikumsverantwortlichen intendiert sind.

Wirksamkeit 2 beschreibt das Maß, zu dem die Lern-Outcomes der Studierenden mit den vorher formulierten Zielen der Lehrveranstaltung übereinstimmen (PSILLOS & NIEDDERER, 2002, S. 21). PSILLOS & NIEDDERER gehen in ihrer Modellierung also stark auf die Studierendenperspektive ein und machen den Praktikumserfolg

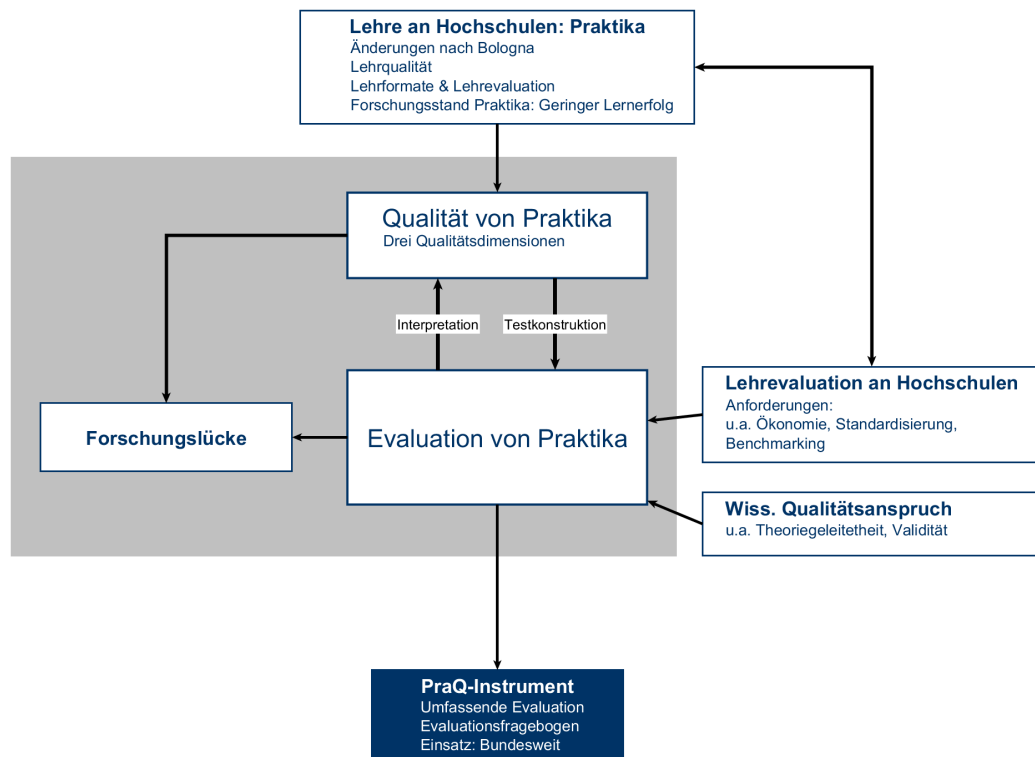


Abbildung 6.1.: Leitgrafik: Eine Exploration der bisherigen Modellierungen der »Qualität von Praktika« gemeinsam mit dem bereits festgestellten Desiderat eines theoriebasierten Instruments zur »Evaluation von Praktika« führt zu zwei Forschungslücken.

vom Erfolg derselben abhängig. Abb. 6.2 zeigt das Modell und seine angenommenen Einflussfaktoren. Beide Wirksamkeiten werden beeinflusst durch Einstellungen und Sichtweisen der beteiligten Akteure. So spielt die Sicht des/der Betreuenden auf Naturwissenschaft und Lernen eine Rolle für die Lehrintentionen (A, Abb. 6.2) und die Gestaltung der Aufgabenstellung (B). Die Sicht der Studierenden auf Naturwissenschaft und Lernen beeinflusst wiederum das tatsächliche Handeln (C) der Studierenden und ihren tatsächlichen Lernerfolg (D). Für beide Einflussmechanismen spielt auch der Kontext mit dem aktuellen Experiment eine Rolle.

6.2. Praktikums-Lehrqualität nach LAGOWSKI (2002)

Aus dem Forschungsraum der Chemie-Praktika stammt ein Modell der Praktikums-Lehrqualität von LAGOWSKI (2002).¹ Er baut hierbei auf dem Cognitive-Apprenticeship-Ansatz (COLLINS et al., 1987) auf, also verkürzt formuliert einem »learning by doing«-Ansatz unter Expert*innenaufsicht und -methoden. Darin werden von COLLINS et al. (1987) auch Dimensionen einer guten Lernumgebung definiert, die LAGOWSKI wiederum aufgreift und auf die bestehenden Lernumgebungen von Chemie-Praktika anwendet: (1) Die *inhaltliche Dimension* ist auf die Studierenden bezogen. Sie besteht aus den vier Komponenten *Fachwissen*, *heuristische Strategien*, *Kontrollstrategien* und *Lernstrategien*.

Fachwissen meint hier praktikumsspezifisches Domänenwissen, etwa das BOHRsche Atommodell, so wie es auch im Lehrbuch dargestellt würde.

Unter *heuristischen Strategien* werden Expert*innenstrategien verstanden, die oft als Erfahrungsstrategien bezeichnet werden könnten, da sie nie schriftlich festgehalten wurden. Dies können bestimmte haptische Manipulationen eines Experimentieraufbaus oder auch kognitive Vereinfachungen eines Modells sein. COLLINS et al. (1987) nennen diese »tricks of the trade« (Lagowski, 2002, S. 8).

Kontrollstrategien sollen Studierende dazu befähigen, ihren eigenen Lern- und Experimentierprozess und -fortschritt im Blick zu behalten und zu regulieren. So können schwierigkeiterzeugende Quellen ausfindig gemacht und diese dann in ihrer Komplexität reduziert werden. Dies sind z. B. komplexe experimentelle Aufbauten, die dann auseinandergebaut und Schritt für Schritt wieder zusammengesetzt werden, um den Gesamtaufbau verstehen zu können.

Lernstrategien helfen den Studierenden, die für sie in einem spezifischen Inhalts- oder Experimentiergebiet passende Quelle als Ansatzpunkt für ihr selbstbestimmtes Lernen zu finden. Für Manche ist dazu eher Lesen hilfreich, andere bevorzugen

¹Der Blickwinkel geht allerdings manchmal über Anfängerpraktika hinaus, etwa, wenn über die Motivation der Studierenden berichtet wird, zu publizieren (ebd., S. 12).

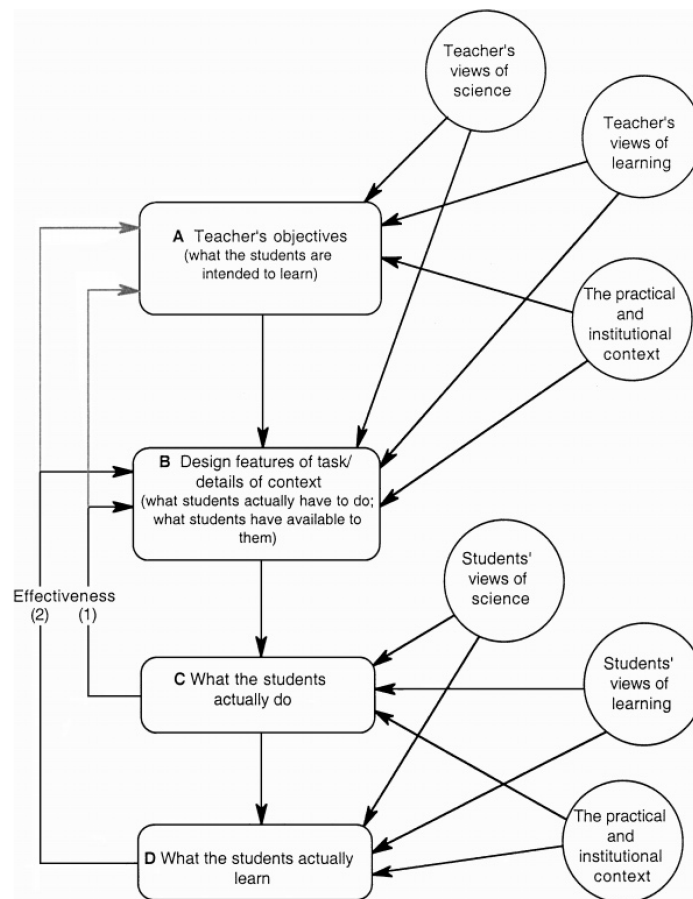


Abbildung 6.2.: Wirksamkeitsmodell für Praktika nach PSILLOS & NIEDDERER (2002) und TIBERGHIEEN et al. (1998, S. 488)

das Zuhören in Vorlesungen. Andere Lernformen sind beispielsweise die Beobachtung und das Nachahmen von peers oder das schriftliche Festhalten lernrelevanter Informationen.

(2) Zweite Dimension ist die *methodische Dimension*, bestehend aus den Komponenten *Nachahmungslernen*, *Betreuung*, *Fachkommunikation*, *Reflexion* und *Exploration*. *Nachahmungslernen*, *Betreuung* und *Exploration* sind auf die Betreuer*innen-Studierenden-Interaktion bezogen, *Fachkommunikation* und *Reflexion* betreffen die Studierenden.

Nachahmungslernen meint, dass Studierende von erfahreneren Experimentatoren lernen können, indem sie diese bei ihren Tätigkeiten beobachten. Gleichzeitig folgen die Studierenden den Äußerungen der Erfahrenen beim Experimentieren oder Erarbeiten eines Sachverhalts. Meist sind dies informelle Situationen im Praktikum, also innerhalb der Experimentierzeit nach den Anfangsgesprächen verortete Interaktionen.

Mit *Betreuung* ist Gegenteiliges gemeint: Nun beobachtet der/die Betreuende die Studierenden beim Handeln und macht inhaltliche, experimentelle oder methodische Vorschläge und gibt Feedback (scaffolding). Dieses scaffolding wird dann stetig reduziert, bis der/die Studierende die Aufgabe ohne Hilfestellung bewältigen kann (fading) (LAGOWSKI, 2002, S. 9f).

Die dritte Komponente wird definiert durch die *Fachkommunikation* der Studierenden. Es sei eine Aufgabe des Praktikums, dass die Studierenden es erlernen, klar und effektiv naturwissenschaftlich zu argumentieren, Probleme zu lösen und auch schriftlich zu kommunizieren, um später wissenschaftliche Publikationen verfassen zu können.

Die Komponente *Reflexion* beinhaltet, dass Studierende dazu befähigt werden und es etablieren, ihr Wissen, Argumentieren und ihre Problemlösestrategien mit denen anderer Studierender, Expert*innen und schließlich mit einem internen kognitiven Modell der Expertise zu vergleichen.

Exploration liegt dann vor, wenn Betreuende ihre Studierenden dazu bringen, selbstständig an Problemen zu arbeiten, also eine Art Aufforderung zur Selbsttätigkeit, wie sie auch bei THIEL et. al (2012) im LeKo vorkommt (vgl. Abschnitt 4.5.1).

(3) Dritte Dimension ist die *Sequenzierung* mit den Komponenten *Globale vor lokale Fertigkeiten*, *Erhöhung der Komplexität* und *Erhöhung der Vielfalt*. Alle drei Komponenten weisen klare Bezüge zur Konzeption des Praktikums bzw. zur *Lernumgebung* auf und sind damit den Strukturdaten zuzuordnen, auch wenn die Be-

6. Forschungsstand: Theoretische Modelle zu Praktika

treuenden erneut exekutive Funktionen einnehmen müssen, indem sie diese Forderungen umsetzen.

Globale vor lokale Fertigkeiten meint das Primat des naturwissenschaftlichen Konzepts vor den manuellen Feinheiten der Methode. Das kann beispielsweise das »rezeptartige Abarbeiten« (vgl. 3.4.4) verhindern, da so die Handlung vorher kognitiv durchdacht wurde. Bereits bei der Konzeption eines Praktikumsversuchs sei dies zu bedenken.

Erhöhung der Komplexität liefert die Aufforderung, die Lernumgebung und Aufgaben so zu gestalten, dass die Anzahl an von den Studierenden benötigten Fähigkeiten, Konzepte und deren Integration mit jeder Aufgabe stetig ansteigen.

Erhöhung der Vielfalt wird dann erreicht, wenn auch die Fähigkeiten und Konzepte wechselnd in den Aufgaben gefordert sind (LAGOWSKI, 2002, S. 10f).

(4) Vierte und letzte Dimension des Modells nach LAGOWSKI (2002) ist die *soziologische*. Enthaltene Komponenten sind hier das *situierte Lernen*, die *Kultur der Expertise*, *intrinsische Motivation*, *Kooperation* und *Konkurrenz*. *Situiertes Lernen* und *Kultur der Expertise* sind Eigenschaften der Konzeption der *Lernumgebung*, *intrinsische Motivation*, *Kooperation* und *Konkurrenz* sind der Qualitätsdimension *Lernzuwachs* zuzuordnen.

Situiertes Lernen beschreibt das Ausmaß, zu dem das Praktikum die verschiedenen Anwendungsfelder der dort zu erwerbenden Fähigkeiten verdeutlicht und integriert.

Kultur der Expertise meint, dass die Gestaltung der Lernumgebungen und Aufgaben in Praktika so vorgenommen wird, dass die Studierenden schrittweise lernen, wie Expert*innen zu denken.

Intrinsische Motivation kann sich dann bei den Studierenden ausbilden, wenn sie interessante Ziele verfolgen können. LAGOWSKI (2002) führt hier die Aussicht auf die Entdeckung neuer wissenschaftlicher Erkenntnisse an, bezieht sich aber auf den Doktorierendenstatus.

Kooperation ist eine Metakompetenz, die auch von LAGOWSKI gefordert wird. Sie sei ein starker Motivator und aus der heutigen Forschungskultur nicht wegzudenken, wo Naturwissenschaftler*innen fast immer in Teams zusammenarbeiten.

Mit *Konkurrenz* wird Gegenteiliges angesprochen: Studierende sollen auch lernen, ihren Konkurrenzkampf als Beweggrund nutzen zu lernen. Erneut bezieht sich LAGOWSKI hier auf erste echte Forschungsarbeiten und kommt in diesem Konstrukt etwas vom Anfängerpraktikumskontext ab. Sämtliche vorherigen Konstrukte sind aber gut geeignet, um eine theoretische Modellierung von Praktikumsqualität vorzunehmen, die neben Physik- auch Chemiepraktika gerecht wird.

6.3. Kompetenzmodell nach KREITEN (2012)

Einen ersten Ansatz für ein Kompetenzmodell lieferte KREITEN im Jahr 2012 (Abb. 6.3). Dies ist wünschenswert für ein Modell der Praktikumsqualität, da Kompetenzen bereits in ihrer Anlage eine Messbarkeit und damit auch Operationalisierbarkeit intendieren (HOPBACH, 2005b), was von entscheidender Bedeutung für Evaluationsmessungen ist. KREITEN adaptierte hierfür die Kompetenzmodelle für die nationalen Bildungsstandards in den naturwissenschaftlichen Fächern (KMK, 2005) in Hinblick auf das physikalische Praktikum für Lehramtsstudierende (»experimentelle Übungen«).

Dieses normative Modell besteht demzufolge aus den klassischen Kompetenzbereichen *Fachwissen*, *Erkenntnisgewinnung*, *Kommunikation* und *Bewertung* (erste Dimension). Diese werden in drei *Anforderungsbereiche* ausdifferenziert (zweite Dimension). Die dritte Dimensionierung des Modells ist die *Phase* im Praktikum, also die Vorbereitung, Durchführung und Nachbereitung. Gerade *Kommunikation* und *Bewertung* erfahren für die Adaption auf Praktika die größten Veränderungen im Vergleich zu den schulorientierten Bildungsstandards. So wird der Bereich Kommunikationskompetenz stark auf schriftliche und lesbare Kommunikation ausgerichtet (z. B. »vorgegebene Grafiken nachvollziehen«, Phase Vorbereitung, AB1, KREITEN, 2012, S. 197), da Skripte verstanden und Protokolle geschrieben werden müssen. Der Bereich *Bewertung* wiederum wird um das gesellschaftsrelevante Element beschnitten und nur die innerfachliche Bewertung betrachtet (z. B. »Versuchsergebnisse mit [...] Literaturwerten vergleichen und bewerten«, Phase Auswertung, AB2, ebd.). Eine klare Begründung für diesen Schritt nennt KREITEN nicht. Mutmaßlich wird der Grund in der sehr fachnahen Ausbildung an Hochschulen zu finden sein.

6.4. Zusammenfassung

Mit Blick auf eine flächendeckende Evaluation von Praktika und die »neue« Outputorientierung der Universitäten (vgl. Abschnitt 2.1 auf Seite 19) ist das Wirksamkeitsmodell nach PSILLOS & NIEDDERER (2002) zwar nützlich, um eine Vorstellung der im Praktikum ablaufenden Prozesse zu erhalten und Hypothesen für die Interdependenz von Betreuer*innenmerkmalen und Studierendenmerkmalen zu erhalten. Es eignet sich aber nicht dazu, konkrete Konstrukte der Praktikumsqualität abzuleiten. Das Modell wurde primär für Videoanalysen konstruiert, wo diese Prozesse sichtbar werden, als bei einer Standard-Lehrevaluationserhebung. Die Wirksamkeit 1 nach PSILLOS & NIEDDERER (2002) ist daher als moderat in seiner Bedeutung für die Lehrevaluation einzuschätzen. Erstens ist nicht erforscht, ob die (meist von den Praktikumsleitenden) intendierten Tätigkeiten auch tatsächlich zu relevanten

6. Forschungsstand: Theoretische Modelle zu Praktika

Kompetenz-bereiche		Anforderungsbereiche			
		I	II	III	
Fachwissen		Wissen wiedergeben: Fakten und physikalische Sachverhalte aus der Vorlesung und den Versuchsunterlagen reproduzieren.	Wissen anwenden: Physikalisches Wissen am Versuch anwenden, Sachverhalte identifizieren und nutzen, Modelle und Analogien benennen. Sicherer Umgang mit versuchsrelevanten Bestimmungsgleichungen. <i>Sich bewusst werden über die eigenen bestehenden Vorstellungen und Modelle</i>		VORBEREITUNG
Kommunikation		Mit vorgegebenen Darstellungsformen arbeiten: Sachverhalte schriftlich wiedergeben können, vorgegebene Grafiken nachvollziehen, zuordnen, beschreiben und darstellen können.	Geeignete Darstellungsformen nutzen: Sachverhalte fachsprachlich korrekt und strukturiert darstellen, auf Beiträge und Fragen sachgerecht eingehen, Aussagen sachlich begründen.		
Erkenntnis-gewinnung		Fachmethoden beschreiben: Physiktypische experimentelle Arbeitsweisen nachvollziehen und beschreiben.			
Fachwissen			Wissen anwenden: Kritische Überprüfung der Messwerte, um evt. Fehlerquellen frühzeitig zu beheben. Vorstellungen überprüfen: evt. Fehlvorstellungen erkennen.	Wissen transformieren und verknüpfen: Fehlvorstellungen korrigieren.	DURCHFÜHRUNG
Erkenntnis-gewinnung			Fachmethoden nutzen: Vorgestellte Experimente in unterschiedlichen Graden der Selbstständigkeit planen und durchführen. Wissen nach Anleitungen erschließen. Umgang mit Messgeräten üben, Sicherheitshinweise beachten.		
Fachwissen				Wissen transformieren und verknüpfen: Die Bedeutung des Versuchs beurteilen, physikalische Erkenntnisse für die Bewertung des Versuchs nutzen, Phänomene in physikalischen Kontext einordnen.	AUSWERTUNG
Kommunikation			Geeignete Darstellungsformen nutzen: Sachverhalte fachsprachlich und strukturiert darstellen, Aufgaben sachgerecht bearbeiten, Aussagen sachlich begründen.	Darstellungsformen selbstständig auswählen und nutzen: Darstellungsformen sachgerecht und adressatengerecht auswählen, anwenden und reflektieren.	
Bewertung		Vorgegebene Bewertungen nachvollziehen: Ziele und Bedeutung des Versuches schildern.	Vorgegebene Bewertungen beurteilen: Versuchsergebnisse mit anderen Versuchswerten und/oder Literaturwerten vergleichen und bewerten.	Eigene Bewertungen vornehmen: Systematische und statistische Fehler des Versuches diskutieren, Verbesserungsvorschläge benennen und begründen. Beschreibung von physikalischen Erkenntnissen in historischen und gesellschaftlichen Zusammenhängen.	

Abbildung 6.3.: Kompetenzmodell für Praktika nach KREITEN (2012, S. 197)

Lernzuwachsen führen, wie PSILLOS & NIEDDERER selbst schreiben: »...relation between student activities in doing science and the conceptual, procedural, and epistemological outcomes after labwork is a complex one, which is currently under investigation« (ebd., S. 26).

Auch das Modell selbst sieht zwar indirekte Einflüsse zwischen dem, was Studierende lernen sollen und dem, was sie tatsächlich lernen vor (A vs. D, vgl. Abb. 6.2 auf Seite 70), beschreibt aber nicht deren Gelingensbedingungen. Es besteht qualitativer Forschungsbedarf, um die komplexen Zusammenhänge der Prozesse im Praktikum verstehen zu können. Zweitens ist eine Erfassung der Wirksamkeit 1 (wie z. B. bei HUCKE, 1999 geschehen) mit einem enormen Aufwand verbunden, etwa videoanalytischer Auswertung der Praktikumsstätigkeiten oder zumindest einer möglichst objektiven Beobachtung der Lehrveranstaltungsdurchführung (vgl. PSILLOS & NIEDDERER, 2002, S. 27). Problematisch bleiben also die schwierige Operationalisierbarkeit und der hohe Abstraktionsgrad des Modells. PSILLOS & NIEDDERER schlagen zwar die Messung von veränderlichen Variablen vor, etwa der beliefs der Studierenden (PSILLOS, 2002, S. 24), bleiben aber die Operationalisierung sämtlicher weiterer Aspekte der Wirksamkeit 2 schuldig. Sie selbst schreiben sogar davon, dass Forschende sich für die Erforschung der Wirksamkeit 1 oder 2 oder beider entscheiden könnten (PSILLOS & NIEDDERER, 2002, S. 26). Dies legt nahe, dass es sich um ein prozessstrukturierendes Modell, nicht aber um ein operationalisierbares Qualitätsmodell von Praktika handelt. Es kann daher für eine flächendeckende Evaluation von Praktika als hintergründiges, strukturierendes, nicht aber bestimmendes und operationalisierbares Modell gelten.

Die verbleibenden drei Modelle haben gemein, dass sie konkretisieren, was manifeste oder latente Merkmale eines guten Praktikums (im Sinne des *Lernzuwachses* der Studierenden, des Verhaltens der Betreuenden oder der *Lernumgebung*) sind. Tab. 6.1 fasst die vielfältigen Konstrukte der Modelle in das bekannte Raster der Qualitätsdimensionen.

Der Bereich der Dimension *Lernzuwachs* erscheint hier recht homogen, viele Konstrukte werden mindestens doppelt angesprochen. Es lässt sich also gut eine Auswahl an qualitätsrelevanten Konstrukten ableiten, die auch im Einklang mit den Zielsetzungen für Praktika nach HALLER (1999) stehen (vgl. Abschnitt 3.4.2 auf Seite 30). Der Bereich der *Lehrkompetenz* wird dagegen kaum berührt, selbst unter den verbleibenden drei Konstrukten sind nur *Nachahmung* und *Exploration* übrig, die nicht nur lehrenden-, sondern auch lernenden- und strukturorientiert sind.

Der Bereich des *Lehrens* in Praktika ist übereinstimmend dazu fast noch nicht erforscht. Es herrscht also Nachholbedarf für eine Definition adäquaten Lehrenden-

Tabelle 6.1.: Konstrukt-Strukturierung der theoretischen Modelle zu Praktika. Unterschieden wird in lernzuwachsrelevante, lehrrelevante und lernumgebungsrelevante Konstrukte. Fett gedruckt sind diejenigen Konstrukte, die sich gemäß in mindestens zwei Modellen wiederfinden. So können *heuristische Strategien* und *Erkenntnisgewinnung*, *Studierenden-Zusammenhalt* und *Kooperation*, sowie *Reflexion* und *Bewertung* sinngemäß verknüpft werden, da sie analoge Konstrukte ansprechen oder Teilmengen voneinander sind.

Quelle des Modells	Dimension <i>Lernzuwachs</i>	Dimension <i>Lehrkompetenz</i>	Dimension <i>Lernumgebung</i>
FRASER & McROBBIE (1995)	Studierenden-Zusammenhalt		Offenheit Integration Regelklarheit materielle Umgebung
	Fachwissen	Nachahmung Betreuung Exploration	Global vor lokal Erhöhung Komplexität Erhöhung Vielfalt Situierendes Lernen Kultur der Expertise
	heuristische Strategien		
	Kontrollstrategien Lernstrategien		
LAGOWSKI (2002)	Fachkommunikation		
	Reflexion		
	intrinsische Motivation		
	Kooperation Konkurrenz		
KREITEN (2012)	Fachwissen		
	Erkenntnisgewinnung		
	Kommunikation		
	Bewertung		

verhaltens in Praktika. Die Strukturdaten (*Lernumgebung*) sind gemäß der zwei eingeordneten Quellen zweigeteilt. FRASER & MCROBBIE (1995) sehen vermehrt organisatorische Belange als wichtig an, LAGOWSKI (2002) betrachtet primär die Gestaltung der Aufgaben in der Lernumgebung. Das im deutschsprachigen Forschungsraum als zentral angesehene Praktikumsdokument wird nicht angesprochen. Für eine moderne, kompetenzorientierte Lehrevaluation, mit der alle Facetten von Praktikumsqualität abgedeckt werden, kann eine Fusion dieser Elemente unter Zuhilfenahme des Forschungsstands in Praktika zu einem theoretischen Modell der Praktikumsqualität führen.

6.5. Forschungslücke: Lehrevaluation von Praktika

Clearly, it would be timely to initiate a new line of research which could help us to obtain feedback about students' views of laboratory settings and investigate the impact of laboratory classes on student outcomes.

FRASER & MCROBBIE (1995, S. 290)

Für die Erforschung von Praktika (vgl. Kapitel 5 auf Seite 55) wurden die impliziten und expliziten Ziele vorgestellt (vgl. Abschnitt 3.4.2), (wenige) theoretische Modelle von der Forschungsgemeinschaft erarbeitet und Methoden zur Feststellung des IST-Zustands in Praktika analysiert. Einige dieser Studien wurden bereits im Kapitel 5 auf Seite 55 aus dem Bereich physikalischer und naturwissenschaftlicher Experimentalpraktika vorgestellt und auch damit einige Probleme, die sich über die verschiedenen Standorte und Strukturen als vielleicht verallgemeinerbar ansehen lassen.

Nichtsdestotrotz sind die Studien in ihrer Aussagekraft stark lokal gebunden, da sie nur ein einziges Praktikum ohne Kontrollgruppe erfassen, auch durch das häufig mit der Studie verbundene Ziel, just die vorliegende Praktikumsrealität am Standort zu verbessern. Ebenso aber lassen die diversen Befunde erkennen, dass Praktika teils mit teils individuellen Problemen zu kämpfen haben – trotz vermeintlicher Vereinheitlichung durch die Zielsetzungen. Dies zeigt sich bereits bei der Organisation der Praktika und ebenso in der sehr heterogenen Zielgruppe in einem Physikpraktikum, von Medizinstudierenden bis zu reinen Physikstudierenden. Ist hier eine Analyse gewünscht, so muss auch die Methode dazu spezifischer Natur sein:

The routine collection of students' evaluations does not in itself lead to any improvement in the quality of teaching

RICHARDSON (2005, S. 392)

Daher liefern die in Vorlesungen und Seminaren etablierten Evaluationsinstrumente (vgl. Abschnitt 4.5 auf Seite 48) hier keinen direkten Beitrag, da sie nicht auf die besondere Lehrveranstaltung »Praktikum« ausgerichtet konstruiert wurden (vgl. Abschnitte 3.4 & 5.1).

Die bisherige Praktikumsforschung lieferte bereits gute Ansätze für die Erforschung von Praktika, aber auch diese sind bisher nicht-standardisiert und erfüllen nicht alle Anforderungen an eine valide Interpretierbarkeit der Messung. Die mehrfach eingesetzte Methode der Didaktischen Rekonstruktion (u.a. THEYSSEN, 2000; NEUMANN, 2004) ist ein Ansatz zur Standardisierung des Verfahrens, da es die zu erfassenden Perspektiven (Lerner, Lehrender, Fach) umfasst. Allerdings wird hier nicht die Erhebungsmethode selbst standardisiert. Dies zeigt sich in den unterschiedlichen Ansprüchen, wann auf qualitative und wann auf quantitative Daten zurückgegriffen werden sollte und welche Gütekriterien die Messinstrumente erfüllen müssen.

Sicherlich ist die Nicht-Beachtung dieser Aspekte in den vorliegenden Forschungsarbeiten zu Praktika auch dem Umstand geschuldet, dass diese zumeist in einer Zeit entstanden sind, als sich die empirische Forschung in der Fachdidaktik in ihrer frühen Phase befand (BUCHHAAS-BIRKHOLZ, 2009, S. 29).

Eine einheitliche Methode zur Optimierung lokaler Praktika könnte hier Vergleichbarkeit schaffen, ein Vorschlag hierzu findet sich bei REHFELDT et al. (2015b).

Ein Benchmarking und ein Überblick über viele Praktika wäre lohnenswert, um aus Lehrenden-, Studierenden- oder Leitendensicht die Qualität des eigenen Praktikums einschätzen und vergleichen zu können. Es ist bislang außer der groß angelegten europäischen Studie zu Zielen von Praktika (HALLER, 1999; WELZEL, 1998) keine Studie bekannt, die großflächig im deutschsprachigen Raum Praktika untersucht hat. Noch wesentlicher wäre an dieser Stelle die Erfassung der Zielerreichung (ZWICKL et al., 2013), also, mit Rückblick auf die theoretischen Modelle von Praktika, der von den Studierenden erworbenen Kompetenzen.

Die Kompetenzen wiederum speisen ihren Inhalt aus den definierten Zielen und stehen daher in enger Wechselwirkung zu diesen. Empirisch liegt hier also eine Forschungslücke vor. Es gibt bislang keine flächendeckende Evaluation von Praktika im deutschsprachigen Raum, die Stärken und Schwächen derselben bzw. eine Vergleichbarkeit der Qualität zwischen Praktika aufdecken könnte. Aus theoretischer Perspektive ist hierfür ein Modell der Qualität von Praktika zu definieren (vgl. Anforderung 1, Abschnitt 4.4 auf Seite 43). Dabei sollte die ökonomische Einsetzbarkeit (Anforderung 3, ebd.) der Evaluationserhebung im universitären Regelbetrieb erhalten bleiben, damit die Durchführbarkeit (DEGEVAL, 2008) erhalten bleibt.

Für Vorlesungen und Seminare verschiedenster Fächer wurden solche qualitäts- und kompetenzorientierten Instrumente zur Lehrevaluation über alle Fächer bereits vorgestellt (z.B. BRAUN et al., 2008; THIEL et al., 2012, vgl. Abschnitt 4.5). Das SLEI-Instrument nach FRASER & McROBBIE (1995) kann trotz seiner Ausrichtung auf Praktika aus den genannten Gründen ebenso nur als erste Anregung für die Konstruktion eines theoretischen Modells dienen (vgl. Abschnitt 4.5.4). Naturwissenschaftliche Praktika wurden bisher nicht ausreichend im deutschsprachigen Raum berücksichtigt und eine bloße direkte Übertragung der vorhandenen Instrumente erscheint der Besonderheiten dieser Lehrveranstaltungsform nicht gerecht und wird auch von den Autor*innen selbst bezweifelt (THIEL et al., 2012, S. 22).

Spezifischen Eigenschaften von Praktika, die turnusartige Organisation, der hohe Anteil an Heimarbeit, der intensivere Betreuungsschlüssel und die Betonung experimentell-praktischen Arbeitens unterscheiden diese Lehrveranstaltungsform wesentlich von eher rezeptiv ausgerichteten Vorlesungen und Seminaren. Demzufolge sind auch die Qualitätsanforderungen unterschiedlich. Praktika bilden den Kern der experimentellen Ausbildung in den Naturwissenschaften (PSILLOS & NIEDDERER, 2002) und werden oftmals kritisch in ihrer Lernwirksamkeit betrachtet (Abschnitt 5), weshalb gerade dort eine passende und aussagekräftige Lehrevaluation von besonderem Nutzen wäre.

Der Abschnitt zu bisherigen theoretischen Modellen zur Qualität von Praktika lieferte zudem die Erkenntnis, dass Nachholbedarf für eine Definition adäquaten Lehrendenverhaltens in Praktika besteht, da die Ziele von Praktika (vgl. Abschnitt 3.4.2) vor allem die Dimension des *Lernzuwachses* betreffen. Der *Lernzuwachs* der Studierenden kann aber mittelfristig stark durch die *Lehrkompetenz* des/der Betreuenden beeinflusst werden (HATTIE, 2009). In dieser Arbeit soll zudem ein naturwissenschaftsübergreifendes Konzept der Evaluation realisiert werden.

Evidenz für die Verallgemeinerbarkeit des Lehrveranstaltungstyps »Praktikum« über die Fächer liefert die vielerorts tradierte Organisationsstruktur (Diese beschreibt schon WESTPHAL, 1937), die sich in den Publikationen zu Praktika widerspiegelt (BORAWSKI & HEINKE, 2005; KREITEN, 2012; PLOMER, 2011; LAGOWSKI, 2002) und auch dem Analysestand der Praktikumsverläufe und -Probleme an verschiedenen Fach-Praktika der Freien Universität entspricht (GUTZLER et al., 2014).

Der Forschungsstand kann wie folgt zusammengefasst werden: Weder die theoretischen Modelle noch die standardisierte Evaluation von Praktika sind elaboriert genug, um auf die Frage nach Qualitätsmerkmalen naturwissenschaftlicher Praktika eine befriedigende Antwort zu geben. Da dies aber ein Ausgangspunkt jeglicher

Überlegung zur Bewertung von Praktika sein sollte, erfolgt hier eine umfassende theoretische Modellierung des eben beschriebenen Desiderats.

Im Folgenden soll daher ein theoretischer Rahmen für ein fächer- und praktikumsübergreifend einsetzbares Evaluationsinstrument zur Erfassung der Qualität von Praktika entwickelt werden. Die Modellierung der Qualität von Praktika sollte verschiedene Facetten der Qualität enthalten und auch deren Wechselwirkung berücksichtigen.

Die bisherigen Modelle blicken hierfür zu stark eindimensional und zu vereinzelt auf (ihre) Praktika (vgl. vorheriger Abschnitt). Die Grundausrichtung der folgenden Modellierung soll kompetenzorientiert sein (vgl. Abschnitt 2.1 auf Seite 19), sei es auf Seiten der Studierenden als zu erwerbende studien- oder berufsrelevante Kompetenzen oder auf Seiten des Lehrpersonals im Sinne einer praktikumsspezifischen *Lehrkompetenz*.

Es wird im Zuge dessen deutlich, dass

- a)** ...ein theoretisches Modell der Praktikumsqualität benötigt wird (vgl. Anforderung 1, Abschnitt 4.4 & Kapitel 6).
- b)** ...eine Adaption von Instrumenten allgemeiner Lehrkompetenz für die Betreuungsanforderungen lohnenswert erscheint (vgl. Abschnitte 4.5.1, 5.1 & Kapitel 6).
- c)** ...eine übergreifende, flächendeckend einsetzbare Evaluationsmethode erarbeitet werden sollte (vgl. Anforderungen 4 & 7, Abschnitt 4.4).

Kapitel 7

Strukturierung von Praktikumsqualität

Wie in Tab. 6.1 auf Seite 76 dargestellt, lassen sich Aspekte der Qualität von Praktika in die drei Dimensionen lernenden-, lehrenden- und lernumgebungsorientiert unterteilen. Nach EINSIEDLER (2002, S. 195) ist die Unterrichtsqualität als die Zielerreichung der Ergebnismerkmale (Lernen) auf Grund der Wirkung der Prozessmerkmale (Lehr-Lern-Prozess) zu verstehen, weshalb sowohl Lehrenden- als auch Lernenden-Aspekte in der Erfassung von Praktikumsqualität ihren Wert haben. Die Rolle der Betreuenden in Praktika wurde bereits als kritische Variable herausgestellt (SANDER, 2000, S. 254) und sogar Schulungen für ebenjene gefordert (BOND-ROBINSON & RODRIQUES, 2006; SANDER, 2000, S. 254; SERE, 2000). Es wurden auch empirisch Anzeichen dafür gefunden, dass die einzige Tätigkeit im traditionellen Anfängerpraktikum, die förderlich für die Beschäftigung mit Theorien beim Experimentieren ist, die Interaktion mit dem/der Betreuenden darstellt (HUCKE, 1999, S. 148). Zudem seien die Betreuenden der wesentliche Interventionsfaktor, wenn Praktika sich zwar verbessern wollen, ihre tradierte Struktur aber beibehalten möchten (HUCKE, 1999, S. 100). TIBERGHIE et al. fassen dies zusammen: »At university level, the teacher is still the main initiator.« (TIBERGHIE et al., 1998, S. 500).

Die Wichtigkeit einer Kompetenzorientierung für den Lernerfolg der Studierenden ist bereits verdeutlicht worden.¹ Doch auch Strukturmerkmale (*Lernumgebung*) sind in Praktika von hoher Bedeutung. Es liegt ein vergleichsweise hohes Maß an Material- und Raumabhängigkeit in Praktika vor (vgl. Abschnitt 3.4). Die *Lernumgebung* wird als Voraussetzung für adäquates Lernen bezeichnet (AINLEY, 1990), sie sei gemäß KIND & KIND (2007) zudem eine Quelle für kreatives Handeln.

Empirisch zeigen sich konsistente Zusammenhänge zwischen der Wahrnehmung der *Lernumgebung* und den kognitiven und affektiven outcomes (FRASER & MCROBBIE, 1995, S. 291). Vor allem dem Skript als Bestandteil der *Lernumgebung* kommt

¹ Die derzeitige Kultur der Praktikumsleitenden ist nach persönlichen Erfahrungen bei zweimaligem Besuch der Praktikumsleitertagung »DPG-Schule« (2014 und 2015) als eher lehrzielorientiert zu beschreiben. Es wird viel über die technischen Experimentiermaterialien und Inputs diskutiert, weniger darüber, was Studierende konkret können sollen. Dieser Qualitätsdimension kommt also nach Ansicht des Autors die wichtigste Rolle zu, auch im Sinne eines dringend erforderlichen Perspektivwechsels mancher Praktikumsleitender.

7. Strukturierung von Praktikumsqualität

hier eine wesentliche Rolle zu (vgl. Abschnitte 5.1 & 5.2.2).² Da für naturwissenschaftliche Praktika eine theoretische Modellierung der Lehrveranstaltungsqualität bisher nicht erfolgt ist, lässt sich nur schwierig eine Aussage treffen, welcher der drei Bereiche größeren Einfluss auf die Qualität von Praktika hat. Sicherlich ist theoretisch im Zuge der Kompetenzorientierung dem *Lernzuwachs* eine höhere Bedeutung zuzuweisen. Welche Rolle dabei aber Einflüsse von Prozess und Struktur der Lehrveranstaltungen haben, ist ungeklärt. Für eine erste theoretische Modellierung kann es sich daher lohnen, alle drei Perspektiven zu erfassen. Dies kann einerseits Aufschluss über spezifische Stärke/Schwäche-Profile von Praktika liefern, also ob eher die Struktur oder die Prozesse verbesserungswürdig erscheinen. An diesen beiden Dimensionen kann anschließend eine Veränderung der Lehrveranstaltung ansetzen, etwa durch Optimierung des Skripts oder einer Schulung der Betreuenden. Andererseits kann über den Zusammenhang zwischen *Lehrkompetenz*, *Lernumgebung* und *Lernzuwachs* auch auf die Mechanismen, die Praktikumsqualität ausmachen, geschlussfolgert werden.

Gemäß der Relevanz von *Lernenden*-, *Lehrenden*- und *Lernumgebungs*merkmalen wird das vorzustellende theoretische Modell der Praktikumsqualität als »3L-Modell« bezeichnet.

Wenn es um das *Lernen* geht, so lassen die bisherigen Ausführungen nur zu, dass diese als zu erwerbende *Kompetenzen* der Studierenden formuliert werden. Kompetenzen sind die für Individuen »verfügbaren oder von ihnen erlernbaren kognitiven Fähigkeiten und Fertigkeiten, bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können.« (WEINERT, 2002, S. 27). Die kognitiven Fähigkeiten spiegeln sich z. B. in den Kompetenzbereichen *Fachwissen* oder *Erkenntnisgewinnung* wider, die motivationale Abhängigkeit beispielsweise in *intrinsischer Motivation*. Das erste »L« des 3L-Modells liefert also der kompetenzorientiert zu betrachtende *Lernzuwachs* der Studierenden.

Das *Lehren* in Praktika ist, wie bereits erwähnt, theoretisches und forschungstechnisches Neuland und damit schwierig theoriegeleitet zu füllen. Letztlich muss es um diejenigen Komponenten der *Lehrkompetenz* des/der Betreuenden gehen, die den Anforderungen an die Betreuung von Studierenden im Praktikum gerecht werden. Der LeKo (THIEL et al., 2012) liefert ein Modell der *Lehrkompetenz* für Vorlesungen und Seminare, dass besser als einige andere Instrumente der Evaluation von

²Arbeiten von HUCKE, 1999; ENGELS & HÜLSBUSCH, 2008, vgl. Abschnitt 5.1 und von KREITEN, 2012; NAGEL, 2009, vgl. Abschnitt 5.2.2

Prozessdaten in seinem Konstruktumfang und Theoriebezug aufgestellt erscheint (vgl. SEEQ, Abschnitt 4.5.1). Für die Konstruktauswahl wird dieser einen wichtigen Beitrag liefern können, nebst erforderlichen Ergänzungen für die Interaktion in kleineren Gruppen (vgl. kommender Abschnitt). Das zweite »L« des 3L-Modells liefert daher die *Lehrkompetenz*.

Die *Lernumgebung* kann in Praktika über den Wert des Skriptes oder die Qualität räumlicher Adäquatheit der Experimentierarrangements operationalisiert werden, aber auch organisatorische Qualität spiegelt diese Dimension wider (vgl. kommender Abschnitt). Das dritte »L« des 3L-Modells liefert daher die *Lernumgebung*. Es gilt nun zu spezifizieren, welche relevanten Ergebnis-, Prozess- und Strukturdaten konkret in Konstrukte überführt werden.

7.1. Vorgehen

Aus den bisherigen Abschnitten zum Forschungsstand in Praktika sind bereits einige Konstrukte angesprochen worden, die im Rahmen von Ergebnis-, Prozess- oder Strukturdaten für eine Modellierung von Praktikumsqualität in Frage kommen. Es folgt die Darstellung der Konstrukte, je mit Quellenbezug und Argumentation, warum ebenjenes Konstrukt relevant für Praktikumsqualität ist. Zudem wird nach Bedarf eine Abgrenzung zu verwandten Konstrukten vorgenommen (AERA et al., 2014, S. 11) und der Ausschluss möglicher ergänzender Konstrukte begründet. Prinzipiell handelt es sich für das 3L-Modell immer um solche Konstrukte, die im Einklang mit generalisierbaren Zielen aller naturwissenschaftlichen Praktika stehen oder nach den benannten Autoren stehen sollten.³

7.2. Lernzuwachs der Studierenden

Für den Bereich des *Lernzuwachses* sollte im Hinblick wird »curriculare Validität« (MESSICK, 1995, S. 745) auf die anerkannten Kompetenzbereiche für die Naturwissenschaften und die Hochschule spezifiziert werden. Im Idealfall sollten hierfür bereits Modellierungen für den Hochschulbereich vorliegen (z. B. KREITEN, 2012; PSILLOS & NIEDDERER, 2002) oder die Modellierung in großem Maße übertragbar

³Eine Ausnahme bildet das Konstrukt *Offenheit* nach FRASER & McROBBIE (1995). Forderungen zu offeneren Lernarrangements in Praktika finden sich meist im Ausblick von Forschungsarbeiten zu Praktika (HUCKE, 1999, SANDER, 2000) und könnten durchaus eine Qualitätsfacette von Praktika darstellen. Nach der deutschsprachigen Literatur und den zahlreichen Praktikumserfahrungen des Autors ist gleichzeitig aber im deutschsprachigen Raum für dieses Konstrukt ein klarer Bodeneffekt zu erwarten, da offene Grundpraktika deutschlandweit eine echte Rarität darstellen. Eine bekannte Ausnahme bildet lediglich das Projektpraktikum der Technischen Universität Berlin. Es sollte trotzdem, ggf. auf anderen Wegen als der Evaluation, die Kritik an zu stark geschlossenen Praktikumsveranstaltungen aufrechterhalten werden. Prinzipiell könnte das Konstrukt *Offenheit* auch für Fortgeschrittenen-Praktika relevanter werden (vgl. Ausblick).

auf die Hochschulrealität in den Praktika sein (z. B. bei Schülerpraktika; FRASER & McROBBIE, 1995). Dies geht einher mit der Vereinbarkeit der Qualitätsdimensionen mit den Zielen der experimentellen Ausbildung nach HALLER (1999). Im Folgenden werden dazu die relevanten Subdimensionen der Qualitätsdimension *Lernzuwachs* dargestellt und sowohl die theoretische Einbettung als auch der Praxisbezug zur Lehrrealität in Praktika skizziert.

Subdimension 1.1: Fachwissen: Die Theorie-Praxis-Verknüpfung gilt als Ziel von Praktika mit dem höchsten Konsens (vgl. Abschnitt 3.4.2 auf Seite 30). Das Lernen von fachwissenschaftlicher Theorie wurde in Praktika dabei aber als teils zu gering festgestellt (vgl. Abschnitte 5.1 & 5.2.1)⁴, ebenso wie das Vorwissen der Studierenden (vgl. Abschnitte 5.2.1 & 5.2.5).⁵ Dies begründet die Relevanz der Erfassung des Fachwissens in Praktika.

Definition: Das Konstrukt *Fachwissen* beschreibt die Fähigkeit, im Praktikum behandelte Inhalte, experimentelle Besonderheiten und Modelle sowohl deklarativ (LAGOWSKI, 2002) als auch vernetzend erläutern zu können. Anders als alle anderen Konstrukte des *Lernzuwachses* wird beim Fachwissen keine Veränderung modelliert, sondern ein Zustand zu Ende des Praktikums. Dies ist dem Umstand geschuldet, dass Studierende mit umfangreichen Vorerfahrungen durchaus bereits zu Praktikumsbeginn alle inhaltlichen Aspekte beherrschen können. Würde nun eine »Verbesserung« des Fachwissens als Indikator für Praktikumsqualität gebraucht, so wäre dies kein eindeutiger Indikator mehr für Praktikumsqualität, da bei einigen Studierenden ohnehin keine große Steigerung mehr möglich wäre.

Subdimension 1.2: Experimentelle Kompetenz: Zweitwichtigstes Ziel ist der *Erwerb experimenteller Fähigkeiten* bzw. das Erlernen des gesamten Erkenntnisgewinnungsprozesses, verknüpft mit *Methoden wissenschaftlichen Denkens* (vgl. Abschnitt 3.4.2). Studien zeigten geringe experimentelle Vorkenntnisse der Studierenden (vgl. Abschnitt 5.2.2)⁶ und im Prozess ein unreflektiertes Abarbeiten experimenteller Schritte, deren Lernwirksamkeit angezweifelt wird (vgl. Abschnitt 5.2.3)⁷. Summativ ergaben sich folgerichtig teils mangelhafte experimentelle Fähigkeiten

⁴ Arbeit von HUCKE, 1999, vgl. Abschnitt 5.1 und Arbeiten von ELLIOTT et al., 2008; ENGELS & HÜLSBUSCH 2008; HUCKE, 1999; KREITEN, 2012; PLOMER, 2011 und THEYSSEN, 2000, vgl. Abschnitt 5.2.1

⁵ Arbeiten von KISSMANN et al., 2010; KREITEN, 2012; THEYSSEN, 2000, vgl. 5.2.1 und Arbeit von FRICKE et al. (2011), vgl. 5.2.5

⁶ Arbeiten von KREITEN, 2012; NEUMANN, 2004 und PLOMER, 2011]

⁷ Arbeiten von HUCKE, 1999; KREITEN, 2012; RUICKOLDT, 1996.

(vgl. Abschnitt 5.2.3)⁸. Daher ist die Erfassung experimenteller Kompetenz als Abbild des gesamten Erkenntnisgewinnungsprozesses wichtig für Praktika.

Definition: Das Konstrukt *Experimentelle Kompetenz* beschreibt verschiedene Teilkompetenzen, die im Umgang mit dem Experiment und seiner Vor- und Nachbereitung verbessert werden können (Erkenntnisgewinnungsprozess). Strukturell wurde sich hierbei am Modell experimenteller Kompetenz nach SCHREIBER (2012) orientiert⁹, welches zunächst eine sehr kleinschrittige Modellierung des Experimentierprozesses erlaubt. Für Praktika erscheinen insbesondere die Schritte *Fragestellung klären*, *Erwartungen formulieren*, *Versuchsplan entwerfen*, *Geräte zusammenstellen*, *Versuchsanordnung aufbauen*, *Messungen durchführen*, *Messungen dokumentieren*, *mit Problemen und Fehlern umgehen*, *Rohdaten graphisch & tabellarisch darstellen*, *mit Daten Berechnungen anstellen* und *Ergebnis interpretieren* relevant. Dem Konstrukt *Fragestellung klären* kommt hier nach SCHREIBER (2012, S. 38) eine besondere Wichtigkeit zu, da es sich häufig um vorgegebene Fragestellungen im Praktikum handelt, und nicht durch natürliche Genese des Erkenntnisgewinnungsprozesses von den Studierenden geäußerte. Wird nun die vorgegebene Fragestellung nicht hinreichend geklärt, so wird der Erkenntnisgewinnungsprozess um fast seiner gesamten Vorbereitungsphase beschnitten (SCHULZ, 2011, S. 41). Es ist zudem zu betonen, dass es keine strikte, sprunglose Reihenfolge der einzelnen Schritte gibt.

Das eigenständige Bilden von Hypothesen spielt für Grundpraktika keine bzw. eine untergeordnete Rolle (vgl. Abschnitt 5) und wird meist erst in den F-Praktika forciert, weshalb dieses Konstrukt nicht Bestandteil des Modells ist. Ebenso wenig wurden »nature of science«-Aspekte miteinbezogen, da diese die Metaebene des Experimentierens betreffen und nicht die Handlungsebene.¹⁰ Zudem sind *beliefs* auch durch gute Praktika nur schwer veränderlich, sie benötigen reichhaltigere und zeitüberdauernde Erfahrungen (BRYAN, 2012).

Das Konstrukt der *Experimentellen Kompetenz* erfüllt zudem die Eigenschaften einer *Methodenkompetenz* nach BRAUN & GUSY (2006), da es als Problemlöseprozess angesehen wird (AAPT, 1998, S. 484). Das in der Praktikumliteratur beliebte »Konstrukt« Theorie-Praxis-Vernetzung wurde nicht als eigenständiges Konstrukt modelliert, da es sich erstens zu stark mit anderen Konstrukten überlappt (*Experimentelle Kompetenz & Fachkompetenz*, *Lehrkompetenz: Relevanz verdeutlichen*, *Lernumgebung: Integration*) und zweitens erneut eine Abstraktionsebene über den

⁸ Arbeiten von FRÜHWEIN & HEINKE, 2007; ZASTROW, 2002 und RUICKOLDT, 1996

⁹ Ähnliche Kompetenzmodelle werden bei MAISEYENKA (2014) oder CHANG et al. (2011, S. 1215) dargestellt.

¹⁰ obwohl PSILLOS & NIEDDERER (2002, S. 24) Untersuchungen dazu fordern

bisherigen Konstrukten liegt (Definition bei SANDER, 2000, S. 32), ähnlich wie schon die nature-of-science-Aspekte.

Subdimension 1.3: Kommunikationskompetenz: Die Ziele von Praktika ergaben, dass sowohl die *schriftliche* als auch die *mündliche Kommunikation* in Praktika gefördert werden sollte (vgl. Abschnitt 3.4.2). Befunde zu schlechter Protokollqualität (vgl. Abschnitt 5.2.4)¹¹ stützen die Notwendigkeit der Messung *schriftlicher Kommunikationskompetenz*. Auch hier lässt sich ein Bezug zu den übergeordneten Zielen von Praktika nach HALLER (1999) (vgl. Abschnitt 3.4.2) in der Dimension *Weiterentwicklung der sozialen Kompetenz, Persönlichkeit und Motivation* finden. Hier heißt es z. B. »allgemeine Fähigkeiten der Kommunikation und der Interaktion weiterzuentwickeln« (ebd., S. 213). Deshalb lässt sich die *Kommunikationskompetenz* als Qualitätskriterium von Praktika definieren.

Definition: Als *Kommunikationskompetenz* im Praktikum wird einerseits die schriftliche, andererseits die mündliche Komponente wissenschaftlicher Kommunikation verstanden. Es geht hierbei um die Verbesserung der Fähigkeiten, wissenschaftliche Texte wie Protokolle zu verfassen, sowie um die adressatengerechte Kommunikation von Messergebnissen und Interpretationen. Gerade das Schreiben von Protokollen in den ersten Semestern bildet eine Art erste Übung für das Verfassen wissenschaftlicher Publikationen (LAGOWSKI, 2002) und nimmt viel Zeit im Praktikum in Anspruch.

Subdimension 1.4: Bewertungskompetenz: Einen Spezialfall in Praktika stellt der Umgang mit Messergebnissen in seiner innerfachlichen Interpretation und Bewertung dar. Gemäß KMK (2005) würde man diesen Kompetenzbereich als *Bewertungskompetenz (innerfachlich)* ansehen, welcher in Praktika genauer betrachtet werden sollte. Vergleichend mit HALLER (1999) ergeben sich einige Überschneidungen mit dem von ihr genannten Grobziel (3), *Methoden wissenschaftlichen Denkens kennenlernen*, dort heißt es z. B. »eine kritische Herangehensweise bei der Interpretation von Daten zu entwickeln« (ebd., S. 212). Auch BRAUN & GUSY (2006, S. 62) formulierten Aussagen, die die Bewertungskompetenz tangieren, wie etwa »nicht nur reproduzierbares Wissen aneignen, sondern diese Kenntnisse auch bewerten«, hier im Kompetenzbereich *Fachwissen*. Der Zuwachs an Bewertungskompetenz als Teil naturwissenschaftlicher Handlungskompetenz kann daher als Qualitätskriterium für Praktika modelliert werden.

Definition: Bei der *Bewertungskompetenz* wird, im Unterschied zum schulischen Verständnis (KMK, 2005), in Praktika lediglich auf das innerfachliche Bewerten

¹¹ Arbeit von LAMMERTZ et al., 2014

Bezug genommen (wie bei KREITEN, 2012, S. 67). So umfasst dieser Kompetenzbereich die Verbesserung in kritischer Interpretation von Daten und die Einschätzung der Grenzen der eigenen Ergebnisse. Erneut gab hier KREITEN mit ihrem Kompetenzmodell für Praktika die hauptsächliche Strukturierung.

Subdimension 1.5: Kooperationskompetenz: Metakompetenzen wie *Kooperationskompetenz* sind ebenso in den Zielsetzungen zu finden (vgl. Abschnitt 3.4.2). Praktika sind eine der wenigen universitären Lehrveranstaltungsformen, die hoch kooperativ angelegt sind (viel Tandemarbeit, geringe Gruppengrößen, vgl. Abschnitte 3.4 & 5.1), weswegen anzunehmen ist, dass *Kooperationskompetenz* am ehesten dort erworben wird. Gemäß des Deutschen Qualifikationsrahmens wird Teamarbeit zudem als notwendige überfachliche Kompetenz angesehen (HRK, 2011). Auch in anderen Publikationen wurde die Wichtigkeit der Kooperation für das naturwissenschaftliche Lernen herausgestellt (z. B. HENSE, 2012, S. 76; AAPT, 1998, S. 484; TIBERGHIE et al., 1998, S. 501), weswegen eine Aufnahme in den Konstruktepool der Praktikumsqualität gerechtfertigt werden kann.

Definition: Unter *Kooperationskompetenz* wird das Vermögen definiert, mit Mitstudierenden derselben Praktikumsgruppe kooperativ agieren zu können.

Subdimension 1.6: Personalkompetenz: Des Weiteren wurde das *Selbstmanagement* als Zielsetzung von Praktika genannt (vgl. Abschnitt 3.4.2)¹², wegen des hohen Maßes an nötiger Selbstregulation in den langen Vor- und Nachbereitungsphasen von Praktika (vgl. Abschnitt 3.4). Als *Personalkompetenz* beschreibt ERPENBECK (2003) in diesem Zusammenhang vor allem intrinsisch-motivationale Aspekte als produktive Einstellung des Individuums gegenüber Lernen und Selbstentwicklung, was auch nach WEINERTs Kompetenzdefinition zum notwendigen Repertoire einer kompetenzorientierten Lehre gehört. Kompetenzerwerb geschieht nur dann, »wenn aufgrund entsprechender handlungsbegleitender Kognitionen genügend Motivation zu angemessenem Handeln gegeben ist.« (KLIEME et al., 2003, S. 74ff). Motivation und Interesse bilden also wichtige Voraussetzungen für den Lernprozess (HENSE, 2012, S. 75, BRAUN & GUSY, 2006, S. 63) Dies lässt sich erneut mit Bestandteilen von HALLERS *Weiterentwicklung der sozialen Kompetenz, Persönlichkeit und Motivation* assoziieren (»Interesse zu entwickeln«; »zu motivieren«, ebd., S. 213). Gemäß Forschungsstand zu Praktika liegen Befunde zu geringem Interesse an Praktika (vgl. Abschnitt 5.2.1)¹³, motivationsverringern dem Erfolgsdruck (vgl.

¹²HRK (2011)

¹³Arbeiten von KISSMANN et al., 2010 und THEYSSEN, 2000

7. Strukturierung von Praktikumsqualität

Abschnitt 5.2.1)¹⁴ und dem bemängelten geringen Forschungsstand zu affektiven Merkmalen der Studierenden in Praktika vor (vgl. Abschnitt 5.1)¹⁵. Dies begründet das lernrelevante Konstrukt *Interesse* für die Praktikumsqualität. BRAUN & GUSY (2006, S. 162) beschreiben *Personalkompetenz* darüber hinaus auch mit dem Konstrukt *Selbstorganisation*. Daher wurde speziell im Bereich der Praktika der Bereich des *Zeitmanagements* (WILD & SCHIEFELE, 1994) mit aufgenommen. Die Ausbildung intrinsischer Motivation durch das Praktikum liefert hierbei wichtige Hinweise auf ein lernförderliches Klima, die Verbesserung des eigenen *Zeitmanagements* ist ein essentieller Faktor im Praktikum, da der enorme zeitliche Aufwand der Vor- und Nachbereitung (METZGER & SCHULMEISTER, 2011) je Woche ein gutes Zeitmanagement fordert. Daher sind beide als Bestandteile der Qualität von Praktika anzusehen.

Definition: Unter *Personalkompetenz* werden die beiden Konstrukte *Interesse* und *Zeitmanagement* zusammengefasst. Als (individuelles) *Interesse* wird die individuelle Wertschätzung der Praktikusthemen und -aufgaben verstanden (SCHNEIDER & HASSELHORN, 2008, S. 47).

Zeitmanagement ist definiert durch die Planung und Kontrolle der Arbeitszeit (WILD & SCHIEFELE, 1994, S. 193).

7.3. Lehrkompetenz bei der Betreuung

Für Konstrukte der *Lehrkompetenz* bei der Betreuung wurde sich primär an drei Quellen orientiert.

(1) Das theoretische Modell hinter dem LeKo-Fragebogen (THIEL et al., 2012) lieferte eine Fülle an theoretisch abgeleiteten Konstrukten, eine theoretisch und empirisch fundierte Strukturierung derselben und die Aussicht auf eine ökonomisch realisierbare Adaption einiger Skalen für ein Evaluationsinstrument. Schließlich wird dies verglichen mit empirisch erforschten Qualitätsmerkmalen für das Experimentieren im Unterricht, sowie theoretischen Qualitätsmerkmalen naturwissenschaftlichen Unterrichts, um der naturwissenschaftlichen Perspektive gerecht zu werden. Da der LeKo bereits in Abschnitt 4.5.1 vorgestellt wurde, wird hier die empirisch erforschte Unterrichtsqualität beim Experimentieren und die theoretische Modellierung von Unterrichtsqualität im naturwissenschaftlichen Unterricht nach SCHULZ (2011) vorgestellt und anschließend diese drei Quellen in Bezug zueinander gesetzt.

(2) SCHULZ untersuchte in 2011 u. a. experimentierspezifische Qualitätsmerkmale für den Chemie-Unterricht und erforschte Korrelationen zum *Lernzuwachs* der

¹⁴ Arbeiten von HAAKE & MÜLLER, 2011; KREITEN, 2012 und THEYSEN, 2000

¹⁵ Arbeiten von NAGEL, 2009 und HUCKE, 1999

Schüler*innen, ihrer *Motivation* und ihrem *Interesse* (dreier wichtiger Komponenten auch des 3L-Modells). Ergebnisse für das Lehrendenverhalten waren, dass der *Raum für Schüler*innenerklärungen*, die *mündliche Sicherung durch die Lehrperson*, ihre *Instruktionseffizienz*, die *Offenheit der Auswertung*, *Strukturierungshilfen*, eine *Schüler*innenorientierung* und die *Nachvollziehbarkeit des Experiments* qualitätsrelevante Konstrukte waren. Bezüglich der theoretischen Qualitätsmerkmale naturwissenschaftlichen Unterrichts (SCHULZ, 2011, S. 13ff) sind unter Einnahme einer Lehrendenperspektive die Herstellung eines *guten Klassenklimas*, *Individualisierung der Lernumgebungen*, das *Anbieten von Lernhilfen* und die *Rückmeldung individueller Lernprozesse* qualitätsrelevant.

(3) Ein Blick in das Verständnis von Qualität in der Hochschullehre nach KOCH (2004, S. 64–69, 205–207, 215–220) zeigt aus Lehrendenperspektive die Konstrukte *anschauliche Beispiele bringen*, *strukturierte Planung*, *allgemeine Verständlichkeit*, das *Stellen effizienter Aufgaben zur Wiederholung*, *kritisch mit Theorien umzugehen*, die *Aufmerksamkeit der Studierenden wecken*, *Verbesserungsmöglichkeiten aufzeigen*, *Enthusiasmus und Engagement ausstrahlen*, die *Variabilität des Lehrstils*, das *zur eigenen Wissensüberprüfung und zum Nachdenken anregen*, die *Eigenaktivität fördern*, sowie *Überforderung diagnostizieren* zu können. Tab. 7.1 stellt die angesprochenen drei Quellen gegenüber.

Bis auf für *Überforderung diagnostizieren können* finden sich für alle Konstrukte nach KOCH (2004) oder SCHULZ (2011) Entsprechungen in den LeKo-Konstrukten, weshalb dieser als Primärquelle für die Modellierung einer *Lehrkompetenz* in Praktika geeignet erscheint und die enthaltenen Konstrukte gemäß Mehrfachnennung in ihrer Relevanz gestützt werden. Eine *Diagnostische Kompetenz* sollte zusätzlich mit erfasst werden, bedingt durch die enge Zusammenarbeit mit den Studierenden innerhalb vergleichsweise kleiner Experimentiergruppen. Damit wird das Konstrukt *Überforderung diagnostizieren können* nach KOCH (2004) mit modelliert. Es wurden unter Berücksichtigung empirisch erforschter Charakteristika von Praktika (vgl. Abschnitt 5) solche Konstrukte ausgeschlossen, die die Betreuung im Praktikum mutmaßlich nicht tangieren und solche ergänzt, die eine Erhöhung der Betreuungsqualität gerechtfertigt annehmen lassen.

Subdimension 2.1: Verständlich erklären (adaptiert nach THIEL et al., 2012): Das Konstrukt *verständlich erklären* umfasst die Fähigkeit des/r Betreuenden, Begriffe und Sachverhalte nachvollziehbar veranschaulichen und erklären zu können, auch bei Verständnisproblemen oder Nachfragen. Schlüsselzeitpunkte sind hierbei das Vorgespräch und etwaige Nachbesprechungen im Praktikum. Die Rele-

Tabelle 7.1.: Vergleich möglicher Konstrukte für eine *Lehrkompetenz* in Praktika.

Konstrukte der <i>Lehrkompetenz</i> nach...		
LeKo (THIEL et al., 2012)	KÖCH (2004)	SCHULZ (2011)
Klar und strukturiert darstellen	Anschauliche Beispiele bringen	Strukturierungshilfen
Verständlich erklären	allgemeine Verständlichkeit	Instruktionseffizienz
Zusammenfassen und Wiederholen	Stellen effizienter Aufgaben zur Wiederholung	mündliche Sicherung durch die Lehrperson
Verknüpfungen herstellen	Kritisch mit Theorien umgehen	
Verständnis überprüfen	Aufmerksamkeit der Studierenden wecken	Raum für Lernenden-Erklärungen
Lerndienliches Feedback geben	Verbesserungsmöglichkeiten aufzeigen	Rückmeldung individueller Lernprozesse
Gute Lernatmosphäre herstellen	Enthusiasmus und Engagement ausstrahlen	Herstellung eines guten Gruppenklimas
Monotonie vermeiden	Variabilität des Lehrstils	
zur Selbsttätigkeit anregen	Eigenaktivität fördern	Offenheit der Auswertung
Relevanz verdeutlichen		
Ermutigen und Selbstwirksamkeit stärken		Lernenden-Orientierung
Angemessene Anforderungen formulieren		Individualisierung der Lernumgebungen
Effizient mit Störungen umgehen		
Zeit nutzen		
	Überforderung diagnostizieren können	

vanz liegt darin, dass den Studierenden so das Verständnis der Theorie erleichtert wird, was einen Einfluss auf das Ausbilden von Fachkompetenz nahelegt und somit als Qualitätsbeitrag angesehen werden kann.

Subdimension 2.2: Zusammenfassen & Wiederholen (adaptiert nach THIEL et al., 2012): Das Konstrukt *zusammenfassen und wiederholen* betrifft die Tätigkeiten der/des Betreuenden, die Aspekte der Theorie oder des Experiments zusammenzufassen und zu pointieren, die besonders zentral sind. Dies dient den Studierenden als Organisationshilfe beim Lernen und ist daher auch als Qualitätsaspekt von Lehre in Praktika zu sehen.

Subdimension 2.3: Verknüpfungen herstellen (adaptiert nach THIEL et al., 2012): Das Konstrukt *Verknüpfungen herstellen* umfasst das Herstellen von Bezügen aktueller Anforderungen und Themen zu vergangenen Themen. Dies kann einerseits das Vorwissen der Studierenden sein, andererseits bereits im Praktikum behandelte Sachverhalte. Dies erleichtert erneut vor dem Hintergrund einer konstruktivistischen Perspektive den Lernprozess.

Subdimension 2.4: Verständnis überprüfen (adaptiert nach THIEL et al., 2012): Beim Konstrukt *Verständnis überprüfen* geht es darum, dass der/die Betreuende den Studierenden durch gezielte Fragen ermöglicht, das eigene Verständnis des Sachverhalts zu reflektieren und einen etwaigen Erklärungsbedarf auszumachen. Dies unterstützt die Studierenden in der Selbstdiagnose des Lernprozesses und kann daher vertieftes Lernen begünstigen. Gerade im Vorgespräch kann dies zu einem besseren Verständnis der Theorie und des experimentellen Aufbaus im Praktikum beitragen.

Subdimension 2.5: Lerndienliches Feedback geben (adaptiert nach THIEL et al., 2012): Feedback ist essentiell für die Selbstdiagnose und das nachfolgende Lernen, weshalb gute Betreuung auch bedeutet, lerndienliches, konstruktives Feedback z. B. zum Experimentieren zu geben.

Subdimension 2.6: Gute Lernatmosphäre herstellen (adaptiert nach THIEL et al., 2012): Dieses Konstrukt umfasst die Fähigkeit des/r Betreuenden, ein lernförderliches Klima im Praktikum herzustellen, was Lernprozesse begünstigt. Hierbei wird also die zwischenmenschliche Ebene zwischen Betreuer*in und Studierenden erfasst, was von Belang für die Lernqualität ist.

Subdimension 2.7: Zur Selbsttätigkeit anregen (adaptiert nach THIEL et al., 2012): Dieser Bereich zielt auf eine Aktivierung der Studierenden durch den/die Betreuenden ab, welches zu einer persönlichen Auseinandersetzung mit dem

7. Strukturierung von Praktikumsqualität

Lerngegenstand führt. So kann etwa verhindert werden, dass ein Tandem von Studierenden nicht kooperativ, sondern einzeln abwechselnd und pausierend arbeitet.

Subdimension 2.8: Relevanz verdeutlichen (Adaptiert nach THIEL et al., 2012): Bei oftmals stark innerfachlich orientierten Praktika kann es vorkommen, dass Studierende den Sinn der Versuche in Frage stellen. Lernen geschieht aber auf Basis individueller Erfahrungen und ist ein situativer Prozess (HENSE, 2012, S. 75-76). Gute Betreuer*innen betonen daher die Relevanz der Inhalte und Methoden für Anwendungszusammenhänge bzw. das weitere Studium der Teilnehmer*innen.

Subdimension 2.9: Ermutigen und Selbstwirksamkeit stärken (adaptiert nach THIEL et al., 2012): Gerade bei den ersten Erfahrungen mit einer neuen Experimentierkultur können bei den Studierenden Probleme durch das Nicht-Bewältigen von experimentellen Aufgaben bzw. die Vermutung desselben entstehen. Gute Betreuende stärken daher die Selbstwirksamkeitserwartung durch den Ausdruck von Vertrauen und Ermutigungen.

Subdimension 2.10: Effizient mit Störungen umgehen (adaptiert nach THIEL et al., 2012): Auch in der universitären Lehre und gerade bei ungünstigen räumlichen Bedingungen (vgl. Subdimension 2.6) können Störungen und Unruhe in der Praktikumsgruppe entstehen, welche es von dem/r Betreuenden zu bewältigen gilt.

Subdimension 2.11: Diagnostik vornehmen: Die Praktikumsbetreuung verlangt eine adäquate *Diagnostik des Lernfortschritts* (DITTON & MERZ, 1995), z. B. bei Überforderung der Studierenden (KOCH, 2004), bedingt durch die enge Zusammenarbeit mit den Studierenden innerhalb vergleichsweise kleiner Experimentiergruppen und der Relevanz des Schutzes der teuren Praktikumsgeräte. So sollte ein/e Praktikumsbetreuer/in Probleme im Verständnis und im Experiment frühzeitig erkennen, um dann eine passende Maßnahme ergreifen zu können. Selbst wenn diese gewählte Maßnahme aus Nicht-Helfen besteht, so sollte sie dennoch bewusst geschehen.

Die verbleibenden Konstrukte des Modells nach THIEL et al. (2012) wurden nicht adaptiert, da sie nicht im Kern die Aufgaben eines/r Praktikumsbetreuenden treffen. *Klar und strukturiert darstellen*, *didaktische Hilfsmittel zielführend einsetzen*, *Monotonie vermeiden* und *Zeit nutzen* sind nur sinnvoll bei umfassenden frontalen Phasen, die im Praktikum selten vorgesehen sind. *Angemessene Anforderungen formulieren* wiederum liegt meist nicht in der Macht des/r Betreuenden, da meist die Praktikumsleitenden dies festlegen.

7.4. Lernumgebungsqualität

Für Strukturdaten, also Daten, die die Organisation und die *Lernumgebung* betreffen, gilt ein unmittelbarer Praxisbezug zur Lehrveranstaltungsform Praktikum, da Medien, Räumlichkeiten und Abläufe im Fokus stehen. Im Zuge dessen liefern hauptsächlich praktikumsspezifische Quellen und Studien Evidenz für die Konkretisierung dieser Datenform in Konstrukte.

Im Rahmen der Lehrmedien kommen hierbei das Praktikumsskript, die räumliche Ausstattung und die Gerätschaften zu den Experimenten in Frage.

Das Skript sticht hierbei in der Literatur durch seine Relevanz für den Lernerfolg im Praktikum hervor (HUCKE, 1999, S. 126, 132, 149; NAGEL, 2009, S. 116; ZASTROW, 2001, S. 11, vgl. 5.1). So finden sich diverse Ansätze im Rahmen von Forschungsprojekten, die eine Verbesserung der Praktikumsqualität vor allem durch die Verbesserung des Skriptmaterials erreichen wollen (FRICKE & SCHECKER, 2012; MÜHLENBRUCH et al., 2014; NAGEL, 2009; ZASTROW, 2001).

Die räumliche Ausstattung spielt insofern eine Rolle, da Funktionalität von Geräten, sowie ausreichend Platz und Experimentiermaterial eine Grundvoraussetzung für eine gute Lernatmosphäre darstellen.

Im Rahmen der Organisation von Praktika lässt sich durch verschiedene Ansätze, teilweise sogar innerhalb eines Fachbereichs an der Universität, vor allem ein Aspekt generalisieren, der die Organisation der Verknüpfung von Lehrveranstaltungen betrifft. Nach FRASER & MCROBBIE (1995) liefert die sogenannte *Integration* einen Qualitätsbeitrag zu Praktika. Sie verstehen darunter die organisatorische und inhaltliche Abstimmung der eher theoretischen Grundlagenvorlesungen mit den Praktika. Im Idealfall stimmt sich also etwa die Experimentalphysikvorlesung in ihren Inhalten mit den im Praktikum behandelten Themen ab, so dass eine bessere Verzahnung von Theorie und Praxis (vgl. HALLER, 1999; SANDER, 2000) gewährleistet werden kann:

There is little attempt to make the lecture and laboratory instruction correlate with each other in some pedagogically useful way.

LAGOWSKI (2002, S. 3)

Auch in den Zielen von Praktika (vgl. Abschnitt 3.4.2) findet sich dieser Aspekt wieder.

Subdimension 3.1: Skriptqualität: Das Skript kommt in einer funktional möglicherweise trennbaren Doppelrolle daher. Einerseits soll es als theoretisches Nachschlagewerk zur inhaltlichen Vorbereitung dienen (vgl. Abschnitt 3.4), andererseits

7. Strukturierung von Praktikumsqualität

ist eine meist kleinschrittige Versuchsanleitung enthalten, die das Experimentieren in der Durchführungsphase anleitet (abgegrenzt als Konstrukt *Skriptqualität für den Versuch*). Als Bestandteile der *allgemeinen Skriptqualität* werden funktionelle Ziele definiert, die in der Literatur an ein gutes Praktikumsskript gestellt wurden. Ergänzt werden diese um allgemeine Qualitätskriterien aus der Lehrbuchforschung. Primäre und offensichtlichste Aufgabe des Skripts ist es, fehlerfrei auf die theoretischen Grundlagen des kommenden Versuchs vorzubereiten, also als Medium im Selbststudium eine solche Vorbereitung zu erlauben (FISCHLER, 1979). Natürlich können auch weitere Lehrbücher zur Vorbereitung herangezogen werden, in der Literatur finden sich aber Hinweise darauf, dass das Skript häufig das zentrale Medium für diesen Aspekt darstellt (ZASTROW, 2001, S. 11). In der Nachbereitung des Praktikums bildet das Schreiben des Protokolls einen großen Schwerpunkt der studentischen Tätigkeit. Ein Skript als begleitendes Praktikumsmedium sollte also auch hierfür Informationen bereitstellen. Zu guter Letzt sind Qualitätskriterien aus der Lehrbuchforschung zu nennen, z. B. eine *gute Strukturierung* (MERZYN, 1994) des Skripts oder enthaltene, *transparente Lernziele* (PRENZEL, 1997).

Subdimension 3.2: Skriptqualität für den Versuch: Da der Versuch für die Präsenzzeit im Zentrum des Praktikums steht, sollte ein qualitativ hochwertiges Praktikumsskript auch auf diesen vorbereiten. Dies wird mit dem Konstrukt *Skriptqualität für den Versuch* genauer definiert. Es geht vor allem um die Funktionalität des Skripts als Instrument für die experimentelle Vorbereitung im Selbststudium zu Hause, wie es klassische Praktikumsrealität ist (REHFELDT et al., 2015b). Das Skript dient als eine Art Praxishandbuch, wie die einzelnen Versuche zu bearbeiten sind, mit meist sehr konkreten Hilfestellungen zum Experimentieren und zu den Geräten. Dies kann sich durch eine Unterstützung in der Vorbereitungsphase ausdrücken, wenn das Skript eine gute Vorstellung vom Versuchsaufbau gibt, ebenso können dort neue Geräte vorgestellt werden und in der Durchführung die Schritte begleitet werden (z. B. THEYSEN, 2000, S. 71).

Subdimension 3.3: Räumlich-sachliche Ausstattung: Die *Räumlich-sachliche Ausstattung* definiert sich über ausreichende Materialien, genügend große Räumlichkeiten (TIBERGHIE et al., 1998) und erreichbares Begleitmaterial als formatives Konstrukt. Dies kann als basale Grundvoraussetzung für erfolgreiches Lernen angesehen werden.

Subdimension 3.4: Integration der Vorlesung: Das Konstrukt *Integration der Vorlesung* definiert sich über die organisatorische und inhaltliche Abstimmung der eher theoretischen Grundlagenvorlesungen mit den Praktika. Dies liefert einen

7.5. Zusammenfassung: 3L-Modell der Praktikumsqualität

Beitrag zum Theorie-Praxis-Transfer (HALLER, 1999). Folgt man etwa der Definition von Praktika im Physikstudium der Freien Universität Berlin, so ist dieser Umstand bereits strukturell gewollt:

In Veranstaltungen, die Teil eines Praktikums sein können, soll besonders auf Lehrinhalte in den Praktika eingegangen [...] werden.

FUB (2012b, §5, S. 1179)

Auch im Rahmen des Bologna-Prozesses wird eine solche »innere Lerndramaturgie« (BOSBACH et al., 2007, S. 167–168) gefordert.

Weitere Konstrukte, wie etwa die *Organisation* des Praktikums, sind aus Sicht der Studierenden schwierig zu erfassen, da es hintergründige Prozesse des Praktikums betrifft. Hierfür wäre eher eine tiefenanalytische Methode sinnvoll (wie bei REHFELDT et al., 2015b gezeigt), da dort sowohl die Betreuenden, als auch die Leitenden des Praktikums in die Evaluationserhebungen miteinbezogen werden. Natürlich sind auch weitere Konstrukte der *Lehrkompetenz* (etwa *fachliche Kompetenzen*) oder der *Lernumgebung* (z. B. *Offenheit*) denkbar für ein Modell der Praktikumsqualität. Speziell die *fachlichen Kompetenzen* der Betreuer*innen sind allerdings in noch keiner Studie zu Praktika negativ bewertet worden und die *Offenheit* der Lernumgebung wird wohl gerade in Anfängerpraktika überwiegend gering und damit die Erfassung derselben informationsarm sein (vgl. Argumentation in Abschnitt 7.1). Zur weiteren Absicherung, ob die implementierten Konstrukte auch die Anforderungen an Praktika erschöpfend abbilden, sei auf die Validierungsstudien 1.1 und 1.2 in Kapitel 12 verwiesen.

7.5. Zusammenfassung: 3L-Modell der Praktikumsqualität

Es wurde eine Spezifizierung der Konstrukte für die drei Dimensionen des 3L-Modells vorgenommen (Abb. 7.1). Langfristig könnte das Modell funktionale Beziehungen zwischen den Konstrukten (Kausalbeziehungen) enthalten, die empirisch ermittelt werden können.

Eine zu erwartende Kausalbeziehung ist diejenige zwischen *Lehrkompetenz* und *Lernumgebung* bezüglich des *Lernzuwachses*, also dass sich gutes *Lehrendenverhalten* und eine lerndienliche *Lernumgebung* positiv auf den *Lernzuwachs* der am Praktikum teilnehmenden Studierenden auswirken (output). Diese und weitere Kausalbeziehungen müssen aber empirisch abgesichert werden, um das 3L-Modell weiter zu erforschen.

7. Strukturierung von Praktikumsqualität

Das 3L-Modell der Praktikumsqualität weist durch seinen fächerübergreifenden und generalisierenden Charakter inhaltliche Grenzen auf. Es ist nicht für Demonstrationspraktika, wie sie im Rahmen der Lehramtsausbildung stattfinden, und nicht für F-Praktika per se geeignet, da diese Lehrveranstaltungen i. d. R. offener und inhaltsdifferenzierter gestaltet sind, was andere Indikatoren für die definierten Dimensionen erfordern würde. Sehr praktikumsspezifische Konstrukte wie die konkrete *Organisation der Praktika*, die *Fachkompetenz der Betreuenden* und die *Aufbereitung und Wahl der Experimente* werden allerdings nicht mit berücksichtigt, da sie besser in Tiefenanalysen von Praktika zu erfassen sind (REHFELDT et al., 2015b).

Zusammenfassend liefert das 3L-Modell der Praktikumsqualität einen theoretischen Rahmen für eine fundierte Lehrevaluation in regulären naturwissenschaftlichen Experimental-Hochschul-Praktika.

7.5. Zusammenfassung: 3L-Modell der Praktikumsqualität



Abbildung 7.1.: Theoretisches 3L-Modell der Praktikumsqualität mit den drei Hauptdimensionen *Lernzuwachs*, *Lehrkompetenz* und *Lernumgebung*.

Kapitel 8

Ziele und Forschungsfragen

Many research studies discuss proposals to improve laboratory activities or create new types of laboratory activities rather than analyzing the current practice with regard to laboratory activities.

TIBERGHIE et al. (1998, S. 484)

Aus der Darstellung des Forschungsdesiderats in Abschnitt 6.5 wurde die Notwendigkeit a) eines theoretischen Modells der Praktikumsqualität, b) einer Adaption von Instrumenten für die Beurteilung allgemeiner *Lehrkompetenz* für die Betreuungsanforderungen und c) einer übergreifenden, flächendeckend einsetzbaren Evaluationsmethode deutlich.

Im Zuge der Erarbeitung des 3L-Modells der Praktikumsqualität wurde die Notwendigkeit a) erfüllt, die zweite und dritte Anforderung betreffen die Erarbeitung neuer Instrumente für die Evaluation von Praktika. Dem obigen Zitat von TIBERGHIE et al. folgend, soll es in dieser Arbeit anders als bei vielen Ansätzen zur Verbesserung bisheriger Praktika also nicht darum gehen, teilweise nur schwach geprüfte Idee zur Verbesserung von Praktika zu implementieren und die Wirksamkeit nachzuweisen. Es geht vielmehr darum, standardisierte Methoden und Instrumente zu konzipieren, die zunächst eine fundierte Analyse von Praktika möglich machen. In einem zweiten Schritt können darauf aufbauend gezielt tiefere Analysen der problematischen Bereiche vorgenommen werden.

8.1. Ziele zum PraQ-Instrument

Innerhalb eines Instruments zur Messung von Praktikums-Qualität (PraQ) sollen die Forderungen b) und c) des Forschungsdesiderats aufgegriffen werden. Das flächendeckend einsetzbare Instrument PraQ soll Interpretationen von Qualitätsmessungen in Praktika für den deutschsprachigen Raum ermöglichen (»externe Validität« der Interpretation). Eine Testentwicklung und -evaluation ist notwendig und wichtig, da der Forschungsstand eine durchwachsene Qualität von Praktika bei gleichzeitigem Mangel an ökonomisch einsetzbaren, beforschten Evaluationsinstrumenten und -methoden für diese Lehrform aufgezeigt hat. PraQ forciert eine summativ ausgerichteten Evaluation von Praktika, kann eine Zusammenfassung der Gesamtsituation

8. Ziele und Forschungsfragen

aufzeigen und zielt auf einen Überblick über das Praktikum im Vergleich zu anderen Praktika ab (BERK, 2005). Die konkreten Ziele für den PraQ sind:

Ziel 1: Es soll eine fundierte empirische Erfassung aller für die Lehre relevanten Aspekte der Qualität von Praktika realisiert werden, die zeit- und materialökonomisch verläuft. Darüber soll eine Etablierung des Instruments als reguläres Evaluationsinstrument von Praktika in Hochschulen möglich werden.

Ziel 2: Es sollen verschiedene Interpretationen der Testergebnisse empirisch abgesichert werden (»Validierung«). Demnach soll es auch möglich sein, mit der Interpretation der Evaluationsergebnisse fachdidaktische Forschung betreiben zu können.

Da es sich in beiden Fällen um Methoden zur Evaluation handelt, werden die meisten Forschungsfragen im Rahmen der Testevaluationsstudien diskutiert (vgl. Abschnitt 12). Im empirischen Teil finden sich nach der Vorstellung aller Testevaluationsstudien im Kapitel 13 Beispiele für inhaltliche Forschungsfragen für die Konstrukte des PraQ, die vor allem eine Profilierung von Praktika und lehrrelevante Gruppenunterschiede zum Gegenstand haben. Zur Übersicht sind diese Forschungsfragen hier schon einmal aufgeführt, sie werden in den entsprechenden Abschnitten abgeleitet und näher ausgeführt:

8.2. Forschungsfragen

Die ersten sechs Forschungsfragen betreffen die Testevaluation (»Validierung«) bezüglich des PraQ-Instruments bzw. dessen Testwertinterpretation. Nach klassischer Bezeichnung der Testvalidierung fokussieren Forschungsfrage 1 und Forschungsfrage 2 die Inhaltsvalidität, Forschungsfrage 3 die faktorielle Validität, Forschungsfrage 4 die konvergente und Forschungsfrage 5 die diskriminante Konstruktvalidität. Forschungsfrage 6 erfasst die Multiperspektivität als Vergleich mit Expert*innenurteilen (MESSICK, 1995).

Forschungsfrage 1: Welche Indikatoren des PraQ sind nicht konstruktrelevant für Praktikumsqualität? (»Inhaltsvalidität«)

Forschungsfrage 2: Welche Aspekte von Praktikumsqualität fehlen im PraQ? (»Inhaltsvalidität«)

Forschungsfrage 3: Welche Konstrukte lassen sich beim PraQ theoretisch wie empirisch voneinander getrennt interpretieren? (»faktorielle Validität«)

Forschungsfrage 4: Wie hängen die Skalen des PraQ statistisch mit Kompetenz-Leistungstests und Selbstwirksamkeitserwartung zusammen? (»konvergente Konstruktvalidität«)

Forschungsfrage 5: Wie hängen die Skalen des PraQ statistisch mit der Messung von allgemeinem Praktikums(sozial)klima zusammen? (»diskriminante Konstruktvalidität«)

Forschungsfrage 6: In welchen Facetten der Praktikumsqualität unterscheiden sich Einschätzungen von Studierenden gegenüber Einschätzungen von Expert*innen? (»Multiperspektivität«)

Die, vorweggenommen, guten Ergebnisse der Testevaluationsstudien ermöglichen die Erforschung erster inhaltlicher Forschungsfragen über eine so noch nie realisierte Anzahl von Praktika im deutschsprachigen Raum. Forschungsfrage 7 dient der Exploration sämtlicher Konstrukte, um ein Qualitätsprofil früher naturwissenschaftlicher Experimentalpraktika aufzustellen und ausgehend von einem Durchschnittsprofil bzw. einem Referenzprofil Vergleiche im Sinne eines Benchmarkings (vgl. Anforderung 7, Abschnitt 4.4 auf Seite 43) anstellen zu können. Forschungsfrage 8 stellt die Frage nach Unterschieden zwischen Chemie- und Physikpraktika. Durch den fächerübergreifend möglichen Einsatz des PraQ erscheint es lohnenswert, Stärken und Schwächen dieser beiden Fachpraktika zu explorieren bzw. miteinander zu vergleichen. Dies könnte einem zweiten Schritt zum Wunsch führen, das betreffende Fachpraktikum genauer zu analysieren, um etwaige Interventionsideen für das jeweils andere Fach zu generieren. Schließlich wird mit Forschungsfrage 9 untersucht, ob der Organisationstyp der Praktika (Semesterkurs oder Blockpraktikum) einen signifikanten Einfluss auf die Praktikumsqualität hat. Je nach Ergebnis könnte dies Informationen zu einem zu bevorzugenden Organisationstypus geben.

Forschungsfrage 7: Was sind Stärken, Schwächen und die qualitative Bandbreite von Praktika in Deutschland in Bezug auf die Praktikumsqualität?

Forschungsfrage 8: Was sind Unterschiede zwischen Chemie- und Physikpraktika in Bezug auf die Praktikumsqualität?

Forschungsfrage 9: Was sind Unterschiede zwischen Block- und Semesterpraktika in Bezug auf die Praktikumsqualität?

II.

Empirischer Teil

Kapitel 9

Einleitung

Im empirischen Teil dieser Arbeit wird die Testkonstruktion und -evaluation eines neuen Instruments für die Evaluation von Praktika vorgestellt, das auf dem 3L-Modell der Praktikumsqualität basiert (Abb. 9.1).

Die von der Evaluation »Betroffenen« gemäß Ziel 1 sind die Studierenden, Betreuenden und Leitenden des Praktikums. Gemäß den Anforderungen an Evaluationsinstrumente ist all diesen Einsicht in Evaluationsergebnisse zu gewähren (vgl. Anforderungen 5, 6 und 10, Abschnitt 4.4). Bezogen auf die Leitgrafik wird nun das Ziel »Monitoring und Benchmarking« beim PraQ-Instrument fokussiert. Gemäß den Anforderungen an gute Evaluation wird hierbei zunächst die Testkonstruktion gemäß 3L-Modell (vgl. Abb. 7.1 auf Seite 97) vorgestellt.

Es folgen Ausführungen zur Objektivität und Reliabilität, wie für quantitative Instrumente üblich.¹ Der Punkt »Validität« betitelt sämtliche PraQ-Testevaluationsstudien, die zur Absicherung von Interpretationsrichtungen durchgeführt wurden. Es wird hierbei einem argument-based approach (KANE, 2001) gefolgt. Der empirische Teil schließt mit inhaltlichen Studien zum PraQ im deutschsprachigen Raum.

¹Warum ein quantitatives Instrument gewählt wurde, wird im kommenden Abschnitt erläutert.

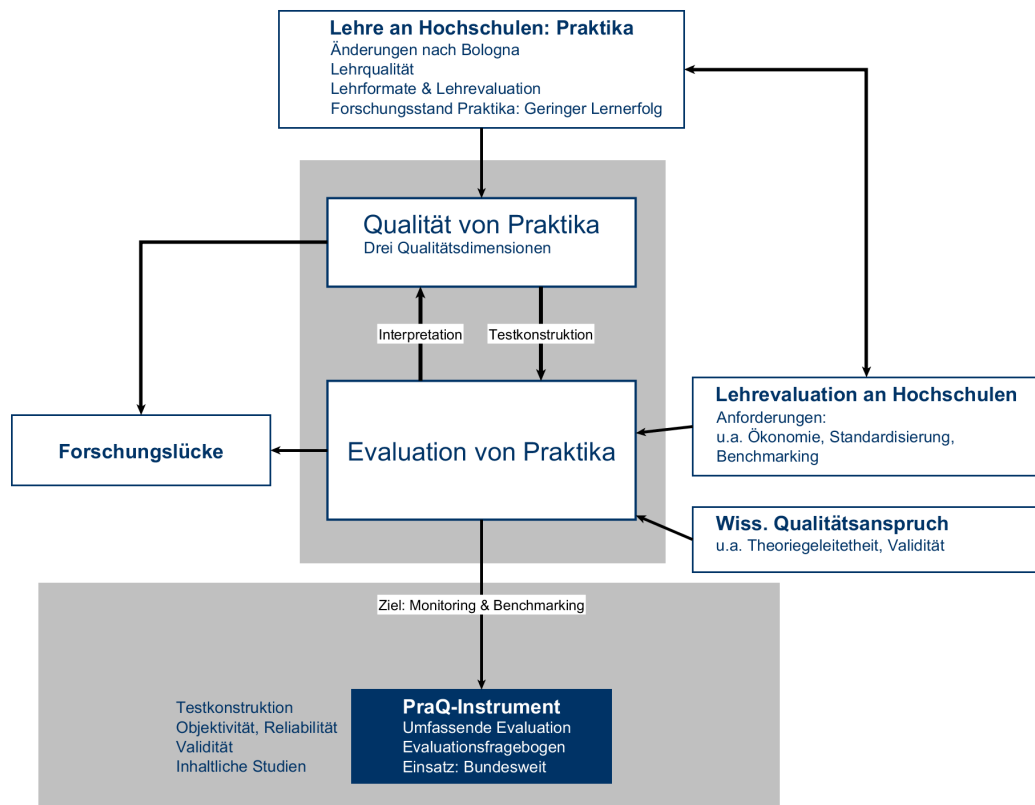


Abbildung 9.1.: Leitgrafik: Die Forschungslücke einer adäquaten »Evaluation von Praktika« wird über die Neumodellierung der »Qualität von Praktika« und die Testkonstruktion geschlossen. Dieses »PraQ-Instrument« wird in seiner Konstruktion, Validierung und Auswertung in den Kapiteln 9 bis 13 dargestellt.

Kapitel 10

Testkonstruktion

10.1. Auswahl der Erhebungsmethode

Auf Grund der Vorteile in Bezug auf die Durchführungs- und Auswertungsökonomie wurde als Messinstrument ein Fragebogen mit Selbst- und Fremdeinschätzungselementen gewählt, die auf einer 6-stufigen Likertskala bzgl. Zustimmung bewertet werden. Hierfür lassen sich auch Akzeptanz- und Implementationsargumente anbringen. Selbsteinschätzungsfragebögen sind als Evaluationsinstrumente von Hochschulen seit Jahren im Einsatz (vgl. Abschnitt 4) und verfügen daher über Vorteile in Implementation und Kosten, da meist Verantwortliche für die Durchführung der Befragungen schon feststehen, die lediglich den Wechsel von einem Fragebogen-Instrument zu einem neuen vornehmen müssten. Es würde keine Einarbeitung in neue Erhebungsformen, eine/n externen Evaluator*in¹ oder neue Auswertungsalgorithmen erfordern, was wiederum auch für eine höhere Akzeptanz und Nutzung des Instruments im Vergleich zu etwa Kompetenztests spräche.

Auch BERK (2005) betont, dass die Messung echter »learning outcomes« für die Evaluation von Lehre wünschenswert, aber sehr umfangreich ist. Eine Evaluation der Kompetenzzuwächse über Selbsteinschätzungen scheint ein guter Kompromiss zu sein, auch da Selbsteinschätzungsfragebögen für standardorientierte Kompetenzmodelle beliebt sind (GRÖSCHNER et al., 2013; SCHNEIDER et al., 2012; FREY, 2005).

Zudem zielt das zu entwickelnde Instrument auf ein deutschlandweites Benchmarking ab, das von einer quantitativen Methode begünstigt wird (vgl. Abschnitt 4.4).

Es wird von manchen Autoren angezweifelt, dass sich Kompetenzen, wie in der ersten Dimension des 3L-Modells festgehalten, über Selbsteinschätzungen verlässlich messen lassen (HARTIG & JUDE, 2007). Es wird auch kritisiert, dass diese Form der Messung nicht Kompetenzen abbildet, sondern eher Selbstwirksamkeitserwartungen bezüglich dieser Kompetenzen (CRAMER, 2010). Dies ist der Grund für einen notwendigen Vergleich des zu konstruierenden PraQ-Instruments mit Kompetenz- und

¹Studierende bzw. allgemein Lernende sind Expert*innen für die Bewertung ihres erhaltenen Unterrichts (SCHULZ, 2011, S. 22), etwa da sie deutlich mehr Zeitstunden mit den Lehrenden verbringen als etwaige externe Evaluator*innen.

Selbstwirksamkeitserwartungsmessungen (vgl. Abschnitt 12.5.1 auf Seite 172). Die Validität von Evaluationsergebnissen, die mittels Einschätzungen durch Studierende zur Qualitätsbeurteilung von Lehre gewonnen werden, wird bei BRAUN et al. (2008, S. 32f) unter Darstellung des Forschungsstands kritisch diskutiert. Sie kommen zu dem Urteil, dass Selbsteinschätzungen auf Grund ihrer guten prädiktiven Validität bezüglich Alternativmessungen geeignet sind, um Kompetenzerwerb darzustellen. Sie zitieren dabei KLIEME et al. (2002, S. 215), dass »Selbstauskünfte ... bei Jugendlichen und Erwachsenen durchaus ein gültiges Bild über zentrale Aspekte ihrer [...] Kompetenz« abgeben.

Allerdings wird auch eine konsequente »Konstruktvalidierung der Instrumente« zur Erfassung der Lehrqualität gefordert (ebd., S. 32). Nach dem argument-based Ansatz der Validität würde man dies als Forderung zur Testevaluation der gewünschten Interpretationsrichtungen übersetzen (KANE, 2001).

RICHARDSON (2005, S. 407) stellt darüber hinaus dar, dass auch die Mehrzahl der Studierenden an die Genauigkeit von Selbsteinschätzungen glauben, was ein Argument für die Akzeptanz der Ergebnisse bei Studierenden liefert. Konkret auf die erstmalige Messung der Entwicklung experimenteller Kompetenz durch Selbsteinschätzungen bezogen, schließen SCHREIBER et al. (2016, S. 3) auf eine »zufriedenstellende Genauigkeit« für Selbsteinschätzungen zum Experimentieren, hier allerdings bei Schüler*innen.

Konkret auf Praktika bezogen setzte STEFANI (1994) Selbstbeurteilungen für Protokolle in Anfängerpraktika der Biologie ein und liefert damit den einzigen Bezugspunkt der aktuellen Forschungslage in Bezug auf die Selbstbeurteilung in Praktika. Die Studie zeigte Korrelationen von .93 zwischen Beurteilungen der Lehrenden und den Selbstbeurteilungen. Dies lässt für Selbstbeurteilungen der eigenen Fähigkeiten in Evaluationen begründet annehmen, dass Selbsteinschätzungen ein verlässliches Maß für Leistungen im Praktikum sind.

Für Fremdeinschätzungen der Lehrqualität, hier der der Betreuenden, zeigt sich ein ähnliches Bild wie für Selbsteinschätzungen. Der im Abschnitt 4.5.1 vorgestellte, international häufig eingesetzte und mit mehreren Testevaluationsstudien bedachte SEEQ-Fragebogen beispielsweise wird trotz seiner Realisierung in Fremdeinschätzungs-Skalen als relativ verzerrungsfrei eingeschätzt (MARSH, 1987). Studien legen zudem nahe, dass Studierendenurteile über ihre Lehre sogar mehr externen Beobachtern ähneln als wenn die Lehrenden sich selbst einschätzen (RINDERMAN, 1996). Ähnlich wie bei Selbsteinschätzungen zeigen sich für Fremdeinschätzungen Evidenzen bezüglich ihrer Akzeptanz bei Lehrenden (SCHULZ, 2011, S. 22).

10.2. Zielgruppe

Nachdem die Erhebungsmethode feststeht, wird die Zielgruppe des zu konstruierenden Instruments definiert, um eine adressatenspezifische Evaluation zu gewährleisten und die gewünschten Interpretationsrichtungen definieren zu können. Zielgruppe sind naturwissenschaftliche Experimental-Grundpraktika an der Hochschule. Als Kriterium für ein »Grund«-Praktikum wurde festgelegt, dass dieses plangemäß vor dem 5. Fachsemester absolviert wird. Als »naturwissenschaftlich« gelten prinzipiell auch alle naturwissenschaftsnahen Fachgebiete wie die Medizin und Ingenieurwissenschaften. Aus ökonomischen Gründen wurden in dieser Arbeit allerdings vorrangig Physik- und Chemiepraktika getestet.

Die Studierenden dieser Praktika befinden sich noch am Anfang ihrer experimentellen Ausbildung und haben daher nur wenig experimentelle Vorerfahrung. Gemäß dem Anspruch eines Überblicks über Praktika in Deutschland und der Antizipation möglicher Synergieeffekte beim Praktikumsvergleich verschiedener Fächer sind die Physik, Chemie, Biologie, Ingenieurwissenschaften und Veterinärmedizin als Fächergruppen Teil der Zielgruppe. Die Gemeinsamkeiten der Praktika dieser Disziplinen und damit eine Generalisierung des Modells lassen sich sowohl (a) auf der organisatorischen Ebene als auch (b) auf der inhaltlichen Ebene begründen.

(a) Organisatorisch weisen die meisten dieser Praktika eine ähnliche personelle und materielle Struktur auf. Die Vorbereitung auf den wöchentlichen Praktikums-termin geschieht oft in Heimarbeit mit einem Praktikumsskript. Die Lehre in den Praktika wird von mehreren Personen als Praktikumsbetreuer*innen durchgeführt, die eine vergleichsweise kleine Studierendengruppe betreuen (Physik z. B. 10:1, HUCKE, 1999, S. 135ff; Chemie 15:1, ENGELS & HÜLSBUSCH, 2008, S. 58). Experimentiert wird hierbei häufig in sehr kleinen Gruppen, z. B. Tandems oder 3er-Gruppen (ebd.). Die Arbeit an den Experimenten folgt in der Regel einem wöchentlichen Turnus, den irgendwann alle Studierenden durchlaufen haben. Ausgewertet wird schließlich meist wieder in Heimarbeit, in der ein Versuchsbericht angefertigt wird. Die strukturellen Gemeinsamkeiten konnten auch in Studien von GUTZLER et al. (2014) bestätigt werden, als die zeitliche Verlaufsstruktur zwischen Praktika der Physik, physikalischen Chemie und Veterinärmedizin verglichen wurden.

(b) Inhaltliche Überschneidungen existieren zunächst auf Basis der übergeordneten Ziele von Praktika (vgl. Abschnitt 3.4.2). Die Fachgebiete überlappen zudem thematisch (LAGOWSKI, 2002, S. 1) und methodisch gibt es einen internationalen Konsens über eine »prototypische Basiskultur experimentellen Arbeitens« (EMDEN, 2011, S. 34; TIBERGHEN et al., 1998, S. 503 argumentiert analog).

Auch in der dem Studium vorangehenden Schulausbildung zeigt sich ein ähnliches Bild. Die KMK (2005) formulierte in den Bildungsstandards identische Kompetenzbereiche für Biologie, Chemie und Physik, die Experimentierkultur ist vergleichbar (EMDEN, 2011).

International ergaben sich auch empirisch über Skriptanalysen der geforderten Tätigkeiten in Praktika Ähnlichkeiten zwischen der Physik, Chemie und Biologie (TIBERGHIE et al., 1998). Die Aufgabenkulturen sind ähnlich, *direkter Report von Beobachtungen* und *indirekte Messung einer Quantität* sind die am stärksten überlappenden Aufgabenbereiche naturwissenschaftlicher Praktika verschiedener Fächer. Unterschiede werden festgestellt für das *Skript* und das *Gesetze bestätigen*, welchem in Physik eine größere Rolle zu Teil wird. Die Tätigkeiten *Etwas herstellen* und *ein Material beobachten* kommen naheliegender Weise nicht in Physik vor (TIBERGHIE et al., 1998, S. 494ff). Auch wenn internationale Befunde nur bedingt auf die deutschsprachige Praktikumskultur übertragbar sind, so geben sie doch Hinweise darauf, dass eine Operationalisierung der Konstrukte des bereits fächerübergreifend angelegten 3L-Modells in Form von Fragebogen-Items auf manche Unterschiede in den Fächern der Zielgruppe eingehen sollte. Geplant ist hier, Items zu wählen, die für alle Fächer der Zielgruppe sinnvoll und konstruktrelevant sind. Abgesichert wird dies u. a. über kognitive Interviews mit Expert*innen aller beteiligter Fächergruppen (vgl. Abschnitt 12.3.1).² Die Expert*inneninterviews sollen die gemeinsame inhaltliche Basis der Items absichern.

10.3. Erstellung des Itempools

10.3.1. Fächerübergreifende Formulierung

Die fächerübergreifende Formulierung von Items lässt sich am besten in Form eines (Gegen-)Beispiels verdeutlichen:

Gegenbeispiel: »Aufgrund dieses Praktikums kann ich statistische Fehlerquellen besser diskutieren.«³

²Prüfungen auf differential item functioning werden ebenso vorgenommen (DIF, Abschnitt 13.2).

³Dieses Gegenbeispiel entstammt dem ersten Itempool, der dem Stand noch vor den Expert*inneninterviews (vgl. Abschnitt 12.3.1) entspricht. Es wurde zurecht, vor allem von den beiden Praktikumsleitenden der Chemie, kritisiert, dass dies eigentlich nur in der physikalischen Chemie vorkäme und auch erst in späteren Praktika Ziel sei. Die Leitenden physikalisch orientierter Praktika stimmten dagegen dem Item in seiner Bedeutung für das Konstrukt *Bewertungskompetenz* zu.

Dies ist ein Beispiel für ein nach Zielgruppe unterschiedlich konstruktrelevantes Item, da statistische Fehlerquellen in chemischen Grundpraktika häufig nicht Teil der Lehre sind.

Beispiel: »Aufgrund dieses Praktikums kann ich den möglichen Einfluss von Fehlerquellen besser diskutieren.«

Dieses Beispielitem aus demselben Konstrukt *Bewertungskompetenz* erfüllt deutlich besser die Anforderungen an eine fächerunabhängige Konstruktrelevanz, da es allgemeiner formuliert ist und dennoch spezifisch die *Bewertungskompetenz* adressiert.⁴

10.3.2. Adaption etablierter Skalen

Bei der theoriegestützten Operationalisierung der Konstrukte des 3L-Modells (vgl. Kapitel 7) in Form von Items wurde darauf geachtet, je Konstrukt mindestens zwei, besser mehr Indikatoren zu finden, die das Konstrukt theoriegemäß abbilden, damit eine Reliabilität schätzbar ist.

Dazu wurde zunächst auf die Adaption etablierter Instrumente zurückgegriffen.

Wie bereits durch die Entstehung des 3L-Modells ersichtlich wird, spielt dabei für die Dimension *Lernzuwachs* der BEvaKomp nach BRAUN & GUSY (2006) eine Rolle, da er das aktuellste deutschsprachige, kompetenzorientierte Evaluationsinstrument studentischer Kompetenzentwicklungen darstellt und in seiner Strukturierung der des 3L-Modells gleicht.⁵

Für die Evaluation des/der Lehrenden ist der Fragebogen zur Erfassung der *Lehrkompetenz* (LeKo) maßgeblicher Bezugspunkt, da er umfassend Konstrukte der *Lehrkompetenz* im Hochschulkontext umfasst. Für die Strukturdaten lag keine Referenz für die Qualität von Lehrbüchern oder Ähnliches vor, so dass die Items zur *Skriptqualität* selbst konstruiert wurden. FRASER & McROBBIE (1995) lieferten die Skala zur *Integration*.

Anpassungen und Ergänzungen der Items auf den Praktikumskontext waren teils notwendig. So wurden beispielsweise für einige Skalen, die sich in ihren Formulierungen nur auf theoretische Inhalte bezogen, Items für die experimentellen Phasen

⁴Das AAPT (1998, S. 484) beispielsweise schreibt, dass es allein schon schwierig wäre, ein gemeinsames Equipment von Physik-Praktika unterschiedlicher Standorte zu beschreiben. HEIDRICH et al. (2013, S. 678) fand bei Analysen von zehn Skripten verschiedener Physikpraktika in Deutschland heraus, dass die Inhalte der Versuche auch innerhalb der Physik so stark variieren, dass ein dies berücksichtigender Test schwierig zu konstruieren sei. Dies bildet zwei Argumente, die einem zu stark kontextualisierten Itempool widersprechen und das Vorgehen in dieser Arbeit rechtfertigen.

⁵International war mit CHANG et al. (2011) ein aktuelleres Instrument zur Erfassung selbst-eingeschätzter Kompetenzen zugänglich. Dieses wurde ebenso mit einbezogen.

des Praktikums ergänzt. Einige Skalen blieben auf Grund ihres fächerübergreifenden Charakters dagegen nahezu unverändert (z. B. *Kooperationskompetenz*), andere wurden stark auf den Praktikumskontext angepasst (z. B. *Zusammenfassen und Wiederholen*: Ergänzung um »Zusammenfassen und Wiederholen experimenteller Schritte«). Bei deutlich mehr als der Hälfte der Konstrukte existierten gemäß Literaturrecherche allerdings keine Operationalisierungen in Form von Selbsteinschätzungen. Hier wurde derart vorgegangen, dass theoretische Modelle zu den Konstrukten recherchiert und daraus eine Ableitung der Indikatoren für das Konstrukt vorgenommen wurden. Die *Experimentelle Kompetenz* beispielsweise kann gemäß Modell nach SCHREIBER (2012) in theoretisch 13 verschiedene Komponenten unterteilt werden. Der nächste Schritt war dann, aus den Modellen diejenigen Bestandteile zu extrahieren, die in Grundpraktika gefördert werden sollen. Hierfür wurden die Ziele von Praktika nach HALLER (1999) zu Rate gezogen. Im Zweifel wurden eher mehr Konstrukte und Indikatoren mit aufgenommen, da diese noch die Testevaluationsstudien durchlaufen würden. Die *Experimentelle Kompetenz* wurde so, in Übereinstimmung mit den Ansichten der Expert*innen (vgl. Abschnitt 12.3.1) um zwei theoretische Komponenten reduziert, so dass elf Komponenten übrig blieben. Alsdann wurden für diese Komponenten Items konstruiert, die sich je an mindestens einer praktikumsorientierten Quelle oder einer Schlussfolgerung aus selbiger orientierten. Für die Experimentelle Kompetenz beispielsweise wurde sich vorzugsweise an SCHREIBERS (2012, S. 35ff) Beschreibung der Modellkomponenten orientiert, allerdings wurden auch weitere Quellen miteinbezogen (vgl. kommender Abschnitt).

Für die Itemformulierungen wurde sich an üblichen Richtlinien orientiert (PORST, 2000; BÜHNER, 2006, S. 46ff). Allein auf Grund der Fülle an Konstrukten, die das 3L-Modell für die Qualität von Praktika definiert, ist die Konstruktion eines vollständig das 3L-Modell abbildenden Instruments nicht möglich, ohne Bearbeitungszeiten jenseits üblicher Dauer für Evaluationserhebungen (15min) zur Konsequenz zu haben. Dieses würde dem Anspruch der Durchführungsökonomie widersprechen, Akzeptanzprobleme auslösen und gerade gegen Ende des Fragebogens zu Ermüdungseffekten der Teilnehmenden führen, die verzerrte Ergebnisse liefern würden.

Die Operationalisierung wurde daher in zwei Fragebögen aufgeteilt. Dies geschah im ersten Entwurf mit insgesamt 140 Items in 40 Skalen und mündete in zwei Instrumente PraQ-A und PraQ-B. Teil A untersucht den *Lernzuwachs*, Teil B *Lehrkompetenz* und *Lernumgebung*. Im Folgenden wird je Konstrukt die Itemzuordnung unter Bezugnahme auf deren Bezugsquellen vorgenommen. Konkrete Beispielitems sind darin ebenso zu finden.

10.3.3. Itemzuordnung je Konstrukt

Da das 3L-Modell eine Neuentwicklung darstellt und etwa die Hälfte der Items einer Neukonstruktion bedarf, stellt die nun vorzustellende Itemzuordnung je Konstrukt nur eine unsichere Zuordnung dar. Zur Absicherung dieser Unsicherheit wird später in Form explorativer Faktorenanalysen die empirische Konstruktzugehörigkeit der stark veränderten oder neu konstruierten Items ermittelt (vgl. Abschnitt 12.4). Die Beispielimts wurden im Verlauf der Testevaluationsstudien teils sprachlich noch angepasst. Die hier dargestellten Formulierungen stellen die erste Version dar, eine finale Version ist in Anhang A auf Seite 269 zu finden. Die Beschreibung der Konstrukte wird an dieser Stelle nicht nochmals aufgeführt, da sie im Kapitel 7 bereits ausführlich beschrieben wurden.

10.3.3.1. Dimension Lernzuwachs

1.1: Fachwissen: Die Basis der Items des Fachwissens war die Skala *Fachkompetenz* nach BRAUN & GUSY (2006). Für den Umfang des Konstrukts wurde sich an LAGOWSKIS »domain knowledge« (LAGOWSKI, 2002), STAUFENBIEHLS »FEPPRA« (STAUFENBIEL, 2000, Anhang) und dem Effektivitätsmodell von PSILLOS & NIEDERER (2002, S. 23) orientiert. Itemideen stammten z. B. von FRASER & MCROBBIE (1995, S. 315, Item 28). Das erste der folgenden Beispielimts ist daran orientiert, das zweite Item ist orientiert an BRAUN & GUSY (2006):

Beispielimt 1: Aufgrund dieses Praktikums verstehe ich die abgedeckte Theorie besser.

Beispielimt 2: Aufgrund dieses Praktikums kann ich behandelte, komplizierte Sachverhalte anschaulich darstellen.

1.2: Experimentelle Kompetenz: Bisherige Operationalisierungen zu Selbsteinschätzungen experimenteller Fähigkeiten sind entweder zu stark auf Physik spezifiziert (z. B. »Abhängige und unabhängige Variable«, VAN DER JAGT et al., 2012) oder werden in der Schule eingesetzt, was nur eine bedingte Übertragbarkeit auf den universitären Kontext gewährt (SCHREIBER et al., 2016). SCHREIBERS (2016) Instrument zur Selbsteinschätzung *experimenteller Kompetenz* ist daher nur bedingt übertragbar, da es sich um eine situative, hoch kontextualisierte Selbstevaluation des gerade erlebten Experimentierverhaltens während einer spezifischen Aufgabe, im Schul- und nicht Universitätskontext handelt. Die Indikatoren sind dadurch so stark auf die vorherige Handlung bezogen und auch nur durch diese nachvollziehbar⁶, dass sie für PraQ nicht geeignet sind.

⁶z. B. »Ich konnte die Messungen ohne Probleme durchführen«, ebd., S. 7

Für die konkrete Operationalisierung der Konstrukte wurde sich daher wie beschrieben an Arbeiten von SCHREIBER (2012) für die verschiedenen Skalen der Experimentellen Kompetenz orientiert, mit der Notwendigkeit einer Neukonstruktion zugehöriger Items. MAISEYENKA (2014, S. 16, 22) lieferte Anregung für Beispiellitem 1 in der Skala *Erwartungen formulieren*, CHANG et al. (2011, S. 1229) gaben Anregung für die Operationalisierung in der Skala *Versuchsplan entwerfen* mit dem zweiten Beispiellitem:

Beispiellitem 1 (*Erwartungen formulieren*): Aufgrund dieses Praktikums sind meine Erwartungen an die Ergebnisse eines Experiments konkreter geworden.

Beispiellitem 2 (*Versuchsplan entwerfen*): Aufgrund dieses Praktikums kann ich einen Versuch besser planen.

1.3: Kommunikationskompetenz: Für die Operationalisierung der Konstrukte wurde sich am Kompetenzbereich *Kommunikation* des Kompetenzmodells für Praktika von KREITEN (2012, S. 197) orientiert, etwa beim Beispiellitem 1 (vgl. auch ZWICKL et al., 2013, S. 8–10: »argumentation«). LAGOWSKI (2002, S. 4, 9) und BRAUN et al. (2008) bildeten die Orientierung für Indikatoren wie *Wortbeiträge verständlicher formulieren* oder *Protokolle verständlicher schreiben*. Für die Formulierung von Items zur *schriftlichen Kommunikation* bildeten CHANG et al. (2011, S. 1219) die Grundlage. Die AAPPT (1998, S. 484) lieferte die Formulierung für Beispiellitem 2.

Beispiellitem 1 (*mündliche Kommunikation*): Aufgrund dieses Praktikums kann ich auf Fragen fachsprachlich besser eingehen.

Beispiellitem 2 (*schriftliche Kommunikation*): Aufgrund dieses Praktikums kann ich meine Versuchsprotokolle präziser schreiben.

1.4: Bewertungskompetenz: Für die Operationalisierung in Form von Items wurde sich an Arbeiten von KREITEN (2012, Item B3, Beispiellitem 1), LAGOWSKIS »higher order cognition« (LAGOWSKI, 2002, S. 4), KMK (2005) im Vergleich mit MAISEYENKA (2014, S. 16) und HALLER (1999, Anhang, Item C4, Beispiellitem 2) orientiert.

Beispiellitem 1: Aufgrund dieses Praktikums kann ich Verbesserungsvorschläge zu Experimenten besser begründen.

Beispiellitem 2: Aufgrund dieses Praktikums kann ich Ergebnisse besser kritisch diskutieren.

1.5: Kooperationskompetenz: Für das Konstrukt *Kooperationskompetenz* wurde eine vollständige Adaption der gleichnamigen Skala von BRAUN et al. (2008) vorgenommen. Obwohl auch im praktikumsbezogenen SLEI-Fragebogen (FRASER & McROBBIE, 1995) eine Skala zu Aspekten der *Kooperation* enthalten ist, wurde sich dagegen entschieden, diese zu adaptieren. Dies liegt einerseits darin begründet, dass sich eher die affektive Komponente, der *Gruppenzusammenhalt* bei FRASER & McROBBIE dahinter verbirgt. Andererseits ist die Skala im Gegensatz zur Skala bei BRAUN et al. nicht kompetenzorientiert entwickelt worden. Bei der Adaption der Skala nach BRAUN et al. wurden lediglich praktikumsnähere Formulierungen wie »Praktikumsgruppe« (statt »Arbeitsgruppe«, Beispielitem 1) und »Ergebnisse« (statt »Ergebnis«, Beispielitem 2) bei einigen Items eingefügt.

Beispielitem 1: Durch die Teilnahme an meiner Praktikumsgruppe gelingt es mir leichter, mich an der Aufgabenverteilung innerhalb einer Gruppe zu beteiligen.

Beispielitem 2: Mit den Ergebnissen meiner Praktikumsgruppe kann ich mich voll und ganz identifizieren.

1.6: Personalkompetenz: Die konkrete Operationalisierung der beiden Konstrukte erfolgte für das *Interesse* erneut über eine Adaption des Instruments nach BRAUN et al. (2008), Skala *Personalkompetenz*. Für das *Zeitmanagement* wurde die Skala nach WILD & SCHIEFELE (1994) für Praktika adaptiert.

Beispielitem 1 (*Interesse*): In diesem Praktikum habe ich Sachen gelernt, die mich begeistern.

Beispielitem 2 (*Zeitmanagement*): Aufgrund dieses Praktikums kann ich meine experimentelle Arbeit besser zeitlich organisieren.

10.3.3.2. Dimension Lehrkompetenz

2.1: Verständlich erklären: Für das Konstrukt *verständlich erklären* wurde eine fast vollständige Übernahme der Skala nach THIEL et al. (2012) vorgenommen, ergänzt um Items zum Erklären von Experimenten. Lediglich für das letzte Item der Skala (Beispielitem 1) wurde der Begriff »Studien« entfernt, da diese meist nicht Bestandteil von Praktika sind.

Beispielitem 1: Der/die Betreuende erläutert auch komplexe Experimente so, dass das methodische Vorgehen verständlich wird.

2.2: Zusammenfassen & Wiederholen: Hier wurde eine vollständige Adaption der entsprechenden LeKo-Skala (THIEL et al., 2012) vorgenommen. Ergänzt wurde lediglich die praktikumsspezifische Gegebenheit, dass oft microteaching in Form von Gesprächen mit dem/der Betreuenden stattfindet.

Beispielitem 1: Der/die Betreuende wiederholt und vertieft besonders schwierige Aspekte ausreichend (z. B. in Gesprächen).

2.3: Verknüpfungen herstellen: Die Items dieser Skala sind vollständig von THIEL et al. (2012) übernommen.

Beispielitem 1: Der/die Betreuende stellt immer wieder Bezüge zu dem bereits behandelten Stoff her.

2.4: Verständnis überprüfen: Bei den Items für diese Skala wurden wieder Anpassungen und Neukonstruktionen vorgenommen. Neu konstruiert wurden Items, die auch das Prüfen des experimentellen Verständnisses beinhalten (Beispielitem 1). Bezüglich der Formulierung wurde analog zu den Items zu inhaltlichen Aspekten nach THIEL et al. (2012) vorgegangen. Bei einem Item wurde die ursprüngliche Formulierung »bevor er/sie im Stoff weitergeht« durch die universell auch für den Fortgang des Experimentierens passende Formulierung »bevor es weitergeht« ersetzt (Beispielitem 2):

Beispielitem 1: Der/die Betreuende vergewissert sich, dass die Studierenden zentrale Aspekte des Experiments verstanden haben, bevor sie weitermachen.

Beispielitem 2: Der/die Betreuende vergewissert sich, dass die Studierenden zentrale inhaltliche Aspekte verstanden haben, bevor es weitergeht.

2.5: Lerndienliches Feedback geben: Feedback geschieht immer in Bezug auf eine Handlung oder ein Produkt, daher sollten die Items dieser Skala praktikumsspezifisch formuliert sein. Erreicht wurde dies über eine Adaption der Skala nach THIEL et al. (2012) unter Hinzufügen praktikumsrelevanter Tätigkeiten wie Experimentieren (Beispielitem 1) und Produkten wie dem Ergebnis der Vortestats (Beispielitem 2).

Beispielitem 1: Der/die Betreuende gibt konstruktives Feedback zum Experimentieren der Studierenden.

Beispielitem 2: Der/die Betreuende gibt aussagekräftiges Feedback zum Vortestat der Studierenden.

2.6: Gute Lernatmosphäre herstellen: Die Lernatmosphäre stellt ein übergreifendes Konstrukt dar und konnte daher in seinen Items komplett vom LeKo übernommen werden.

Beispielitem 1: Der/die Betreuende geht auf die Interessen der Studierenden ein.

2.7: Zur Selbsttätigkeit anregen: Die Items dieses Konstrukts waren erneut so zu übernehmen.

Beispielitem 1: Der/die Betreuende regt die Studierenden an, sich mit den Lehrinhalten auch außerhalb der Veranstaltung auseinanderzusetzen.

2.8: Relevanz verdeutlichen: Auch die Skala *Relevanz verdeutlichen* konnte zu einem Großteil aus dem LeKo mit Items ausgestattet werden. Es wurde lediglich ein Item von HENSE (2012, S. 86) adaptiert hinzugefügt, dass der Diversität der Inhalte in Praktika Rechnung trägt (Beispielitem 1). Zudem wurde das Item zur Nützlichkeit für die Berufspraxis vereinfacht, da dies stärker in Fortgeschrittenen-Praktika mit Forschungsbezug vorkommt (Beispielitem 2).

Beispielitem 1: Der/die Betreuende zeigt, welchen Nutzen die einzelnen Inhalte haben.

Beispielitem 2: Der/die Betreuende zeigt in Ansätzen die Nützlichkeit des Stoffs für die spätere Berufspraxis auf.

2.9: Ermutigen und Selbstwirksamkeit stärken: Diese Skala wurde auf Grund ihres übergreifenden Charakters aus THIEL et al. (2012) so beibehalten.

Beispielitem 1: Der/die Betreuende ermutigt die Studierenden bei der Aneignung schwieriger Inhalte.

2.10: Effizient mit Störungen umgehen: Diese Skala wurde ebenso auf Grund ihres übergreifenden Charakters aus THIEL et al. (2012) so beibehalten.

Beispielitem 1: Der/die Betreuende beweist bei Unruhe und Störungen das nötige Durchsetzungsvermögen.

2.11: Diagnostik vornehmen: Diese Items wurden operationalisiert nach der Skala der *Diagnostik des Lernfortschritts* nach DITTON & MERZ (1995) und mussten von der Schul- in die Hochschulformulierung überführt werden (Beispielitem 1). Zudem wurden erneut analoge experimentierspezifische Items ergänzt (Beispielitem 2):

Beispielitem 1: Der/die Betreuende nimmt sich die Zeit, um einzelnen Studierenden Dinge zu erklären, die sie nicht verstanden haben.

Beispielitem 2: Der/die Betreuende erklärt das Experiment so oft, bis es alle verstanden haben (z. B. in Gesprächen; im Vorgespräch; Nachbesprechungen etc.).

10.3.3.3. Dimension Lernumgebung

3.1: Skriptqualität: Die Skala zur Skriptqualität musste wie beschrieben gänzlich neu konstruiert werden. Orientiert wurde sich zunächst an den Hauptkonstrukten des LeKo (Beispielitem 1), da das Skript in der Vorbereitung und Durchführung des Praktikums eine stark steuernde Rolle erhält (TIBERGHIE et al., 1998, vgl. Abschnitt 5). Hinzu kamen Items, die orientiert an Qualitätskriterien für Schulbücher nach ABEL (2014, Beispielitem 2) waren. Spätere Analysen (REHFELDT et al., 2015a) zeigten, dass auch die Anforderungen an gute Lehrbücher nach MERZYN (1995) erfasst werden (Studie 1.2).

Beispielitem 1: Das Skript enthält angemessene Modelle, Graphiken oder Schemata, die das Verständnis komplexer Sachverhalte erleichtern.

Beispielitem 2: Das Skript ermöglicht selbstständiges Lernen (Vorbereitung, Nachbereitung etc.).

3.2: Skriptqualität für den Versuch: Aus den Ergebnissen der Evaluation des physikalischen NP-Praktikums der Freien Universität Berlin (REHFELDT et al., 2015; MÜHLENBRUCH et al., 2015) stammt die Skala *Skriptqualität für den Versuch*, die im Gegensatz zur *allgemeinen Skriptqualität* auf die experimentelle Vorbereitung abzielt. Ein Zusammenschluss dieser Skala mit der Skala *Skriptqualität* ist natürlich denkbar und wird statistisch-strukturell untersucht werden (vgl. Abschnitt 12.4).

Beispielitem 1: Das Skript unterstützte mich dabei, eine gute Vorstellung vom Versuchsaufbau zu erhalten.

3.3: Räumlich-sachliche Ausstattung: Die Räumlich-sachliche Ausstattung wurde über praktikumsrelevante Einzelitems operationalisiert.

Beispielitem 1: Die Lernmaterialien (z. B. Skript, Platzskripte, ggf. PC-Plätze) waren leicht verfügbar.

3.4: Integration der Vorlesung: Gemessen wird dieses Konstrukt über eine Adaption der Skala von FRASER & McROBBIE (1995) unter Übersetzung ins Deutsche und unter Entfernen zu stark inhaltsgleicher Items.⁷

Beispielitem 1: Die Themen der Vorlesung passen zur Arbeit im Praktikum.

Mit dem so erstellten Itempool konnten die Fragebögen PraQ-A und PraQ-B in einer ersten Version erstellt werden.

Es folgen Ausführungen zur Gestaltung derselben, danach werden Objektivitäts- und Reliabilitätsmaßnahmen und die Testevaluationsstudien zur Optimierung der Interpretierbarkeit der Ergebnisse von PraQ-Messungen vorgestellt.

10.4. Vorbereitung der Durchführung

Um die Verständlichkeit der PraQ-Items für Studierende zu erhöhen, wurden zwei kognitive Interviews mit studentischen Hilfskräften der AG Didaktik der Physik der Freien Universität Berlin geführt. Diese befanden sich im Bachelor ihres Physik-(Lehramts)-Studiums und gehörten damit zur Zielgruppe des PraQ.

Die sprachlichen Anpassungen fielen minimal aus, lediglich die teils verwirrenden unterschiedlichen Begriffe aus den LeKo- und BEvaKomp-Items wurden vereinheitlicht: Aus den Begriffen »Konzepte« und »Inhalte« wurde der gemeinsame Begriff »Sachverhalte«. Daraufhin wurden die zwei Fragebögen zusammengestellt. Diese enthielten außer den 3L-Modell-relevanten Items noch folgende Items:

Gruppencode: Der Gruppencode ermöglicht die anonyme Zuordnung der Studierenden zu Experimentiergruppen. Dies kann in der Auswertung für hierarchische Modelle nützlich sein, wenn Studierende einer Gruppe je denselben/dieselbe Praktikumsbetreuer*in bewerten.

Personencode: Üblicher Code zur anonymen Identifikation Studierender, die zufälligerweise doppelt an PraQ-Erhebungen teilnahmen.

Mindestbesuch: Gemäß RICHARDSON (2005, S. 403) sollten Evaluationserhebungen vorzugsweise zu Ende der ganzen Lehrveranstaltung erfolgen. Für PraQ-Erhebungen wurde das Mindestkriterium von 40% absolvierter Praktikumszeit gesetzt und mit diesem Item danach gefragt. Dies wurde als annehmbare Erfahrung mit der Lehrveranstaltung gewertet, um Auskunft über dessen Qualität geben zu können.

⁷z. B. »My laboratory work and regular science class work are unrelated.« bzw. »The laboratory work is unrelated to the topics that I am studying in my science class.« (FRASER & McROBBIE, 1995, S. 314ff)

Blockveranstaltung: Manche Praktika sind als Blockveranstaltung realisiert, meist in der vorlesungsfreien Zeit. Um mögliche Einflüsse dieser Organisationsform zu untersuchen, wurde dichotom danach gefragt, ob es sich um eine Blockveranstaltung handelt.

Partner- oder Gruppenarbeit: Zur genaueren Beschreibung der teilnehmenden Praktika wurde danach gefragt, ob vorzugsweise Partner- und Gruppenarbeit oder die Arbeit allein praktiziert wird. Ein zusätzliches Item erfasste, ob die Gruppenmitglieder einem Wechsel unterliegen.

Baugleiche Experimente: Interessant in Bezug auf Unterschiede zwischen Physik- und Chemiepraktika ist, wie die Experimente organisiert sind. In der Physik sind sie oft baugleich, in der Chemie traditionell nicht.

Universität: Item zur Erfassung der Universität, an der das Praktikum stattfindet.

Vorwissen: Mit je einem Item zum *theoretischen* und *praktischen Vorwissen* wird als Kontrollvariable erhoben, ob dies Einflüsse auf das Evaluationsverhalten hat.

Geschlecht: Das Geschlecht wurde primär zur Beschreibung der Stichprobe erhoben.

Alter: Das Alter wurde primär zur Beschreibung der Stichprobe erhoben.

Studiengang: Der Studiengang der Studierenden wurde zur Beschreibung der Stichprobe erhoben, aber auch für die Möglichkeit zur Auswertung nach fachfremd und nicht-fachfremd das Praktikum besuchenden Studierenden.

Fachrichtung des Praktikums: Diese wichtige, das Praktikum kennzeichnende, Eigenschaft wird erhoben.

Fachsemester: Das Fachsemester wurde primär zur Beschreibung der Stichprobe erhoben.

Bemerkungen: Jeder Fragebogen enthielt ein offenes Kommentarfeld. Dieses wurde in der Auswertung bei gravierend auf die Auswertung auswirkenden Kommentaren (etwa ein invertiertes Antwortverhalten) berücksichtigt, ansonsten gebündelt an Praktikumsleitende weitergegeben.

Kapitel 11

Objektivität

Zur Sicherung der Durchführungsobjektivität (BORTZ & DÖRING, 2006, S. 212) wurden bei der Konstruktion der PraQ-Fragebögen einige Maßnahmen ergriffen.

Zunächst stehen alle für das Ausfüllen notwendigen Informationen im Fragebogen selbst, es sollte keine offensichtlichen Gründe für notwendige Nachfragen bei den Testleitenden geben.

Die erste Seite des Fragebogens wurde als Titelseite ohne Items realisiert, um nicht von außen einsehbar zu sein, was dem Datenschutz und der Verringerung von Effekten der sozialen Erwünschtheit dient. Für Erhebungen mit dem PraQ-B (*Lehrkompetenz* und *Lernumgebung*) wurde antizipiert und im Durchführungsmanual festgehalten, dass die Betreuenden möglichst den Raum verlassen oder abgelenkt sein sollten. Dies stellt eine Vorsichtsmaßnahme dar, obwohl lediglich der/die Betreuende des letzten, nicht des aktuellen Praktikumstermins evaluiert wird. Es sind also eigentlich keine Effekte der sozialen Erwünschtheit auf anwesende Personen zu erwarten.

Die Erhebungen selbst wurden innerhalb der Praktikumsräume vorgesehen, also einer zum Testgegenstand konvergenten Umgebung. Für Erhebungen im Berlin-Brandenburger Raum und in Wuppertal war der Autor vor Ort, andere Praktika erhielten die erforderlichen Materialien per Post, darunter eine ausreichende Anzahl Fragebögen, Süßigkeiten als Durchführungsanreize (incentives) und Durchführungsmanuale für die Praktikumsleitenden und -betreuenden. Zusätzlich erhielten sie telefonisch die notwendigen Informationen, auch wenn diese bereits im Manual enthalten waren. Kontaktperson war je der/die Praktikumsleitende. Zudem waren Zettel mit randomisierten Gruppencodes anbei und enthielten die Aufforderung, den Teilnehmer*innen einer Praktikums-Experimentalgruppe (in der Regel ca. acht Personen) denselben Gruppencode zukommen zu lassen (vgl. vorheriger Abschnitt). Der Gruppencode wurde in Chemiepraktika nicht eingesetzt, wo Praktikums-Experimentalgruppen nicht vorgesehen waren.

Bei Nachfragen zu Itemformulierungen sollten die erhebenden Betreuenden derart reagieren, dass sie die Antwort »Bitte beantworten Sie es so, wie sie es verstehen« nutzen. Dies sollte eine hohe Durchführungsobjektivität gewährleisten.

11. Objektivität

Die Datenaufbereitung und -auswertung wird in den einzelnen Studien genauer beschrieben und zielt auf Transparenz ab (vgl. Kapitel 12 und 13). Der interessierte Leser kann zudem auf Anfrage beim Autor die vollständige kommentierte Syntax der Bearbeitung der Datensätze, von der ersten Datenbereinigung bis zur Interpretation des Ergebnisses, erhalten. Dies soll die Prüfung der Auswertungs- und Interpretationsobjektivität (BORTZ & DÖRING, 2006, S. 212) ermöglichen.

Kapitel 12

Testevaluation (Validität und Reliabilität)

12.1. Forschungsfragen zur Validierung

Die Erfassung der Praktikumsqualität kann verschiedenen Zwecken dienen und somit der Begriff der Qualität auch unterschiedlich definiert werden. Einerseits kann diese Grundlage für genauere Analysen und Lehrverbesserungen sein, also als Rückmeldung für die Beteiligten über die Lehr- und Lernqualität dienen (vgl. Ziel 1, Kapitel 8 auf Seite 99). Andererseits soll das Instrument psychometrisch adäquate Messungen bzw. Interpretationen praktikumsrelevanter Konstrukte ermöglichen, welche dann im (fachdidaktischen) Forschungskontext wertvoll sind (Ziel 2), was letztlich auch dem Ziel 1 zu Gute kommt. Gemäß argument-based approach der Validierung (AERA et al., 2014; KANE, 2001; MESSICK, 1995) wird an dieser Stelle festgelegt, welche Schlussfolgerungen aus den Ergebnissen von Erhebungen mit dem PraQ-Instrument gezogen können werden sollen, um darauf ausgerichtet Forschungsfragen für die Testevaluation zu formulieren. Es werden die klassisch geprägten Begriffe (Inhaltsvalidität, Konstruktvalidität, etc.) zur Orientierung mit genannt, obwohl diese Validität als eine Eigenschaft des Tests kennzeichnen und nicht als eine Eigenschaft der Interpretation von Testergebnissen im Rahmen spezifischer Kontexte und Zielgruppen:

Validation logically begins with an explicit statement of the proposed interpretation of test scores, along with a rationale for the relevance of the interpretation to the proposed use.

AERA et al. (2014, S. 11)

Erste Anforderung an die Interpretierbarkeit der Testergebnisse ist, dass die vollständige lernrelevante Praktikumsqualität in möglichst allen Facetten und ohne qualitätsirrelevante Aspekte bestimmt werden kann, die einem *Lernzuwachs* der Studierenden dienlich ist (»Inhaltsvalidität«). »Construct irrelevant variance« (MESSICK, 1995, S. 742) meint hier den Fall, dass konstrukt fremde Einflüsse oder konstrukt fremde Variablen mit gemessen werden, was es auszuschließen gilt.

Ebenso zu vermeiden ist die »construct underrepresentation« (ebd.). Gemeint ist hiermit das »Vergessen« von wesentlichen Aspekten der Qualität von Praktika, etwa

12. Testevaluation (Validität und Reliabilität)

das Nicht-Erfassen von experimentellen Fähigkeiten, obwohl es zentrales Ziel von Praktika ist (vgl. AAPT, 1998, HALLER, 1999, ZWICKL et al., 2013).

Die sich hieraus ergebenden Forschungsfragen für die Validierung sind:

Forschungsfrage 1: Welche Indikatoren des PraQ sind nicht konstruktrelevant für Praktikumsqualität? (»Inhaltsvalidität«)

Forschungsfrage 2: Welche Aspekte von Praktikumsqualität fehlen im PraQ? (»Inhaltsvalidität«)

Des Weiteren sollen Ergebnisse der Messungen verschiedener Konstrukte der Praktikumsqualität voneinander trennbar interpretiert werden dürfen (»faktorielle Validität«). Damit könnte nämlich eine etwaige Verbesserung des Praktikums an just denjenigen Konstrukten ansetzen, die sich als problematisch ergeben haben. Dazu gehört auch eine klare Item-Konstrukt-Zuordnung (AERA et al., 2014, S. 16).

Dies kann dann in einem zweiten Schritt zu gezielteren Analysen oder Interventionen führen. Es soll also in der Interpretation der Testergebnisse möglich sein, die Qualität einzelner Dimensionen bzw. qualitätsrelevanter Konstrukte voneinander getrennt untersuchen zu können. Theoretisch wie statistisch bedeutet dies eine Trennbarkeit der Konstrukte (»structural aspect«, MESSICK, 1995).

Forschungsfrage 3: Welche Konstrukte lassen sich beim PraQ theoretisch wie empirisch voneinander getrennt interpretieren? (»Faktorielle Validität«)

Für jedes neue Messinstrument liegt ein Vergleich mit bereits etablierten Instrumenten nahe (»Konvergente Konstruktvalidität« bzw. »external aspect«, MESSICK, 1995). Zudem sollte die subjektiv geprägte Erhebungsform »Selbsteinschätzungen von teilnehmenden Studierenden« kritisch auf Störeinflüsse geprüft werden (»diskriminante Konstruktvalidität«; vgl. Abschnitt 10.1).

So ist etwa nicht bekannt, wie stark Messungen mittels PraQ Kompetenzselbsteinschätzungen mit tatsächlicher Kompetenz zusammenhängen, oder ob eher Bestandteile von *Selbstwirksamkeitserwartung* gemessen werden (CRAMER, 2010). Wenn die Messung mittels Selbsteinschätzung als tatsächlicher Kompetenzzuwachs interpretierbar sein soll, so müsste sich demnach empirische Evidenz finden lassen, die Zusammenhänge zwischen dem PraQ und etablierten Kompetenzmessinstrumenten je desselben Konstrukts aufzeigt. Analoges gilt, falls eine Interpretation gemäß *Selbstwirksamkeitserwartung* gelten können soll.

Forschungsfrage 4: Wie hängen die Skalen des PraQ statistisch mit Kompetenz-Leistungstests und Selbstwirksamkeitserwartung zusammen? (»Konvergente Konstruktvalidität«)

Zur Untersuchung subjektiver, affektiver Einflüsse auf das Testverhalten von Praktikumsteilnehmenden kann zudem der Zusammenhang der Konstruktausprägungen mit dem »Praktikumssozialklima« herangezogen werden. Dieses Konstrukt beschreibt die soziale Atmosphäre im Praktikum und stellt einen möglichen Störeinfluss für eine valide Interpretation der Messung von Aspekten der Praktikumsqualität dar. Studierende könnten beispielsweise aus sozialer Frustration das Skript schlechter beurteilen, als es rein kognitiv von ihnen beurteilt würde. Ähnlich also wie der konvergente Zusammenhang zu einem Leistungstest sollte sich ein diskriminanter Zusammenhang zum Praktikums-klima ergeben.

Forschungsfrage 5: Wie hängen die Skalen des PraQ statistisch mit der Messung von allgemeinem Praktikums(sozial)klima zusammen? (»diskriminante Konstruktvalidität«)

Schließlich soll die Evaluation von Praktika langfristig allein aus Sicht der Studierenden erhoben werden, wie es institutionell verankert ist (vgl. Abschnitt 4.2). Um eine zu einseitige Interpretation der Messergebnisse zu vermeiden, kann es in dieser Arbeit daher lohnen, auch Urteile von dritter Stelle mit denen der Studierenden zu vergleichen (JUSSIM, 2012, S. 202).

Forschungsfrage 6: In welchen Facetten der Praktikumsqualität unterscheiden sich Einschätzungen von Studierenden gegenüber Einschätzungen von Expert*innen? (»Multiperspektivität«)

Darüber hinaus wurden wie bereits erwähnt kognitive Interviews mit Studierenden geführt, um das homogene Verständnis der Items zu prüfen (vgl. Abschnitt 10.4). Dies kann als ein Beitrag zur Testevaluation der kognitiven Prozesse der Studierenden bei der Testabsolvierung verstanden werden (»Kognitive Validität«; AERA et al., 2014, S. 15). Häufig werden auch kriteriale Maße wie Noten herangezogen, um sie mit Testergebnissen zu vergleichen (ebd.). Dies ist hier nur bedingt sinnvoll, da Noten meist auf Grund einer sozialen oder kriterialen Bezugsnorm vergeben werden, die *Lernzuwachss*skalen des PraQ aber die individuelle Bezugsnorm ansprechen, durch die Ausrichtung als Veränderungsmaß (»Aufgrund dieses Praktikums. . . kann ich besser. . . «).

Die inhaltlichen Forschungsfragen 7–9 werden sich an die Validierungs-Forschungsfragen 1–6 anschließen (für einen Überblick über alle Forschungsfragen vgl. Kapitel 8 auf Seite 99).

12.2. Validierungsdesign

Diese Forschungsfragen wurden im Rahmen einer Testevaluations-, klassisch »Validierungsstudie« genannt, untersucht. Die erste Studie bezüglich der Forschungsfragen Forschungsfrage 1 und Forschungsfrage 2 wurde über Expert*inneninterviews zur Relevanz, Passung und Vollständigkeit der Indikatoren des PraQ realisiert, wie es bei am inhaltlichen Aspekt orientierten Testevaluationsstudien üblich ist (AERA et al., 2014).

Zusätzlich wurden speziell die Skalen zur *Skriptqualität* im Rahmen eines Expert*innenratings bewertet (REHFELDT et al., 2015a). Die Forschungsfrage Forschungsfrage 3 zur Indikatoren-Konstrukt-Struktur (»faktorielle Validität«) wurde im Rahmen einer zweistufigen Studie (Studien 2.1 und 2.2) aller PraQ-Indikatoren und Konstrukte bei einer großen Stichprobe über Erhebungen im deutschsprachigen Raum untersucht. Die Studie 2.1 diente der Exploration der Strukturierung der Items zu Faktoren. Die Studie 2.2 untersuchte die explorativ gefundene Item-Faktoren-Struktur konfirmativ an einer neuen Stichprobe. Am Ende dieses Zweischritts konnten verlässliche Reliabilitätsschätzer berechnet werden.

Die Forschungsfragen 4 und 5 wurden im Rahmen einer klassischen Kreuzvalidierungsstudie erforscht. Hierbei wird ein Vergleich von *Experimentelle-Kompetenz-PraQ*-Skalen und *Erkenntnisgewinnungs-Kompetenz-Leistungstests* realisiert. Ein weiterer Vergleich mit *experimentierbezogener Selbstwirksamkeitserwartungen* wurde vorgenommen.

Darüber hinaus wurde diskriminant ein Vergleich aller PraQ-Skalen mit dem *Praktikumssozialklima* vorgenommen. Für die Erforschung von Forschungsfrage 6 wurden ein Expert*innenrating und ein Studierendenrating zur *Skriptqualität* durchgeführt und verglichen. Diese Studie gliederte sich an die Erhebungen der Studie nach REHFELDT et al. (2015a) an.

Tab. 12.1 fasst alle Studien respektive deren adressierte Forschungsfragen und Konstrukte zusammen.

Tabelle 12.1.: Forschungsfragen und Konstrukte für die Testevaluationsstudien.

Studie	FF	»Validität«	Konstrukte
1.1	1	Inhalt	alle aus PraQ
1.2	2	Inhalt	Skriptqualität
2.1	3	faktoriell	alle außer etablierter Skalen
2.2	3	faktoriell	alle aus PraQ
3.1	4	konvergent	Experimentelle Kompetenz
3.2	5	diskriminant	alle aus PraQ; Praktikumsclima
4	6	multiperspektivisch	Skriptqualität

12.3. Studie 1: Inhaltlicher Umfang des PraQ («Inhaltsvalidität»)

Innerhalb der Studie 1 werden die Forschungsfragen Forschungsfrage 1 und 2 mit Expert*inneninterviews untersucht. In Studie 1.1 wurden hierfür Interviews zur Relevanz und Vollständigkeit des Itempools mit ExpertInnen für naturwissenschaftliche Praktika durchgeführt, die alle Konstrukte des PraQ-A und PraQ-B zum Gegenstand hatten. Auf den inhaltlichen Aspekt abzielend fanden hierbei zwei Schritte statt.

Den ersten Schritt bildete eine umfassende Literatursichtung zu Qualitätsaspekten von Hochschullehre und Praktika (fordern u. a. AERA et al., 2014, S. 12), kombiniert mit kollegialen Diskussionen zu den relevanten Facetten und Indikatoren hochqualitativer Praktika (vgl. Abschnitt 10.3). Dies sollte sicherstellen, dass eine umfassende, gesättigte Menge an Indikatoren zur Verfügung steht, die dann in den zweiten Schritt getragen werden kann.

Diesen zweiten Schritt bildete eine Akzeptanzstudie mit kognitiven Interviews (PRÜFER & REXROTH, 2005) mit Praktikumsexpert*innen aus den verschiedenen Praktika der Naturwissenschaften. Dies diente einerseits der Exploration der Passung des Instruments zur Lehrrealität, andererseits der Identifikation möglicher Akzeptanzprobleme (REHFELDT et al., 2014). Beides führt ggf. zu einer Reduktion der Itemanzahl. Dies war mit ein Grund für die vergleichsweise große Anzahl von ursprünglich 140 Items des Itempools.¹

Für Studie 1.2 sollte darüber hinaus noch Evidenz speziell zum Konstrukt *Skriptqualität* gewonnen werden, da Relevanz-Beurteilungen der Items von Expert*innen für die lerngerechte Gestaltung von Medien (hier: Physikdidaktiker*innen) verfügbar waren. Wie aus dem Theorieteil hervorging (vgl. Kapitel 5 auf Seite 55) wird dem Skript eine hohe Bedeutung für den *Lernzuwachs* der Studierenden attestiert,

¹Die erste Fassung des LeKo bestand aus 63 Items (THIEL et al., 2012), der BEvaKomp bestand in 2008 aus 26 Items (BRAUN et al., 2008).

weshalb gerade dort eine gültige Interpretation der Skriptqualität wichtig für die Evaluation des Praktikums erscheint. Die Studie 1.2 untersucht daher Forschungsfrage 1 und Forschungsfrage 2 speziell für die PraQ-Skalen *Skriptqualität* und *Skriptqualität für den Versuch*.

12.3.1. Studie 1.1: Relevanz und Vollständigkeit: PraQ

12.3.1.1. Einleitung

Anspruch der PraQ-Messungen ist es, ein umfassendes und repräsentatives Bild der Qualität von Praktika in ihrer Interpretation zu liefern. An diesem Anspruch wurde sich in der theorie- und forschungsgeleiteten Konstruktion des Itempools (vgl. Abschnitt 10.3) orientiert. Dies liefert theoretische Evidenz (MESSICK, 1995) für die Adäquatheit der Inhalte des PraQ in Bezug auf die intendierten Konstrukte im 3L-Modell. Ziel der nun folgenden Studie ist es, auch empirische Evidenz dafür zu finden (»inhaltlicher Validitätsaspekt«, MESSICK, 1995, S. 745). Um die Repräsentativität von Items bezüglich ihres zugeordneten Konstrukts zu erfassen, wird oft mit Expert*innenbeurteilungen gearbeitet (AERA et al., 2014, S. 14).

Forschungsfragen Gemäß Validierungsdesign werden folgende Forschungsfragen gestellt:

Forschungsfrage 1: Welche Indikatoren des PraQ sind nicht konstruktrelevant für Praktikumsqualität? (»Inhaltsvalidität«)

Forschungsfrage 2: Welche Aspekte von Praktikumsqualität fehlen im PraQ? (»Inhaltsvalidität«)

Die Studie zielt auf die fächerübergreifende Schnittmenge der relevanten Aspekte für Praktikumsqualität ab. Der fächerübergreifende Aspekt kommt aus der Zielsetzung eines gemeinsam für die naturwissenschaftlichen Praktika einsetzbaren Instruments.

12.3.1.2. Methode

Stichprobe Die Auswahl der Expert*innen geschah über eine Anfrage an diverse Praktikumsexpert*innen der Freien Universität Berlin und eines Praktikumsleitenden an der Technischen Hochschule Wildau.² Es nahmen fünf Praktikums-

²Angefragt wurden die beiden Praktikumsleitenden der Physik, zwei Praktikumsleitende der Chemie (organische und physikalische Chemie), der IT-Learning-Leiter im Veterinärmediziner*innenpraktikum und die Praktikumsleitung der Pflanzenphysiologie der Freien Universität Berlin, sowie der Praktikumsleitende des Biotechnologiepraktikums an der TH Wildau.

leiter*innen der Fachrichtungen Physik, physikalische Chemie, organische Chemie, Biotechnologie und Veterinärmedizin an den Expert*inneninterviews teil.

Vorbereitung der Erhebung Methodisch handelte es sich bei den Expert*inneninterviews um kognitive Interviews zu den Items des PraQ (PRÜFER & REXROTH, 2005). Der zusätzliche Interview-Leitfaden wurde nach KRÜGER et al. (2013, S. 122) erstellt und zielte auf ein ermittelndes Interview unter Verwendung von Alltagssprache ab. Der Leitfaden (vgl. Anhang D auf Seite 283) bestand aus einem allgemeinen Einstieg in das Thema der Verbesserung von Praktika, gefolgt von einer erzählgenerierenden Aufforderung³. Dies diente dem Vertrautmachen mit dem Thema und erster Exploration von neuen Aspekten für Forschungsfrage 2. Darauf folgten die Leitfragen des kognitiven Interviews mit der think-aloud-Technik, etwa für die Items zum *Lernzuwachs*:

Welche dieser Bestandteile der Fach-/Methoden-/Kommunikations-/Kooperations-/Personal-/Bewertkompetenz sind für ihr Praktikum relevant, welche nicht? Warum?

So wurde jedes Item besprochen. Bei Unklarheiten wurde mit der Leitfrage »Ich weiß, Sie haben das schon einmal erläutert, aber Sie helfen mir, das besser zu verstehen, wenn Sie das jetzt nochmal beschreiben.« reagiert. Für den Fall langer Sprechpausen wurde mit »Bitte sagen Sie, woran Sie jetzt denken.« interveniert. Die Abschlussfrage für Forschungsfrage 2 lautete

Haben wir etwas vergessen, was Sie gerne noch ansprechen möchten?

In Absprache mit den Interviewten wurden die Interviews zur besseren Auswertbarkeit audiographisch aufgezeichnet und im Nachhinein je Item ausgewertet.

Auswertungsmethode Die Auswertungsmethode fokussierte einerseits auf diejenigen Skalen und Items, die Widerstände oder Ablehnung bei den Interviewten im Rahmen des kognitiven Interviews hervorriefen.

Andererseits boten die erste und letzte Frage des Leitfadens den/r Interviewten die Möglichkeit, neue für relevant befundene Aspekte der Praktikumsqualität zu ergänzen (Forschungsfrage 2). Für Forschungsfrage 2 wurden alle Vorschläge für neue Indikatoren gesammelt, zur späteren Sichtung und Diskussion.

Die Stichprobe der Praktikumsleitenden stellt zwar eine Expert*innengruppe für den Ablauf, die Inhalte und die Handlungen in Praktika sind, jedoch nicht zwangsläufig eine Expert*innengruppe für pädagogisches und didaktisches Handeln inner-

³»Beschreiben Sie bitte einmal kurz, was Ihnen zum Thema »gutes Praktikum« einfällt.«

halb des Praktikums dar. Da jedoch einige Skalen des PraQ pädagogische und didaktische Inhalte abdecken, kann es passieren, dass in den Expert*innenbeurteilungen Praktikumsleitende Items auf naiver Lehrerfahrung beruhend ablehnen. Hier oblag es der Einschätzung des, fachdidaktisch und pädagogisch ausgebildeten, Auswertenden, dies zu beurteilen.

Je abgelehntem Item je Experte/Expertin wurde dann eine Zuordnung zu entweder einem echten Passungsproblem oder einer nicht auf Expertise beruhender Ablehnung vorgenommen. Wurde ein Item nach einem Interview als echtes Passungsproblem eingestuft, so erhielt es 1 Ablehnungspunkt. Lagen keine solchen Probleme, aber ein Akzeptanzproblem vor, so erhielt es 0.5 Ablehnungspunkte. Nach fünf erfolgten Interviews wird dann bei über 2.5 Ablehnungspunkten das Item aus dem Itempool ausgeschlossen, da es oft (begründete) Ablehnung in den einzelnen naturwissenschaftlichen Praktika erfahren hat. Für Items kleiner oder gleich 2.5 und größer 1 Ablehnungspunkt wurde in späteren Erhebungen die Antwortoption »passt nicht« ergänzt, da es sich um ein in manchen Praktika unpassendes Item handelt. Es wurden auch für nicht-entfernte Items nützliche Verbesserungsvorschläge für die Itemformulierungen mit eingearbeitet.

Durchführung Die Durchführung verlief gemäß Leitfaden (vgl. Anhang D auf Seite 283) und wurde im August und November 2014 vorgenommen.⁴ Ort der Durchführung waren je die Büros der Praktikumsleitenden, die Interviews dauerten zwischen 45 und 90 Minuten.

12.3.1.3. Ergebnisse

Die Ergebnisse der Zuordnung fassen die Tab. 12.2 und 12.3 für die Items mit mindestens einem Ablehnungspunkt zusammen.

Es wird ersichtlich, dass von 140 Items lediglich 15 Items nicht direkt von den meisten Praktikumsleitenden angenommen wurden. Von diesen 15 Items wurden sechs mit einem Ablehnungsscore von über 2.5 ausgeschlossen, die verbleibenden neun bildeten nunmehr Items mit »passt nicht«-Option.

Gegenargumente der Interviewten bezogen sich mehrheitlich darauf, dass der Inhalt eines Items »nicht vorkomme« oder, dass dies erst in Fortgeschrittenenpraktika forciert würde (vgl. Abschnitt 14.4.4). Es gab auch Argumente, die auf geringe Akzeptanz pädagogischer oder didaktischer Lernprinzipien schließen ließen. So argumentierte ein Experte etwa bei der *Lehrkompetenz*-Skala *Gute Lernatmosphäre*

⁴Es wurden auch praktikumsvergleichende Filterfragen gestellt. Diese werden hier allerdings nicht ausgewertet.

Tabelle 12.2.: A) Items mit hoher Ablehnungsquote bei den interviewten Praktikumsleitenden. Grau gedruckt sind Items, die wegen zu hoher Ablehnung entfernt wurden, die restlichen dargestellten Items erhielten fortan eine »passt nicht«-Antwortoption.

Itemwortlaut: Auf Grund dieses Praktikums	Konstrukt	# Passung	# Akzeptanz	# gesamt
...kann ich Widersprüche und Ähnlichkeiten behandelte Sachverhalte (z. B. Widersprüche zwischen verschiedenen Modellen oder Verfahren) herausarbeiten.	Fachkompetenz	1	2	2
...weiß ich in der Planung besser, welche Messgrößen ich während des Experimentierens verändern muss.	EXP: Versuchsplan entwerfen	2	0	2
...hat sich meine Fähigkeit, einen Versuchsaufbau theoretisch zu entwerfen, verbessert.	EXP: Versuchsplan entwerfen	3	2	4
...hat sich meine Fähigkeit, einen Versuchsaufbau praktisch zu entwerfen, verbessert.	EXP: Versuchsplan entwerfen	3	0	3
...kann ich einen Versuch besser planen.	EXP: Versuchsplan entwerfen	2	2	3
...kann ich zum Experiment gehörende Gerätschaften besser auswählen.	EXP: Geräte zusammenstellen	1	2	2
...gelingt mir das Finden geeigneter Gerätschaften leichter.	EXP: Geräte zusammenstellen	1	4	3

#: Ablehnungspunkte

grau: entfernte Items

EXP: *Experimentelle Kompetenz* (PraQ)

Tabelle 12.3.: B) Weitere Items mit hoher Ablehnungsquote bei den interviewten Praktikumsleitenden.

Itemwortlaut: Auf Grund dieses Praktikums	Konstrukt	# Passung	# Akzeptanz	# gesamt
... kann ich Experimente besser nach meinem eigenen Versuchsplan durchführen.	EXP: Messungen durchführen	2	0	2
... kann ich beim Experimentieren besser auf unerwartete Probleme reagieren.	EXP: mit Problemen und Fehlern umgehen	2	1	2.5
... kann ich eher Verbesserungsmöglichkeiten für Experimente benennen.	Bewertungskompetenz	1	1	1.5
... kann ich statistische Fehlerquellen besser diskutieren.	Bewertungskompetenz	3	0	3
... eröffnet den Studierenden Möglichkeiten, sich mit interessanten Sachverhalten eingehender zu beschäftigen.	zur Selbsttätigkeit anregen	0	3	1.5
... verdeutlicht die Relevanz der Experimente für unterschiedliche Anwendungszusammenhänge.	Relevanz verdeutlichen	0	3	1.5
Das Skript macht den Stellenwert der Sachverhalte im Praktikum für das weitere Studium klar.	Skriptqualität	1	2	2
Die Praktikumsgeräte sind auf dem neuesten Stand.	Räumlich-sachliche Ausstattung	2	2	3
#: Ablehnungspunkte grau: entfernte Items EXP: <i>Experimentelle Kompetenz</i> (PraQ)				

herstellen bezüglich der Studierenden, dass »...die im Praktikum arbeiten und nicht sich wohlfühlen sollen...«.

Aus den Interviews, vornehmlich bei der Eingangs- und Abschlussfrage, neu gewonnene Indikatoren (Forschungsfrage 2) sind Tab. 12.4 zu entnehmen. Diese wurden nach Abschluss der fünf Interviews erneut den Expert*innen vorgelegt, diesmal in Form einer Email, mit analogen Aufforderungen zur Bewertung wie beim Interview. Keines der neuen Items erhielt einen Ablehnungsscore von über 1.

12.3.1.4. Diskussion

Die sechs ausgeschlossenen Items bilden die Antwort auf die Forschungsfrage 1. Diese betreffen zu einem Großteil das Planen eines Experiments, was in den fünf assoziierten Grundpraktika anscheinend keine Zielsetzung der Veranstaltungen darstellt. Häufig wird hier auf Fortgeschrittenenpraktika verwiesen. Die Herausnahme dieser Items nur auf Basis von Expert*innenurteilen kann kritisch gesehen werden, da die Evaluation nicht nur Ziele der Expert*innen, sondern vielleicht davon abweichende Ziele des Moduls, des Studiums oder der Gesellschaft erfassen sollte (vgl. Abschnitt 4.1 auf Seite 39). Die Modularisierungsfestlegungen der Freien Universität Berlin (FUB, 2012a, S. 7) beispielsweise fordern die »selbstständige[n] Erarbeitung von Fragestellungen und Lösungsmöglichkeiten« im Rahmen naturwissenschaftlicher Praktika, was durchaus auch das Planen von Experimenten mit einschließen kann. Für zukünftige Analysen des inhaltlichen Aspektes von Messungen mit dem PraQ-Instrument wird daher empfohlen, eine gemischte Expert*innengruppe aus Praktikumsleitenden, Studierenden, Modulverantwortlichen, Fachdidaktiker*innen und Abgeordneten lehrrelevanter Gremien zu interviewen. Zumindest für die Items der Skriptqualität wurde daraufhin eine zweite Expert*innenperspektive untersucht (Studie 1.2). Bezüglich der Forschungsfrage 2 ergaben sich zehn neue Indikatoren im Konstruktbereich »Experimentelle Kompetenz«. Dies betont erneut die Relevanz dieses Kompetenzbereichs für naturwissenschaftliche Praktika und stärkt die Annahme, dass diese im Rahmen einer Lehrevaluation von Praktika erfasst werden sollte. Sämtliche neuen Items zeigten sich bei der erneuten Begutachtung durch die Expert*innen als akzeptiert, sind also als gewinnbringende inhaltliche Ergänzung des Itempools anzusehen. Wenige weitere Items des Itempools wurden im Rahmen informeller Gespräche mit Methoden-Expert*innen⁵ verändert oder entfernt, wenn starke Argumente dafür herausgearbeitet werden konnten. Dies geschah etwa, wenn

⁵Persönliche Kommunikation des Autors mit Dr. Blüthmann, M.A. Nowakowski und Dr. Bergann des Fachbereichs Erziehungswissenschaft und Psychologie der Freien Universität Berlin. Berlin, 9. September, 2014.

Tabelle 12.4.: Gemäß Forschungsfrage 2 neu gewonnene Qualitätsindikatoren für Praktika in Itemform, den vorhandenen Konstrukten zugeordnet. Die Expert*innenvorschläge resultierten vornehmlich aus der Eingangs- und Abschlussfrage. Keines der neuen Items erhielt einen Ablehnungsscore von über 1.

Konstrukt	Neues Item gemäß Expert*innenvorschlag
EXP: Fragestellung klären	... kann ich die Fragestellung besser auf den konkreten Versuch beziehen.
EXP: Erwartungen formulieren	... kann ich konkrete Erwartungen an den Ausgang eines Experiments formulieren.
EXP: Geräte zusammenstellen:	... erkenne ich besser, welches Gerät ich vor mir habe. ... weiß ich besser, welche Geräte ich konkret für den Versuch/das Experiment benötigen werde.
EXP: Versuchsanordnung aufbauen	... kann ich Bestandteile des Versuchsaufbaus besser anbringen. ... erkenne ich Aufbaufehler besser.
EXP: Messungen durchführen	... kann ich die einzelnen Geräte besser montieren. ... beherrsche ich mehr Messverfahren als vorher.
EXP: Messungen dokumentieren	... bin ich besser in der Lage, eine vollständige Dokumentation der Datenaufnahme zu gewährleisten. ... habe ich einen besseren Überblick, welche Daten im Messprotokoll/Laborbuch dokumentiert werden müssen.
EXP: Experimentelle Kompetenz (PraQ)	

es sich um ein inhaltlich gedoppeltes Item handelte oder das Abstraktionsniveau der Itemformulierung ungleich dem der anderen Items der Skala war. Zusammenfassend kann nach der erfolgten Überarbeitung des PraQ also davon ausgegangen werden, dass der PraQ hinreichend viele, aber keine unnötigen und unpassenden Aspekte der Qualität von Praktika enthält und zudem als Lehrevaluationsinstrument eine gewisse Akzeptanz mitbringt. Die Interpretation von Messergebnissen mit PraQ als »Praktikumsqualität« scheint damit gerechtfertigt.

12.3.2. Studie 1.2 Relevanz und Vollständigkeit: Skriptqualität

12.3.2.1. Einleitung

In Kooperation mit einem Masterstudierenden der Freien Universität (REHFELDT et al., 2015a) wurden weitere Expert*innen zur Relevanz von PraQ-Items befragt. Thematisch beschränkte sich dies aus ökonomischen Gründen auf den Bereich der *Skriptqualität*. Besonderheit war die veränderte Stichprobe, es handelte sich nun um Fachdidaktiker*innen der Physik und nicht mehr um Praktikumsleitende. Damit wurde den Erfahrungen aus Studie 1.1 begegnet, nach denen unterschiedliche Expert*innengruppen auch für die inhaltliche Betrachtung des PraQ wichtig sind.

Forschungsfragen Die Forschungsfragen sind analog zur Studie 1.1, hier lediglich zu den Items der Skalen *Skriptqualität* und *Skriptqualität für den Versuch*.

Forschungsfrage 1: Welche Indikatoren der Skriptqualität sind nicht konstruktrelevant für Praktikumsqualität?

Forschungsfrage 2: Welche Aspekte von Skriptqualität fehlen im PraQ?

12.3.2.2. Methode

Stichprobe Als Stichprobe wurden $N = 30$ wissenschaftliche Mitarbeiter*innen aus Arbeitsgruppen deutscher Hochschulen gewonnen, die sich mit Physikdidaktik beschäftigten, also mit Theorien des Lehrens und Lernens, sowie des Einsatzes von Medien zur Unterstützung des Lernens von Physik vertraut sind. Es handelte sich um eine freiwillige Teilnahme, angeschrieben wurden deutschlandweit 55 Physikdidaktiker*innen.

Instrument Bei der ausgewählten Methode handelte es sich um ein Expert*innen-rating. Die einzelnen PraQ-Items der *Skriptqualität* wurden durch die Expert*innen auf ihre Relevanz für die *Skriptqualität* hin auf einer 6-stufigen Likertskala eingeschätzt. In einem offenen Kommentarfeld sollten Ergänzungen für Items getätigt werden, falls Aspekte der *Skriptqualität* nicht bedacht worden seien.

Auswertungsmethode Als Maß für die Relevanz der Items sollte der Median der Ratings dienen: Bei einem Median von 4 oder mehr wurde das betreffende Item von der Mehrheit der Probanden für relevant befunden, da dies einem Wert über der Mitte der Skala entspricht (3.5 ist der Neutralitätspunkt).

Durchführung Die Befragung wurde online-basiert im Juli und August 2015 durchgeführt. Die Teilnehmer*innen wurden über Emails kontaktiert.

12.3.2.3. Ergebnisse

Wie Abb. 12.1 zu entnehmen ist, lag der Median der Ratings bei allen Items bei einem Wert von 4 oder mehr, sie wurden also alle als relevant für die *Skriptqualität* befunden. Darüber hinaus ergab die qualitative Auswertung der Kommentare (»Kritisieren und Ergänzen«), dass die Expert*innen vor allem ihre bereits quantitativ ausgefüllten Einschätzungen verbalisierten. Darüber hinausgehende Kritik oder Ergänzungsvorschläge für etwaige neue Indikatoren wurden nicht vollzogen, allerdings wurde die Kommentarfunktion selten genutzt.

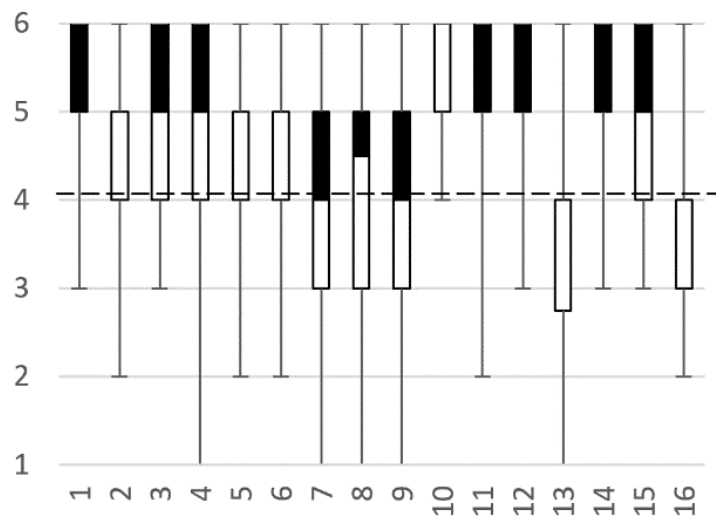


Abbildung 12.1.: Expert*innenrating-Ergebnisse, dargestellt im Boxplot. Der Median aller Items liegt über oder im Cutoff-Wert 4. Die Items 1 bis 3 gehören zur Skala *Skriptqualität für den Versuch*, Items 4 bis 16 sind der Skala *Skriptqualität* zugeordnet. Itemformulierungen finden sich im Anhang A und in Studie 4, die dieselbe Erhebung nutzte.

12.3.2.4. Diskussion

Die Forschungsfrage 1 lässt sich im Rahmen dieser Studie derart beantworten, dass alle vorliegenden Items zur *Skriptqualität* von Expert*innen der Fachdidaktik Physik für relevant befunden wurden. Die Skalen decken demnach keine irrelevanten Indikatoren ab.

Forschungsfrage 2 ist auf Grund von nicht getätigten Verbesserungsvorschlägen der Expert*innen bedingt damit zu beantworten, dass anscheinend keine relevanten Indikatoren fehlen. Bedingt daher, als das das Nicht-Ausfüllen des offenen Kommentarfelds auf eine geringe Motivation zum Ausfüllen dieses aufwändigeren Fragetyps zurückzuführen sein kann. Insgesamt zeigen sich die Skalen der *Skriptqualität* in dieser Studie inhaltlich solide.

12.4. Studie 2: Strukturelle Eigenschaften (»faktorielle Validität«)

Zweite Stufe der Testevaluation ist die Analyse der strukturellen Eigenschaften der Messung des PraQ. Das 3L-Modell der Praktikumsqualität ist als theoretisches Modell konstruiert worden und bedarf daher einer empirischen Prüfung der Messbarkeit der enthaltenen Konstrukte. Für den strukturellen Aspekt (MESSICK, 1995, S. 745) ist hierbei die Bestimmung der Beziehung zwischen Konstrukt und Messstimulus (hier: Item) entscheidend bzw. der Vergleich mit der theoriegegebenen Struktur.

Aus der Natur der Modellentstehung ergibt es sich hierbei, dass einige Messbeziehungen bereits empirisch beforscht sind (z. B. einige LeKo- oder BEvaKomp-Skalen), andere wiederum einer solchen bedürfen.

Auf jeden Fall muss bei einer Neuzusammenstellung von etablierten und nicht-etablierten Skalen die Beziehung dieser untereinander empirisch beforscht werden, wenn die Interpretation als verschiedene Konstrukte möglich sein soll, was Anspruch der Evaluation und fachdidaktischen Forschung ist (Ziele 1 und 2, vgl. Abschnitt 8).

Die Studie ist hierbei in zwei aufeinander aufbauende Studien untergliedert. Innerhalb der ersten Studie wird die Item-Faktor-Struktur der Messung mittels PraQ explorativ untersucht. Im zweiten Schritt wird die dort gefundene Struktur genutzt, um die Adäquatheit des Messmodells konfirmatorisch zu prüfen. AERA et al. (2014, S. 26–27) schlagen als Methode Faktorenanalysen vor, um Interpretationen der Testergebnisse, die von der Item-Konstrukt-Struktur abhängen, zu evaluieren (Standard 1.13). Die im Rahmen dieser beiden Studien erhobenen Daten sind auch Grundlage für die Untersuchung inhaltlicher Forschungsfragen in Kapitel 13.

12.4.1. Studie 2.1: Explorative Analyse

12.4.1.1. Einleitung

Der erste Schritt der strukturellen Testevaluationsstudie ist die explorative Analyse der Faktorstruktur der PraQ-Messung. Da die disjunkten Messungen zweier Fragebögen PraQ-A und PraQ-B zu analysieren sind, ist diese Studie in der Darstellung der Ergebnisse und der Diskussion zweigeteilt.

Forschungsfragen Ergebnisse der Messungen verschiedener Konstrukte der Praktikumsqualität sollen voneinander trennbar interpretiert werden dürfen. Dazu gehört auch eine klare Item-Konstrukt-Zuordnung (AERA et al., 2014, S. 16). Es soll also in der Interpretation der Testergebnisse möglich sein, die Qualität einzelner Dimensionen bzw. qualitätsrelevanter Konstrukte voneinander getrennt untersuchen zu können.

Forschungsfrage 3: Welche Konstrukte lassen sich beim PraQ theoretisch wie empirisch voneinander getrennt interpretieren? (»faktorielle Validität«)

Die Forschungsfrage impliziert hierbei auch, dass Items klar einem Konstrukt zuzuordnen sein müssen, da sonst das entsprechende Item der Trennbarkeit der Konstrukte widersprechen würde. Dies bedeutet im Idealfall, dass ein Item stets nur einem Konstrukt zugeordnet wird und dies empirisch, nicht nur theoretisch.⁶

Die empirisch nachzuweisende, eindeutige Zuordnung von Items zu genau einem Konstrukt (Einfachstruktur) dient dem Ziel, tatsächlich das zu untersuchende Konstrukt nach der Messung in seiner Ausprägung interpretieren zu können. Die Zuordnung von Items zu Konstrukten sollte andererseits mit einer entsprechend messbaren »Zuordnungsstärke«, der Faktorladung ausfallen. Diese kann im Zusammenhang mit konstruktgleichen Items erforscht werden und zeigt an, inwieweit das Item zur

⁶Zum Beispiel sollte ein Item für *schriftliche Kommunikationskompetenz* keine allzu großen empirischen Zusammenhänge mit dem Konstrukt *mündliche Kommunikationskompetenz* aufweisen. Ist das Gegenteil der Fall, so kann dies mehrere Ursachen haben.

Erstens könnte es sein, dass die beiden Konstrukte empirisch gar nicht trennbar sind, d. h. etwa, dass eine Änderung in der mündlichen Kommunikationskompetenz sehr häufig auch mit einer analogen Änderung der schriftlichen Kommunikationskompetenz einherginge.

Zweitens ist denkbar, dass das Item selbst bzw. seine Operationalisierung in sprachlicher Form inhaltlich mit beiden Konstrukten überlappt. In unserem Beispiel der schriftlichen und mündlichen Kommunikationskompetenz könnten dies z. B. Items sein, die die Verbesserung der Fähigkeit »zu Formulieren« beinhalten, was sowohl auf das Mündliche als auch auf das Schriftliche bezogen werden könnte.

Drittens ist es möglich, dass die Operationalisierung in Form eines Items Methodeneffekte bzw. eine Spezifität mit dem Fremdkonstrukt teilt, dies könnte eine sehr ähnliche Satzformulierung sein. Weitere Gründe sind denkbar.

Bestimmung des Konstrukts messtechnisch beiträgt. Das hat für eine Interpretation der Messergebnisse den Vorteil, dass nicht-konstruktbestimmende Items bereits in der Studie 2.1 entfernt werden können. Nebenbei wird eine Reduktion des sehr großen Itempools der PraQ-Instrumente erreicht, die zur Durchführungsökonomie beiträgt.

Ein Konstrukt bzw. Faktor sollte zudem eine gewisse Anzahl von bestimmenden Items aufweisen, idealerweise mehr als zwei, um die Messgenauigkeit (statistisch: Reliabilität) schätzen zu können (BÜHNER, 2006).⁷

Es sind folgende gewünschte Eigenschaften festzuhalten:

- a) Items sollen sich genau einem Konstrukt zuordnen lassen.
- b) Items sollen einen wesentlichen Beitrag zur Konstruktmessung leisten (Faktorladung).
- c) Die Itemanzahl je Faktor sollte größer als zwei sein, bei guter Reliabilität der Messung.

12.4.1.2. Methode

Stichprobe Mittels PraQ-Instrument wurden für die Studien 2.1 und 2.2 teilnehmende Studierende von 18 naturwissenschaftlichen Grundpraktika an zwölf Standorten im deutschsprachigen Raum befragt.⁸

Notwendiges Kriterium für die Teilnahme an der Studie sind dem Abschnitt 10.2 zu entnehmen.⁹

⁷ Dies begründet sich zudem dadurch, dass Konstrukte verschiedene Facetten ihrer Realisierung aufweisen, so zeigt sich *schriftliche Kommunikationskompetenz* in mehreren Tätigkeiten im Praktikum, z. B. »Protokolle gliedern« und »fachsprachlich schreiben«, nicht nur einer einzigen. Bezogen auf eine spätere Interpretation der Ergebnisse bedeutet das, dass die Interpretation auf einer facettenreichen, ausgeglichenen Erfassung des Konstrukts fußt und somit inhaltlich gehaltvoll ist. Der Miteinbezug der Messgenauigkeit in die Interpretation kann einen Beitrag dazu liefern, messfehlerbehaftete Ergebnisse nicht überzuinterpretieren und umgekehrt.

⁸ Auf Grund des Datenschutzes werden die Standorte und die Fächergruppen getrennt dargestellt, eine Auswertung im Sinne eines Praktikumsrankings ist auf Grund des noch nicht validierten Status des Instruments an dieser Stelle inhaltlich unzulässig und nicht mit den Datenschutzvereinbarungen der Universitäten konform. Für die Gewinnung der Praktika wurden mehrere Zugangswege beschritten: Es wurden Websites der Universitäten besucht und praktikumsleitendes Personal direkt per Email angeschrieben, teilweise auch betreuendes Personal. Über persönliche Kontakte innerhalb der Fachdidaktiken (hauptsächlich Physik) konnten weitere Praktika rekrutiert werden. Eine Teilnahme an der Praktikumsleitertagung »DPG-Schule« der AG Physikalische Praktika mit entsprechender Präsentation des PraQ in 2015 lieferte weitere Kontakte.

⁹ Für den strukturellen Aspekt ist diese Positivselektion allerdings von geringer Bedeutung, sofern nicht ein gänzlich anderes »Funktionieren« der Items in Abhängigkeit von der Kooperationsbereitschaft der Praktikumsleitenden angenommen wird.

12. Testevaluation (Validität und Reliabilität)

Die insgesamt zwölf beteiligten Standorte waren Berlin, Potsdam, Wildau, Kiel, Bielefeld, Wuppertal, Tübingen, Trier, Köln, München, Aachen und Wien (vgl. Tab. 12.5).

Eine Stichprobenverzerrung kann sich dadurch ergeben haben, dass die Teilnahme freiwillig und ohne monetäre Anreize verlief, allerdings lassen Zusagequoten von 95% der Physik-Praktika und 66% der Chemiepraktika, sowie Rücklaufquoten der Studierenden von immer über 70% je Praktikum auch Gegenteiliges vermuten.¹⁰

Für die Studie 2.1 des PraQ-A¹¹ waren drei Berliner Praktika der Chemie beteiligt, zudem jeweils ein physikalisches Praktikum aus Berlin, Kiel, Wien und Wuppertal, sowie ein biologisch orientiertes Praktikum aus Wildau. Der PraQ-B umfasste in dieser Studie Daten aus den drei Praktika der Physik aus Berlin und Wien.

Alle übrigen beschriebenen Praktika nahmen an der Studie 2.2 teil.¹²

Schlussendlich konnten für die Studie 2.1 $N_A = 237$ und $N_B = 241$ Teilnehmer*innen gewonnen werden. Die Stichproben beschreibt Tab. 12.6 genauer.

Auswertungsmethode Da die Itemkonstruktion z. T. induktiv bzw. per Adaption bzw. Ergänzung etablierter Skalen erfolgte, wurde für die strukturelle Validierung dieser Instrumentteile zunächst eine explorative Faktorenanalyse durchgeführt (EFA, BROWN, 2006, S. 194).¹³

Die benötigte Stichprobengröße zur Durchführung einer EFA ist nicht pauschal anzugeben, da sie von vielen je Studie verschiedenen Faktoren abhängt (Itemanzahl, Faktoranzahl, Itemanzahl je Faktor etc.). Als grobe Richtlinie können aber Empfehlungen aus der Literatur dienen, beispielsweise $N \approx 300$ (BORTZ, 2010, S. 396; ROUQUETTE & FALISSARD 2011; HENSON & ROBERTS 2006, S. 402), für Faktoren mit vielen Items aber auch $N \approx 150$ (BORTZ, 2010, S. 396).

Kern der Auswertung stellen die Ladungen der Items auf die Faktoren dar. Gemäß HAIR et al. (1998, S. 111) sind Ladungen erst ab einer Höhe von .40 als interpretierbar zu bezeichnen, ab .50 gelten sie als praktisch relevant.

Für den PraQ-A konnten die Items der Skalen *Kooperationskompetenz* (Fast unverändert nach BRAUN & GUSY, 2006) und *Personalkompetenz: Interesse* (ebd.)

¹⁰RICHARDSON (2005, S. 405) sieht Rücklaufquoten als ausreichend, wenn sie 50% überschreiten, was in allen Fällen vorlag.

¹¹damals noch als »PraKo A« bezeichnet

¹²Wobei auch einige Praktika mit zwei Kohorten je in Studie 2.1 und Studie 2.2 insgesamt vertreten waren.

¹³Die EFA dient hierbei der Exploration von Itemclustern zu Faktoren, die als dahinterliegende Konstrukte interpretiert werden können (FIELD, 2013, S. 628) und am Ende einen übersichtlichen Blick auf die Praktikumsqualität gewähren.

Die EFA basiert auf Korrelationshäufungen der Items und gibt Aufschluss darüber, wie stark ein Item zu einem Faktor gehört.

Tabelle 12.5.: Standorte der Praktika, in denen die Erhebungen für Forschungsfrage 3 stattfanden.

Standort	Physikpraktika	Chemiepraktika	sonstige Praktika
Berlin	3 (FU, HU, TU)	3 (FU, HU)	1
Potsdam	1		
Wildau			1
Kiel	1		
Bielefeld		1	
Wuppertal	1	1	
Tübingen	1		
Trier	1		
Köln	1		
München (LMU)	1		
Aachen	1		
Universität Wien	1		

aus der EFA ausgeschlossen werden, da sie bereits etablierte Maße darstellen. Sie werden aber Bestandteil der CFA in Studie 2.2 sein.

Folgende Konstrukte des PraQ-B konnten als bereits etablierte Skalen operationalisiert werden und bedürfen daher keiner explorativen, sondern nur einer späteren konfirmatorischen Untersuchung: *Verständlich erklären*, *Verknüpfungen herstellen*, *gute Lernatmosphäre herstellen*, *zur Selbsttätigkeit anregen*, *Effizient mit Störungen umgehen* (in THIEL et al., 2012); *Skriptqualität für den Versuch* (etabliert in REHFELDT et al., 2014) und *klassische räumlich-sachliche Ausstattung* (formatives Konstrukt). Die restlichen der Konstrukte des PraQ-B wurden explorativ faktorenanalytisch untersucht.

Konkret wurde sich für eine Analyse mittels Hauptkomponentenanalyse entschieden, die eine näherungsweise multivariate Normalverteilung der Items erfordert (FIELD, 2013, S. 650).¹⁴ Diese bietet den Vorteil, robust gegenüber Multikollinearität zu sein (FIELD, 2013, S. 658).

Da für die meisten Konstrukte aus lerntheoretischer Sicht kausale Zusammenhänge erwartet werden, wurde in der EFA eine Korrelation der Faktoren zugelassen (»oblimin-Rotation«).

Für die Faktorenextraktion stand als Richtlinie das KAISER-GUTTMANN-Kriterium (KAISER, 1960), bei leichtem Unterschreiten des Eigenwertes 1 wurde bei inhaltlicher Plausibilität der betreffende Faktor trotzdem extrahiert, das Ziel war eine Ein-

¹⁴ Aus Gründen sprachlicher Vereinfachung wird der Terminus »EFA« aber weiterhin verwendet, da gemäß FIELD (2013, S. 650ff) ohnehin wenig Unterschiede in den Ergebnissen einer EFA und einer PCA zu erwarten sind

Tabelle 12.6.: Stichprobenbeschreibung für die Studie 2.1.

Studien- gänge	Sem	Fach- praktikum	fachfremd Physik	fachfremd Chemie	Block	weibl	Tan- dem
PraQ-A	45% Phy	58% Phy					
	33% Chem	31% Chem	67%	0%	31%	31%	80%
	5% Bio	5% Bio					
	18% sonst.	6% sonst.					
PraQ-B	52% Phy	88% Phy					
	29% Bio	4% Chem	37%	30%	2%	34%	92%
	5% Chem	2% Bio					
	14% sonst.	6% sonst.					

Studiengänge TN: Studiengänge der teilnehmenden Studierenden

Sem: Fachsemester: $M(SD)$

Fachpr: Fachpraktikum

fachfremd Physik: Anteil fachfremder Studierender in Physikpraktika

fachfremd Chemie: Anteil fachfremder Studierender in Chemiepraktika

Block: War es ein Blockpraktikum?

weibl: Anteil weiblicher Studierender

Tandem: Wurde in Tandems gearbeitet?

fachstruktur zu finden. Technisch wurde die Analyse über SPSS realisiert (Version 22).

Durchführung Es wurde den Praktika je wunschgemäß entweder der PraQ-A oder der PraQ-B zur Durchführung der Erhebung ausgehändigt. Die Erhebung fand in der Praktikumszeit statt und dauerte ca. 15 min. Zeitpunkt war meist im Anschluss an das theoretische Kolloquium.¹⁵

Durch eine Erhebung in den Praktikumsräumen war eine konvergente Umgebung für die Evaluation des Praktikums gegeben. Für Erhebungen im Berlin-Brandenburger Raum und in Wuppertal wurde durch den Autor sprachlich standardisiert erhoben. Alle weiteren Standorte bekamen das Material per Post oder Email, inklusive Durchführungsmanualen für Praktikumsleiter*innen (Koordination) und Praktikumsbetreuer*innen (Erhebung), die auch Leittexte zur Ankündigung der Erhebung enthielten. Die Manuale enthielten Anmerkungen zur Form der Erhebung, Verteilung der Durchführungsanreize und zum Datenschutz und sind in Anhang E auf Seite 285 einzusehen. Darüber hinaus wurde auch telefonisch vom Autor instruiert. Um Effekte sozialer Erwünschtheit zu minimieren, wurde bezüglich des PraQ-B stets der/die letzte Woche erlebte Betreuende bewertet, also nicht der/die bei Erhebung anwesende.

Zudem wurde darauf geachtet, dass der/die aktuell Betreuende nach Möglichkeit nicht im selben Raum befindlich war. Darüber hinaus enthielt das Deckblatt des Fragebogens keine zu bearbeitende Frage, was erneut auf Grund des Datenschutzes geschah.¹⁶

Die Erhebungen der Studie fanden zu Ende des Wintersemesters 2014/2015 statt, als die Ergebnisse der Studie 1 bereits in den PraQ eingearbeitet waren.

12.4.1.3. Ergebnisse

PraQ-A Bei der deskriptiven Vorabanalyse der Items wurden zwei Items entfernt. Ein Item¹⁷ wies einen starken Bodeneffekt auf, da es Zitierregeln adressierte, die in Praktika fast nie Bestandteil der Lehre sind. Ein weiteres Item¹⁸ wies zu wenige

¹⁵ Dieser Zeitpunkt wurde gewählt, da er genau zwischen theoretischem und praktischem Präsenzterminanteil liegt und man die Studierenden daher nicht aus dem Arbeitsfluss bringt, was sich positiv auf die Testbearbeitung und Durchführungsbereitschaft auswirken kann. Des Weiteren herrscht direkt vor dem mündlichen Kolloquium meist eine hohe Anspannung und Erfolgsdruck, weshalb der Zeitraum zu Beginn der Präsenzzeit ebenso nicht in Frage kam.

¹⁶ Vor allem bei Erhebungen, bei denen die Bögen ggf. von Betreuenden eingesammelt und erst dann zum/r Praktikumsleitenden weitergegeben und erst danach an den Forschenden übergeben werden, kann dies wichtig sein.

¹⁷ »... beherrsche ich Zitierregeln besser.«

¹⁸ »... kann ich eher Verbesserungsmöglichkeiten für Experimente benennen.«

Korrelationen (< 12) mit anderen Items auf und wurde daher entfernt (FIELD, 2013, S. 648).

Ein weiterer Schritt betraf den Umgang mit fehlenden Werten. Die gemäß Studie 1.1 definierten Items mit »passt nicht«-Option wiesen überdurchschnittlich oft fehlende Werte auf. Es wurde sich für eine Imputation dieser Werte mit dem EM-Schätzer entschieden (WEAVER & MAXWELL, 2014, S. 147ff), da dieser unabhängig von MCAR-Daten ist und für Faktorenanalysen dann zulässig ist, wenn keine Signifikanztests zwischen Faktorscores vorgenommen werden (BROWN, 2006, S. 368). Gemäß EID (2013, S. 875) liefert dieser Schätzer meist ähnliche Ergebnisse wie eine multiple Imputation und war zum damaligen Zeitpunkt die einzige in SPSS verfügbare Möglichkeit der rechnerischen Imputation fehlender Werte.

Zur weiteren Vorbereitung der EFA wurden die Voraussetzungen geprüft (BORTZ, 2010; FIELD, 2013, S. 607). Für alle Items außer denjenigen zu »Zitierregeln« (entfernt, vgl. Tab. 12.7) lagen hier Werte für die Schiefe vom Betrag kleiner 2 und für die Kurtosis kleiner 7 vor, was für näherungsweise normalverteilte Daten spricht (URBAN & MAYERL, 2014, S. 142).

Als Faktorlösungen wurden in einem iterativen Verfahren für die theoretisch 16 zu erwartenden Konstrukte sowohl 16- als auch 15-Faktoren-Lösungen getestet. Itemausschlüsse ergaben sich hier auf Basis schlechter Kennwerte (Tab. 12.7). Von den anfangs 80 Items des PraQ-A blieben so 68 im Rahmen der EFA erhalten. Die Eigenwerte der finalen unrotierten Lösung mit 15 extrahierten Faktoren waren 22.1 für den ersten Faktor, 5.6 und 4.7 für die folgenden Faktoren. Die Faktoren 4 bis 14 liegen im Bereich [1.0; 2.7], Faktor 15 bei 0.95, Faktor 16 bei 0.85. Die rotierte Lösung weist nahezu eine Einfachstruktur auf (Tab. 12.8 und 12.9), alle noch enthaltenen Items laden auf ihrem Zielfaktor. Lediglich für die Items 40 bis 42 und 65 ergaben sich minimale Nebenladungen.

Bei der finalen EFA lag das KMO-Kriterium bei exakt .90, was eine ausreichende Stichprobengröße für die Analyse nahelegt, der BARTLETT-Test fiel signifikant aus (FIELD, 2013, S. 607). Die Varianzaufklärung des Modells lag bei 75.4%.

Durch die Itemausschlüsse beim PraQ-A (Tab. 12.7) ist ersichtlich, dass keine Analyse der Skala *Fragestellung klären* der *Experimentellen Kompetenz* mehr möglich ist, da die betreffenden Items keinen Faktor bildeten.

Der Faktor *Messungen durchführen* hat auch einige Itemausschlüsse erfahren. Der große anfängliche Itempool dieses Konstrukts wurde mit vier entfernten Items stark verkleinert.

Bei den Faktoren selbst fanden Umstrukturierungen statt. Das Konstrukt *Fragestellung klären* ergab keinen Faktor und entfiel. Das Konstrukt *Fachkompetenz*

Tabelle 12.7.: A) Itemausschluss und Gründe: PraQ-A.

Itemwortlaut: Aufgrund dieses Praktikums...	D	r	KL	EI	K	N	V	Konstrukt
...kann ich Widersprüche und Ähnlichkeiten behandelnder Sachverhalte (z.B. Widersprüche zwischen verschiedenen Modellen oder Verfahren) herausarbeiten.	1	1					gering ^a	Fachkompetenz
...kann ich mich besser in experimentelle Fragestellungen einarbeiten.	1	1					Skala entfernt	Fragestellungen klären
...kann ich die inhaltlichen Versuchsziele eines Experiments besser nachvollziehen.				1			Skala entfernt	Fragestellungen klären
...kann ich die Fragestellung besser auf den konkreten Versuch beziehen.	1	1					Skala entfernt	Fragestellungen klären
...kann ich Experimente besser nach Anleitung des Skripts durchführen.				1			gering ^b	Messungen durchführen
...kann ich Experimente besser nach Anleitung durch den/die Betreuende/n durchführen.	1						gering ^c	Messungen durchführen
Aufgrund dieses Praktikums weiß ich beim Experimentieren besser, welche Messgrößen ich verändern muss.	1						mäßig ^d	Messungen durchführen
... bin ich besser in der Lage, verschiedene experimentelle Verfahren für die Datenaufnahme zu verwenden.	1						gering ^e	Messungen durchführen
...kann ich graphische Darstellungen besser auswerten.	1	1					gering ^f	Daten auswerten
...kann ich graphische Darstellungen besser verstehen.	1						gering ^g	Kommunikation (schriftlich)

zutreffend = 1; D: Deskriptive Auffälligkeit; r: # Korrelationen mit anderen Items (zu gering);

KL: Keine Ladung > .40; EI: Einzelitem bildete eigenen Faktor; K: Kommunalität N: mind. eine Nebenladung > .40; V: Validitätsgefährdung; BE: Bodeneffekt

^ada stark auf der Metaebene^bda durch Skala "Skriptqualität für den Versuch" abgedeckt^cda durch Skala "verständlich erklären" abgedeckt^dFür Physik-Praktika entfält die mathematische Komponente.^eda vor allem in F-Praktika forciert.^fDurch Items "Zusammenhänge in Daten" und "Messwerte interpretieren" abgedeckt.^ginhaltlich gut durch andere Items der Skala abgedeckt

12. Testevaluation (Validität und Reliabilität)

Tabelle 12.8.: A) Mustermatrix der explorativen Faktorenanalyse zum PraQ-A.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Item 1	.82														
Item 2	.69														
Item 3	.84														
Item 4	.70														
Item 5		-.58													
Item 6		-.74													
Item 7		-.64													
Item 8		-.60													
Item 9			.74												
Item 10			.80												
Item 11			.80												
Item 12			.56												
Item 13			.41												
Item 14				.60											
Item 15				.62											
Item 16				.60		.35									
Item 17				.77											
Item 18				.44											
Item 19					-.84										
Item 20					-.65										
Item 21					-.66										
Item 22						.76									
Item 23						.67									
Item 24						.76									
Item 25						.66									
Item 26							-.70								
Item 27							-.55								
Item 28							-.61								
Item 29							-.46								

A: Fachwissen (theoretisch) | B: Fachwissen (praktisch) | C: EXP: Erwartungen formulieren | D: EXP: Versuchsplanung & Geräte zusammenstellen | E: EXP: Gerätekenntnisse deklarativ | F: EXP: Versuchsanordnung aufbauen | G: EXP: Messungen durchführen
EXP: *Experimentelle Kompetenz* (PraQ)

Ladungen vom Betrag $< .30$ wurden aus Lesbarkeitsgründen entfernt. Interpretation der Ladungsgrößen: unwichtig $< .30$ | minimal: $[.30, .40]$ | OK: $[.40, .50]$ | praktisch bedeutsam: $> .50$ (HAIR et al., 1998, S. 111). Etwaige negative Vorzeichen sind Artefakte der Methode.

12.4. Studie 2: Strukturelle Eigenschaften (»faktorielle Validität«)

Tabelle 12.9.: B) Mustermatrix der explorativen Faktorenanalyse zum PraQ-A.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Item 30								-.73							
Item 31								-.68							
Item 32								-.67							
Item 33									.65						
Item 34									.79						
Item 35									.70						
Item 36									.55						
Item 37										.83					
Item 38										.77					
Item 39										.57					
Item 40										.49	.40				
Item 41										.47			-.42		
Item 42										.52			-.4		
Item 43											.51				
Item 44											.55				
Item 45											.52				
Item 46											.44				
Item 47											.56				
Item 48											.52				
Item 49												.72			
Item 50												.80			
Item 51												.76			
Item 52												.78			
Item 53												.73			
Item 54													-.79		
Item 55													-.82		
Item 56													-.78		
Item 57													-.79		
Item 58													-.76		
Item 59													-.44		
Item 60														-.79	
Item 61														-.91	
Item 62														-.87	
Item 63														-.80	
Item 64															.49
Item 65									.33						.62
Item 66															.80
Item 67															.66
Item 68															.62

H: EXP: Messungen dokumentieren | I: EXP: Mit Problemen und Fehlern umgehen | J: EXP: Rohdaten graphisch/ tabellarisch darstellen | K: EXP: Daten interpretieren | L: Kommunikationskompetenz (mündlich) | M: Kommunikationskompetenz (schriftlich) | N: Zeitmanagement | O: Bewertungskompetenz
EXP: *Experimentelle Kompetenz* (PraQ)

zerfiel in zwei Faktoren, die *Fachwissen (theoretisch)* und *Fachwissen (praktisch)* benannt wurden, da sich alle Items des zweiten Faktors auf experimentelle Sachverhalte bezogen.

Für die Komponenten der *Experimentellen Kompetenz*, *Versuchsplan entwerfen* und *Geräte zusammenstellen* ergab sich ein gemeinsamer Faktor *Versuchsplanung & Geräte zusammenstellen*. Allerdings ergab sich für einige Items, die ursprünglich diesen beiden Konstrukten zugeordnet waren, eine Clusterung zu einem weiteren neuen Faktor, der inhaltsgemäß *Gerätekenntnisse deklarativ* benannt wurde. Die Korrelation der auf den Faktoren aufbauenden Skalen ergab eine Spannweite von 0.16 (*Rohdaten auswerten* vs. *Gerätekenntnisse deklarativ*) bis 0.65 (*Daten interpretieren* vs. *Bewertungskompetenz*).

Aufbauend darauf fand eine Itemanalyse statt. Die Trennschärfen der Items je Skala waren stets größer als .51, fast immer größer als .60. Die ersten Reliabilitätsschätzungen mittels CRONBACHS α ergaben eine Spannweite von $\alpha = .79$ für *Gerätekenntnisse deklarativ* bis $\alpha = .93$ für *Kommunikationskompetenz (schriftlich)*.

PraQ-B In der zunächst durchgeführten deskriptiven Betrachtung der Itemkennwerte des PraQ-B traten keine Auffälligkeiten auf.

Aufgrund vieler designbedingter missings¹⁹ wurde sich für einen paarweisen Fallausschluss entschieden, um die größtmögliche Information zu erhalten.

Für alle Items lagen Werte für die Schiefe vom Betrag kleiner 0.5 und für die Kurtosis kleiner 1 vor.

Als Faktorlösungen wurden in einem iterativen Verfahren für die theoretisch 10 zu erwartenden Konstrukte 7- bis 10-Faktor-Lösungen getestet.

Itemausschlüsse ergaben sich hier auf Basis schlechter Kennwerte (Tab. 12.10).

Von den anfangs 45 Items des PraQ-B blieben so 37 im Rahmen der EFA erhalten.

Die Eigenwerte der finalen unrotierten Lösung waren 12.9 für den ersten Faktor, 5.0 und 2.0 für die folgenden Faktoren. Die Faktoren 4 bis 7 liegen im Bereich [1.1; 1.7], Faktor 8 bei 0.91, Faktor 9 bei 0.78. Die rotierte 8-Faktor-Lösung weist nahezu eine Einfachstruktur auf (Tab. 12.11), alle noch enthaltenen Items laden auf ihrem Zielfaktor. Hierbei zeigten sich nur sehr wenige und nur nicht-substanzielle Nebenladungen (vgl. Tab. 12.11).

Bei der finalen EFA lag das KMO-Kriterium bei .90, was eine ausreichende Stichprobengröße für die Analyse nahelegt, der BARTLETT-Test fiel signifikant aus (FIELD, 2013). Die Varianzaufklärung des Modells lag bei 74%.

¹⁹Designbedingte missings lagen z. B. bei der »passt nicht«-Option vor. Eine Wiederholung der EFA ohne die besonders von missings betroffene Skala *Integration* ($N_{neu} = 152$) lieferte eine identische Ladungsstruktur und fast identische Ladungsausprägungen.

Der Ausschluss betraf nur Items, die durch inhaltsähnliche verbliebene Items hinreichend abgedeckt werden (vgl. Tab. 12.10). Eine Gefährdung der inhaltlichen Validitätsaspekte ist so unwahrscheinlich. Allerdings entfiel das gesamte Konstrukt *lern-dienliches Feedback geben*, es ergab sich empirisch kein Faktor, obwohl die Items der Skala substanzielle Korrelationen untereinander aufwiesen. Auch eine erzwungene Erhöhung bzw. Verringerung der Faktorenzahl ergab keine Verbesserung. Allerdings deckten die Skalen zur *Diagnostik* das Konstrukt teilweise ab.

Der erwartete Faktor *Diagnostik* zerfiel statistisch in zwei Faktoren *Diagnostik: Grundeinstellung* und *Diagnostik: situativ*. Alle weiteren Konstrukte blieben erwartungstreu.

Die Skalen-Korrelationen lagen im moderaten bis mittleren Bereich ($r \in [.21; .64]$), was für die statistische Trennbarkeit der Faktoren spricht.

Die Ergebnisse der darauf aufbauenden Itemanalyse zeigen acht reliable Skalen ($\alpha > .80$) mit trennscharfen Items, je bei über $r_{it} > .60$ für alle Items in allen Skalen.

12.4.1.4. Diskussion

Dieser Beitrag hatte nach den erfolgreichen inhaltlichen Testevaluierungen des neuen PraQ-Fragebogens die strukturelle Seite zum Gegenstand, mit der Forschungsfrage 3, die hier mittels explorativer Faktorenanalyse untersucht wurde.

PraQ-A Bezüglich des PraQ-A ergab sich ein inhaltlich wie statistisch tragbares Bild. Nahezu alle durch das 3L-Modell zu erwartenden Konstrukte konnten empirisch in Form von Faktoren wiedergefunden werden, es gab nur sehr wenige strukturelle Änderungen. Die erhaltene, statistisch adäquate 15-Faktor-Lösung enthält hierbei nur wenige kleine, inhaltlich zu rechtfertigende Nebenladungen, sowie vier veränderte Konstrukte und ein neues Konstrukt. Bezüglich der Nebenladungen ergab sich für Item 40²⁰, dass es nicht nur auf den Faktor *Rohdaten auswerten*, sondern auch auf *Daten interpretieren* lud. Da die Nebenladung mit .40 noch als minimal betrachtet werden kann und die Doppelladung inhaltlich sinnvoll ist – gute Dateninterpretationsfähigkeiten können eine gute Musterinterpretation bewirken – wird das Item trotzdem beibehalten und in der Studie 2.2 analysiert, ob in einer latenten konfirmatorischen Analyse die Trennbarkeit möglich ist. Ähnliches gilt

²⁰ »...erkenne ich Zusammenhänge/Muster innerhalb meiner gewonnenen Daten («Zahlen«) besser.«

Tabelle 12.10.: A) Itemausschluss und Gründe: PraQ-B.

Itemwortlaut: Der/die Betreuende...	D	r	KL	EI	K	N	V	Konstrukt
... gibt konstruktives Feedback zum Experimentieren der Studierenden.	1	1					Skala entfernt	lerndienliches Feedback geben
... gibt konstruktives Feedback zum Vortest/Vorgespräch der Studierenden.	1	1					Skala entfernt	lerndienliches Feedback geben
... macht Verbesserungsvorschläge zum experimentellen Aufbau.		1					Skala entfernt	lerndienliches Feedback geben
... unterstützt gezielt einzelne Studierende oder Studierendengruppen, die einen besonderen Unterstützungsbedarf haben.			1				gering	Ermutigen und SWE stärken
... erklärt die Theorie so oft, bis es alle verstanden haben (z. B. in Gesprächen; im Vorgespräch; Nachbesprechungen etc.).		1		1			gering	Diagnostik
... erklärt das Experiment so oft, bis es alle verstanden haben (z. B. in Gesprächen; im Vorgespräch; Nachbesprechungen etc.).	1						gering	Diagnostik
Das Skript (ggf. Zusatzmaterial) hilft mir, meine Protokolle/Versuchsberichte klarer zu gliedern.				1			gering	Skriptqualität
Im Skript wird der Unterschied zwischen Theorie und Experiment klar.				1			mäßig	Skriptqualität

Deskriptive Auffälligkeit; r: # Korrelationen mit anderen Items (zu gering);

KL: Keine Ladung > .40; EI: Einzelitem bildete eigenen Faktor; K: Kommunalität

N: mind. eine Nebenladung > .40; V: Validitätsgefährdung; BE: Bodeneffekt

zutreffend = 1; D:

12.4. Studie 2: Strukturelle Eigenschaften (»faktorielle Validität«)

Tabelle 12.11.: Mustermatrix der explorativen Faktorenanalyse zum PraQ-B.

	A	B	C	D	E	F	G	H
Item 1	.62							
Item 2	.83							
Item 3	.79							
Item 4	.56							
Item 5	.75							
Item 6	.66							
Item 7		-.82						
Item 8		-.69						
Item 9		-.57						
Item 10			.77					
Item 11			.80					
Item 12			.80					
Item 13				.78				
Item 14				.79				
Item 15				.84				
Item 16					.55			
Item 17					.53	.32		
Item 18						.85		
Item 19						.84		
Item 20						.74		
Item 21							.86	
Item 22							.86	
Item 23							.81	
Item 24							.77	
Item 25					-.34		.66	
Item 26					-.36		.62	
Item 27							.86	
Item 28							.64	
Item 29							.75	
Item 30							.55	
Item 31							.74	
Item 32							.79	
Item 33			.37				.41	
Item 34								.82
Item 35								.90
Item 36								.81
Item 37								.73

A: Zusammenfassen & Wiederholen | B: Verständnis überprüfen | C: Relevanz verdeutlichen | D: SWE erhöhen | E: Diagnostik: Grundeinstellung | F: Diagnostik: situativ | G: Skriptqualität | H: Integration der Vorlesung

Ladungen vom Betrag < .30 wurden aus Lesbarkeitsgründen entfernt. Interpretation der Ladungsgrößen: unwichtig < .30 | minimal: [.30, .40] | OK: [.40, .50] | praktisch bedeutsam: > .50 (Hair et al., 1998, S. 111). Etwaige negative Vorzeichen sind Artefakte der Methode.

für die Items 41²¹ und 42²². Item 41 weist eine inhaltlich gut nachvollziehbare Nebenladung mit dem Faktor *Kommunikationskompetenz (schriftlich)* auf, geeignete Darstellungsformen erstellen ist auch ein Zeichen guter schriftlicher Kommunikation. Genauso kann das Repertoire unterschiedlicher Darstellungsformen, gerade in den Naturwissenschaften, die Darstellung von Graphen bedeuten, was positiv von der Kompetenz, *Rohdaten auszuwerten* beeinflusst sein kann (Item 42). Item 65 schließlich weist eine unbedeutende, da sehr kleine Nebenladung auf.

Das Konstrukt *Fragestellung klären* der *Experimentellen Kompetenz* entfiel, wodurch inhaltlich diese Komponente der *Experimentellen Kompetenz* nicht mehr Berücksichtigung findet, eventuell wird dieses Konstrukt durch *Erwartungen formulieren* tangiert, es verbleibt aber die Notwendigkeit einer Neukonstruktion der Skala, wenn diese inhaltliche Komponente im Praktikum Berücksichtigung finden soll.

Beim Konstrukt *Messungen durchführen* ist trotz vierfachem Itemausschluss ein inhaltlich aussagekräftiger Itemstamm erhalten geblieben (vgl. Anhang A.1 auf Seite 270), was hier auch dem groß angelegten Itempool zu verdanken ist. Die Aufteilung des Konstrukts *Fachkompetenz* in einen *theoretisch* und einen *praktisch* orientierten Faktor ist inhaltlich sinnvoll und trägt zu einer weiteren Spezifizierung dieser Kompetenz im Rahmen des 3L-Modells bei.

Die Zusammenführung von *Versuchsplan entwerfen* und *Geräte zusammenstellen* zu *Versuchsplanung & Geräte zusammenstellen* stellt den umgekehrten Fall dar, ist inhaltlich auf Basis der Nähe der Komponenten im Modell SCHREIBERS (2012) aber haltbar.

Der neue Faktor *Gerätekenntnisse deklarativ* wurde dem Kanon der Konstrukte experimenteller Kompetenz zugeordnet, da dies als Voraussetzung für experimentelles Handeln gewertet werden kann. Diese Neuzuordnungen von Items zu Konstrukten machen deutlich, dass ein zunächst exploratives Vorgehen sinnvoll sein kann, da die empirische Struktur Abweichungen von der theoretisch erwartbaren Struktur zeigt. Diese Neuordnung kann in Studie 2.2 dann empirisch auf Stabilität geprüft werden.

Die aus den Faktoren gewonnenen Skalen weisen gute Itemkennwerte mit hohen Reliabilitäten auf, was für die Messgenauigkeit der Skalen spricht. Eine genauere Analyse der Reliabilität wie auch der Korrelationen der Faktoren wird im Rahmen der Studie 2.2 messfehlerbereinigt vorgenommen.

²¹ »... fällt es mir leichter, geeignete Darstellungsformen (Grafiken, Tabellen, Schemata) zu erstellen.«

²² »... nutze ich mehr unterschiedliche Darstellungsformen für die Darstellung meiner Ergebnisse (»Repertoire«).«

PraQ-B Die Studienergebnisse unter Verwendung der explorativen Faktorenanalyse zeigen auf, dass auch Interpretationen struktureller Art für die Ergebnisse des PraQ-B zulässig sind. Die Zuordnung der Items zu inhaltlich sinnvollen Faktoren ist geglückt, die empirische Modellierung gleicht der theoretischen nahezu, lediglich der Bereich der *Diagnostik* für die *Lehrkompetenz* wurde weiter ausdifferenziert, *Diagnostik: Grundeinstellung* steht für eine grundsätzlich aufgeschlossene Haltung gegenüber einer Diagnostik des Lernstands. *Diagnostik: situativ* beinhaltet die direkte Reaktion der Lehrenden während der Interaktion mit den Studierenden.

Manifeste Korrelationen dieser beiden Skalen fallen dabei mittelmäßig hoch aus ($r = .54$), was für eine statistische Trennbarkeit spricht. Dies kann allerdings erst sicher nach der konfirmatorischen Faktorenanalyse in Studie 2.2 bestätigt werden.

Die ohnehin schon in der ursprünglichen Quelle (THIEL et al., 2012, S. 16) mit wenigen Items versehene Skala *Lerndienliches Feedback geben* musste dagegen komplett entfernt werden, die Betreuungsqualität weist daher in genau diesem Bereich eine Interpretationslücke auf.

Es ergeben sich Reliabilitätsschätzungen, die für eine hohe Messgenauigkeit und Replizierbarkeit der Ergebnisse sprechen. Die Trennschärfe lag je bei $r_{it} > .60$ für alle Items in allen acht Skalen. Dies kann allerdings nur als erster Hinweis auf Messgenauigkeit interpretiert werden, da eine konfirmatorische, messfehlerbereinigte Schätzung noch erfolgt (vgl. Ausblick).

Das Ergebnis ist insgesamt als vorläufig zu bezeichnen, da eine vollständige Analyse erst über die konfirmatorisch ausgerichtete Studie 2.2 gewährleistet ist.

12.4.2. Studie 2.2: Konfirmatorische Analyse

12.4.2.1. Einleitung

Für Forschungsfrage 3 wurde im vorherigen Abschnitt eine explorative Faktorenanalyse für diejenigen hypothetischen Skalen durchgeführt, die neu konstruiert oder stark von der Ursprungsquelle abgewandelt wurden. Damit wurden neue Strukturen entdeckt und vermutete untermauert.

Gemäß BROWN (2006, S. 12) folgt daraufhin eine konfirmatorische Faktorenanalyse, die das in der EFA etablierte Modell á priori in die Analyse implementiert und die Passung des Modells zu einer neuen Stichprobe prüft.

Forschungsfragen Die Forschungsfrage lautete:

Forschungsfrage 3: Welche Konstrukte lassen sich beim PraQ theoretisch wie empirisch voneinander getrennt interpretieren? (»faktorielle Validität«)

Die konfirmatorische Analyse kann Aufschluss geben über a) die Güte der Messmodelle der einzelnen Konstrukte (latente Faktoren), b) die Passung des Modells auf die vorliegenden Daten insgesamt und c) die empirische Trennbarkeit der Faktoren (messfehlerbereinigt).²³

12.4.2.2. Methode

Stichprobe Bezüglich des PraQ-A konnten für die konfirmatorische Studie die Fachrichtungen Veterinärmedizin (Berlin), Physik (Tübingen, Köln, Berlin, Potsdam, Kiel, Aachen) und Chemie (Berlin) gewonnen werden. Für den PraQ-B konnten Erhebungen in Praktika der Veterinärmedizin (Berlin), Chemie (Wuppertal, Berlin) und Physik (Berlin, Potsdam, Tübingen, Trier, Köln) realisiert werden. Die Stichprobengröße für den PraQ-A betrug $N_A = 576$ und $N_B = 260$ und lag damit jeweils über den angestrebten 250. Tab. 12.12 beschreibt die Studierenden der Stichprobe.

Auswertungsmethode Je Fragebogendatensatz wurde eine eigene konfirmatorische Faktorenanalyse (CFA) angestrebt. In dieser wird die Item-Faktor-Zuordnung, die Faktorstruktur, á priori definiert. Die Faktorstruktur stammte aus den explorativen Faktorenanalysen der Studie 2.1. Die nur wenig veränderten Skalen nach THIEL et al. (2012) wurden zudem in die CFA mit aufgenommen.

Bei der konfirmatorischen Analyse wird im Wesentlichen geprüft, ob die theoretische Kovarianzmatrix der empirischen gleicht. Erstere beschreibt die theoretischen korrelativen Zusammenhänge zwischen den Items gemäß Struktur der erfolgten EFA, letztere umfasst die empirisch in der Validierungserhebung gemessenen Zusammenhänge.²⁴ Voraussetzungen für eine CFA mit dem Maximum-Likelihood-Schätzer (URBAN & MAYERL, 2014, S. 106) und den Umgang mit fehlenden Werten mittels Full Information Maximum Likelihood (FIML) oder multipler Imputation (BROWN, 2006, S. 370) ist die Erfüllung der multivariaten Normalverteilungsannahme je Item und das missing-at-random-Kriterium.²⁵

Eine missing-Rate von unter 5% je Proband und je Item wurde darüber hinaus als Kriterium festgelegt. Die Kriterien für eine gelungene Übereinstimmung des theoretischen Modells mit den empirischen Daten werden folgendermaßen festgelegt:

²³Die ursprüngliche Forschungsfrage enthält also statistisch drei Teilkomponenten, die es im Folgenden zu analysieren gilt. Dabei ist a) Voraussetzung für b) (BROWN, 2006, S. 129) und c) kann nur dann analysiert werden, wenn das Modell überhaupt in sich stimmig ist, also die theoretische Zuordnung von Items zu Faktoren sich empirisch untermauern lässt.

²⁴Einen guten Einstieg in die Methode der CFA bietet BROWN (2006).

²⁵Dieses Kriterium ist speziell für die missings der »passt-nicht«-Items nicht erfüllt.

Tabelle 12.12.: Stichprobenbeschreibung für die Studie 2.2.

	Studien- gänge	Sem	Fachpr	fachfremd Physik	fachfremd Chemie	Block	weibl	Tan- dem
PraQ-A	39% Phy		82% Physik					
	27% Chem	2.9 (1.3)	5% Chemie	50%	0%	19%	36%	96%
	7% Bio		5% VetMed					
	27% sonst.		8% sonst.					
PraQ-B	27% Phy		60% Phy					
	27% VetMed		25% VetMed					
	15% Chem	2.8 (1.3)	8% Chem	52%	6%	23%	50%	90%
	8% Bio		7% sonst.					
	23% sonst.							

Studiengänge TN: Studiengänge der teilnehmenden Studierenden
Sem: Fachsemester: M (SD)
fachfremd Phy: Anteil fachfremder Studierender in Physikpraktikum
fachfremd Chem: Anteil fachfremder Studierender in Chemiepraktika
Block: War es ein Blockpraktikum?
weibl: Anteil weiblicher Studierender
Tandem: Wurde in Tandem gearbeitet?
VetMed: Veterinärmedizin

Hervorzuheben sind die fit-indices, welche Auskunft über die Passung des Modells zu den Daten geben. Als akzeptabel angesehen werden ein $CFI > .90$ (KLINE, 2011), $TLI > .90$ (BROWN, 2006, S. 87), $RMSEA < .10$ (ebd.) und ein $SRMR < .10$ (BYRNE, 2001), als gute Ausprägungen der fit-indices werden ein $CFI > .95$, $TLI > .95$, $RMSEA < .06$ und ein $SRMR < .08$ angesehen (je BROWN, 2006, S. 87).²⁶

Die Trennbarkeit der Faktoren wiederum betreffen die Faktorkorrelationen, die für klar trennbare Konstrukte unter .90 liegen sollten (CHANG et al., 2011, S. 1227). Technisch wurde die Analyse über das package »lavaan« in R realisiert (Version 3.4).

Durchführung Die Durchführung geschah analog zur Studie 2.1, jedoch ab dem Sommersemester 2015 und unter Ausschluss der in der Studie 2.1 ermittelten kritischen Items. Die Erhebung mittels PraQ-Fragebögen wurde analog zur Studie 2.1 in diversen Hochschulen im deutschsprachigen Raum vorgenommen.

Für die Erhebung wurde eine Stichprobengröße von $N > 250$ je Fragebogen angestrebt.

12.4.2.3. Ergebnisse

Die Ergebnisse werden erneut für den PraQ-A und PraQ-B getrennt dargestellt. PraQ-A und PraQ-B enthielten gemeinsam 30 zu analysierende Skalen. Bevor das Gesamtmodell getestet werden kann, müssen zunächst die einzelnen Messmodelle je Faktor überprüft werden.²⁷

Ergebnisse PraQ-A

Umgang mit missings values Für den PraQ-A konnte wegen geringer Anteile fehlender Werte die Analyse mit der Full Information Maximum Likelihood Methode (FIML) für den Umgang mit fehlenden Werten gewählt werden.

Einzel-Messmodelle der Faktoren Die Auswertung zeigt bezüglich der angelegten Kriterien an die Modellgüte (Tab. 12.13) viele fit-Parameter im Normbereich und nur wenige leicht über dem Normbereich. Überschreitungen der Obergrenze eines Kriteriums lagen nicht vor. Es wurden im Rahmen der Analyse, auch auf Basis der »modification-Indices«, korrelierte Residuen von Items dann zugelassen,

²⁶Für weitere Informationen zu den fit-indices sei auf die in der Tabelle angegebene Literatur oder BROWN (2006) verwiesen (vgl. auch Tab. G.1 in Anhang G auf Seite 291).

²⁷Dabei ist zu betonen, dass für alle dargestellten Modelle keine $sEPCs < .20$ für Item-Residualkorrelationen vorlagen (SARIS et al., 2009, S. 573).

wenn es inhaltlich plausible Gründe für eine über die Faktorvarianz hinausgehende Interkorrelation der Residuen gab, z.B. analoges wording.²⁸

Items wurden dann entfernt, wenn ihre Ladung .50 unterschritt, korrelierte Residuen mit allen übrigen Items in den modification-Indices vorgeschlagen wurden²⁹ oder gravierende Probleme mit der Modellpassung auf ein einzelnes Item zurückzuführen waren³⁰.

Zusätzlich gab es einen Fall, in dem zwei Items stark analogen wordings, aber unterschiedlicher Faktoren, so viel Varianz teilten, dass die Trennbarkeit der betreffenden Konstrukte nicht mehr gewährleistet war. Die beiden Items luden stark auf beiden Konstrukten und wurden, auch auf Grund des besseren Modellfits der neuen Lösung (vgl. Tab. 12.14, *Daten interpretieren* vs. *Rohdaten darstellen*), entfernt.³¹

Eine starke Veränderung der inhaltlichen Breite eines Konstrukts ist lediglich zu erwarten für die Skala *EXP (Experimentelle Kompetenz): Rohdaten darstellen*. Hier wurden Items zum »Messdaten auswerten«, »Erkennen von Zusammenhängen in Daten« und zum »Repertoire der Darstellungsformen« entfernt. Übrig bleiben die Facetten »graphisch darstellen«, »tabellarisch darstellen« und »geeignete Darstellungen erstellen«.

Moderate Einbußen ergaben sich für das *theoretische* und *praktische Fachwissen*, das *Formulieren von Erwartungen* beim Experimentieren, das *Daten Interpretieren* und die *Kooperationskompetenz* mit je einem Itemausschluss.³²

Gesamtmodell und Alternativmodelle Nach der erfolgreichen Etablierung von Einzelmessmodellen für die Konstrukte des PraQ-A konnte nun das Gesamt-Messmodell inklusive aller Items und Konstrukte des PraQ-A getestet werden. Das Gesamtmodell, das sich aus den Einzelmessmodellen zusammensetzte, erreichte einen guten Modellfit ($N = 576$, $SRMR = .04$, $RMSEA = .04$, $CFI = .94$, vgl. Abb. 12.2).

Die Testung möglicher Alternativmodelle umfasste ein Modell, das die *Experimentelle Kompetenz* als einen übergreifenden Faktor annimmt, sowie ein Modell,

²⁸ Bsp.: Kommunikationskompetenz (mündlich): »...fällt es mir leichter, meine eigenen Eindrücke/Meinungen zu äußern.« und »...fällt es mir leichter, nachzufragen, wenn ich etwas nicht verstanden habe.«

²⁹ überflüssiges Item gemäß BROWN, 2006, S. 203

³⁰ z.B. hohe standardisierte Residuen

³¹ Es handelte sich bei *EXP: Rohdaten darstellen* und *EXP: Daten interpretieren* um die Items »...erkenne ich Zusammenhänge/Muster innerhalb meiner gewonnenen Daten ("Zahlen") besser.« und »...erkenne ich Zusammenhänge/Muster in den grafischen Darstellungen meiner Daten besser.«

³² Für diese Fälle lagen jedoch stets noch vorhandene Items vor, die die inhaltliche Facette im Wesentlichen abdeckten.

Tabelle 12.13.: Konfirmatorische Analyse der Einzel-Messmodelle (je Faktor) des PraQ-A. Dargestellt sind fit-indices und Informationen zu den Items. Auf Grund von identischen wordings von Items wurden teils korrelierte Residuen zugelassen. Die modification-Indices führten weiterhin zum Ausschluss von insgesamt acht Items. Faktoren mit nur drei Items (siehe in Klammern) können mathematisch erst im Gesamtmodell geprüft werden ($df = 0$).

Faktor (Anzahl Items)	<i>SRMR</i>	<i>RMSEA</i>	<i>CFI</i>	CR	# dropped
Fachwissen: Theorie (3)	$df = 0$				1
Fachwissen: Praxis (3)	$df = 0$				1
EXP: Erwartungen formulieren (4)	.01	.04	1.00	1	1
EXP: Versuchsplanung & Geräte (5)	0.3	.08	.98		
EXP: Gerätekenntnisse (3)	$df = 0$				
EXP: Versuchsanordnung aufbauen (4)	.01	.08	.99		
EXP: Messungen durchführen (4)	$df = 0$			1	
EXP: Messungen dokumentieren (3)	$df = 0$				
EXP: Probleme lösen (4)	.01	.08	.99		
EXP: Rohdaten darstellen (3)	$df = 0$				3
EXP: Daten interpretieren (5)	.02	.06	.99		1
Kommunikationskompetenz: schriftlich (5)	.01	.06	1	1	
Kommunikationskompetenz: mündlich (6)	.02	.08	.98	1	
Bewertungskompetenz (5)	.01	.05	.99		
Kooperationskompetenz (4)	.01	.07	.99		1
Interesse (5)	.02	.10	.98		
Zeitmanagement (4)	.01	.08	.99		

L: alle Faktorladungen $> .50$ | CR: korrelierte Residuen zugelassen | # dropped: Anzahl ausgeschlossener Items | $df = 0$: Modell besitzt keine Freiheitsgrade, weshalb der Modellfit der Einzelmessmodells nicht berechnet werden kann | EXP: *Experimentelle Kompetenz* (PraQ)

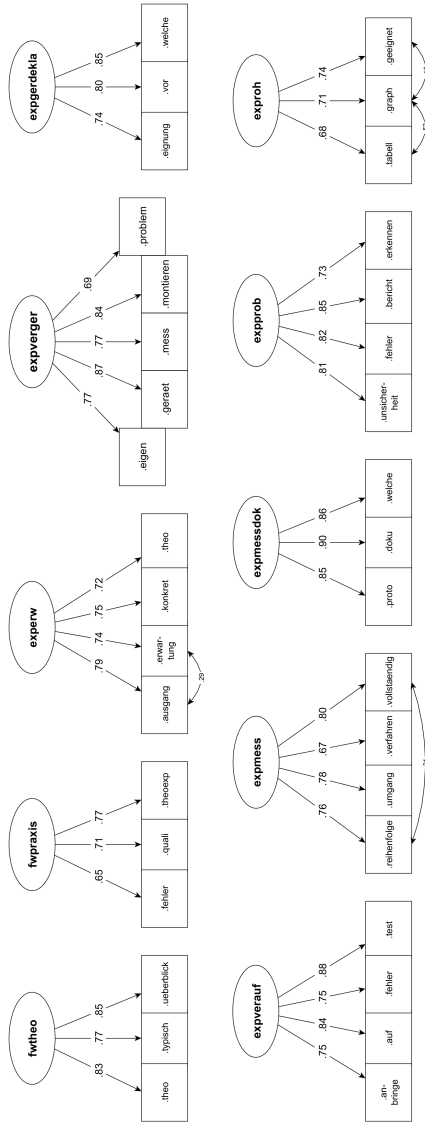


Abbildung 12.2.: Gesamtmodell zur CFA des PraQ-A (Auswahl). Korrelationen zwischen Faktoren werden zwecks Übersichtlichkeit nicht dargestellt.

fwtheo: Fachwissen Theorie; fwpraxis: Fachwissen Praxis; experw: Experimentelle Kompetenz: Erwartungen formulieren; expverger: Experimentelle Kompetenz Versuchsplanung & Geräte zusammenstellen; expgerdeka: Experimentelle Kompetenz: Gerätekenntnisse deklarativ; expwerauf: Experimentelle Kompetenz: Versuchsanordnung aufbauen; expmess: Experimentelle Kompetenz: Messungen durchführen; expmessdok: Experimentelle Kompetenz: Messungen dokumentieren; expprob: Experimentelle Kompetenz: mit Problemen und Fehlern umgehen; expproh: Experimentelle Kompetenz: Rohdaten graphisch tabellarisch darstellen; expdat: Experimentelle Kompetenz: Daten interpretieren; kommschrift: Kommunikationskompetenz schriftlich; kommuendl: Kommunikationskompetenz mündlich; bewfach: Bewertungskompetenz fachlich; koop: Kooperationskompetenz; intr: Interesse; persozeit: Personalkompetenz; Zeitmanagement

Itemformulierungen: vgl. Anhang A auf Seite 269

dass die gesamte *Praktikumsqualität* als einen Faktor annimmt, also sämtliche Items auf nur einem gemeinsamen Faktor laden. Zusätzlich wurden auf Einzelmessebene noch *EXP: Rohdaten darstellen* und *EXP: Daten interpretieren* als ein gemeinsamer Faktor modelliert, da diese die höchste Interfaktorkorrelation aufwiesen. Alle Modellvergleiche fielen gemäß χ^2 -Differenzentest und *CFI*-Differenzkriterium von $\Delta CFI \geq .01$ (CHEUNG & RENSVOED, 2002, S. 251) zu Gunsten der Grundmodelle aus, wie Tab. 12.14 genauer zeigt.

Empirische Trennbarkeit der Faktoren Eine empirische Trennbarkeit der Faktoren ist gemäß CHANG et al. (2011, S. 1227) bei Interfaktorkorrelationen kleiner .90 erfüllt. Im Falle des PraQ-A liegen die Interfaktorkorrelationen im Intervall [.25, .79], sind also allesamt positiv kleiner als .90 (Abb. 12.3). Die mittlere Interfaktorkorrelation beträgt .56.

Reliabilitätsanalyse Es wird der Reliabilitäts-Koeffizient RAYKOVs ρ (BROWN, 2006, S. 338) verwendet, da er messfehlerbereinigte Schätzungen der Reliabilität im Rahmen der CFA ermöglicht. Dieser ergibt sich aus der Summe der unstandardisierten Faktorladungen ($\sum \lambda_i$) und der Summe der unstandardisierten Fehlervarianzen ($\sum \theta_{ii}$):

$$\rho = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + \sum \theta_{ii}} \quad (12.1)$$

Werden korrelierte Residuen im Messmodell zugelassen, so ergibt sich die Änderung zu der Formel zu:

$$\rho = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + \sum \theta_{ii} + 2 \cdot \sum \theta_{ij}} \quad (12.2)$$

Hier wird also $\sum \theta_{ij}$ als Summe der Fehlerkovarianzen eingeführt, was bei Faktoren mit korrelierten Residuen automatisch zu einer Reduktion der Reliabilität führt.

Gemäß CHANG et al. (2011, S. 1227) sind Reliabilitätsmaße $> .60$ annehmbar im Rahmen einer CFA. Tab. 12.15 stellt die Reliabilitätsmaße dar.

Zu erkennen ist im Falle des PraQ-A, dass alle Reliabilitätsmaße im Normbereich liegen, lediglich die Skala *Experimentelle Kompetenz: Rohdaten graphisch, tabellarisch darstellen* fällt mit $\rho = .66$ etwas niedriger aus, was den Itemausschlüssen geschuldet sein kann.

Ergebnisse PraQ-B

Tabelle 12.14.: Modellvergleiche mittels χ^2 -Differenzentest und ΔCFI -Kriterium ($N = 576$).

	Grundmodell	Alternativmodell
	17-Faktoren-Modell	1-Faktor-Gesamtmodell
χ^2	4010.2	13879.7
df	2203	2339
$p_{diff} <$	.0001	
CFI	0.94	0.6
ΔCFI	0.34	
	9-Faktor-Modell-EXP	1-Faktor-Modell-EXP
χ^2	1025.2	4121.7
df	520	556
$p_{diff} <$	.0001	
CFI	0.96	0.73
ΔCFI	0.23	
	Daten interpretieren vs. Rohdaten darstellen	Fusionsmodell: Umgang mit Daten
χ^2	47.1	85.7
df	17	18
$p_{diff} <$	0.0001	
CFI	0.98	0.96
ΔCFI	0.02	

Die **fettgedruckten** Modelle sind in beiden Tests signifikant besser als ihr Pendant.

df : Freiheitsgrade

» $p_{diff} <$ «: χ^2 -Differenzentest für genestete Modelle

12. Testevaluation (Validität und Reliabilität)

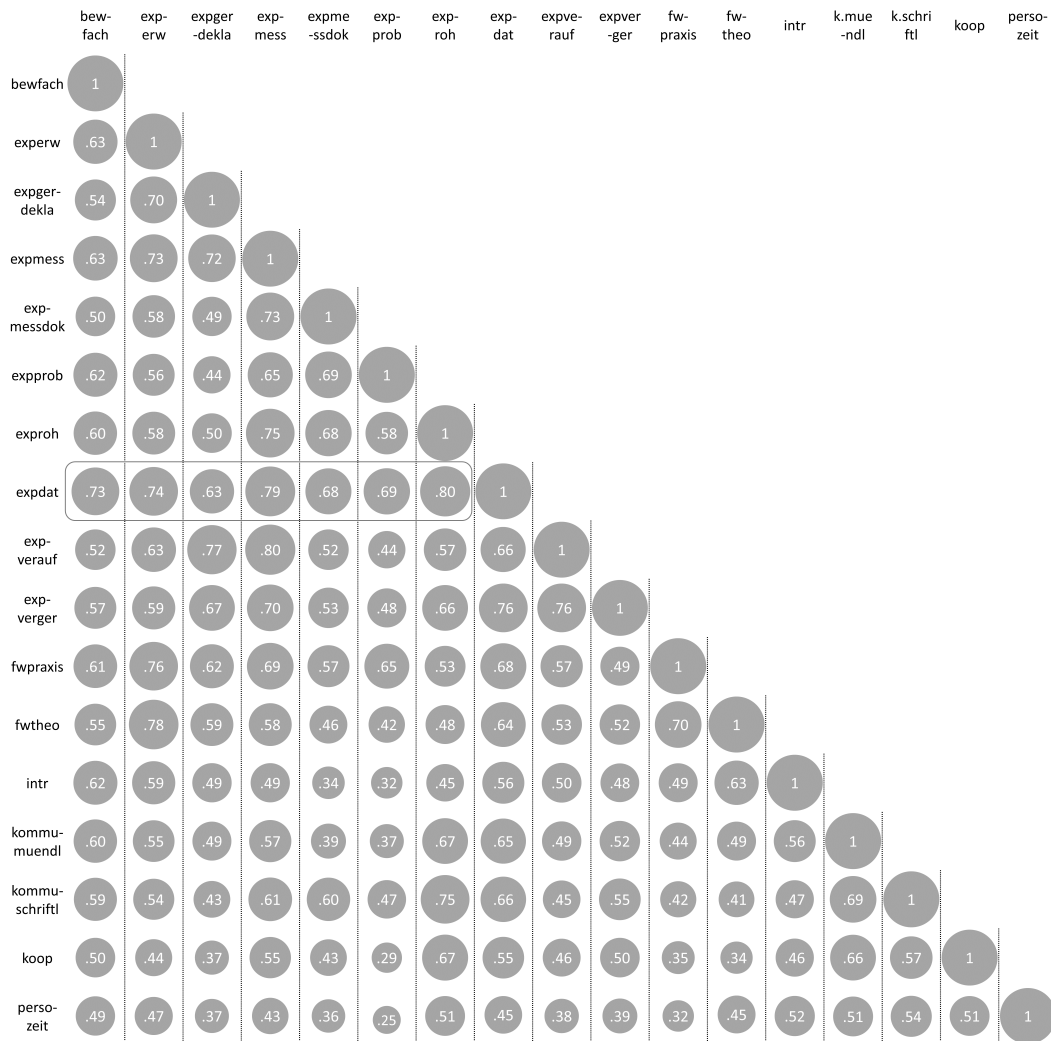


Abbildung 12.3.: Korrelationsdarstellung zu allen Faktoren des PraQ-A. Die grauen Kreisflächen sind vom Flächeninhalt proportional zum Korrelationskoeffizienten r , die Diagonalen mit $r = 1$ dienen als Referenz.
 $ps < .001$

fwtheo: Fachwissen Theorie; fwpraxis: Fachwissen Praxis; experw: Experimentelle Kompetenz: Erwartungen formulieren; expverger: Experimentelle Kompetenz: Versuchsplanung & Geräte zusammenstellen; expgerdekla: Experimentelle Kompetenz: Gerätekenntnisse deklarativ; expverauf: Experimentelle Kompetenz: Versuchsanordnung aufbauen; expmess: Experimentelle Kompetenz: Messungen durchführen; expmessdok: Experimentelle Kompetenz: Messungen dokumentieren; expprob: Experimentelle Kompetenz: mit Problemen und Fehlern umgehen; exproh: Experimentelle Kompetenz: Rohdaten graphisch tabellarisch darstellen; expdat: Experimentelle Kompetenz: Daten interpretieren; kommuschriftl: Kommunikationskompetenz schriftlich; kommue-ndl: Kommunikationskompetenz mündlich; bewfach: Bewertungskompetenz fachlich; koop: Kooperationskompetenz; intr: Interesse; persozeit: Personalkompetenz: Zeitmanagement

Tabelle 12.15.: Messfehlerbereinigter Reliabilitätsschätzer RAYKOV'S ρ mit Standardfehler je Skala. Für Faktoren mit zugelassenen korrelierten Residuen wurde die korrigierte Formel 12.2 verwendet, das Pendant zum Standardfehler blieben die Quellen schuldig (BROWN, 2006, S. 338; KLINE, 2011, S. 242).

Konstrukt	ρ	SE_{ρ}
Fachwissen (theoretisch)	0.86	0.01
Fachwissen (praktisch)	0.75	0.02
EXP: Erwartungen formulieren	0.82 [#]	
EXP: Versuchsplanung/Geräte	0.893	0.007
EXP: Gerätekenntnisse deklarativ	0.86	0.01
EXP: Versuchsanordnung aufbauen	0.906	0.006
EXP: Messungen durchführen	0.82	
EXP: Messungen dokumentieren	0.898	0.007
EXP: Probleme lösen	0.886	0.008
EXP: Rohdaten darstellen	0.66 [#]	
EXP: Daten interpretieren	0.894	0.007
Kommunikation (schriftlich)	0.927 [#]	
Kommunikation (mündlich)	0.882 [#]	
Bewertung fachlich	0.904	0.006
Kooperation	0.905	0.006
Interesse	0.88	0.008
Zeitmanagement	0.911	0.006

[#] ρ mit korrelierten Residuen: Angepasste Formel 12.2

EXP: *Experimentelle Kompetenz* (PraQ)

Umgang mit missings values Aufgrund der Tatsache, dass zwar viele aber nicht alle Praktika ein Skript einsetzten (in der vorliegenden Stichprobe waren es 96%) und dass es nicht immer eine Begleitvorlesung zum Praktikum gab (64%), wiesen die Skalen zur *Skriptqualität* und zur *Integration* eine vergleichsweise hohe Rate an fehlenden Werten auf, eine Imputation war nicht möglich.³³

Gleiches galt für das Item »Der/die Betreuende eröffnet den Studierenden Möglichkeiten, sich mit interessanten Sachverhalten eingehender zu beschäftigen.«, da bei diesem die »passt nicht«-Option mit 18% überdurchschnittlich oft genutzt wurde. Ein listenweiser Fallausschluss bezüglich dieser Items hätte hier einen großen Informationsverlust aller anderen Variablen zur Folge gehabt. Es wurde sich daher dafür entschieden, zweigleisig vorzugehen. Der »Primärdatensatz« enthält keine Items zur *Skriptqualität*, *Integration* und zur *Selbsttätigkeit anregen*, dafür aber eine Stichprobengröße von $N = 219$. Der »Sekundärdatensatz« wurde über einen listenweisen Fallausschluss derjenigen Proband*innen gebildet, die mindestens zwei fehlende Werte in der *Integrationsskala* (gesamt: 3 Items) oder mindestens drei fehlende Werte in der *Skriptskala* (gesamt: 13 Items) aufwiesen, dafür ergab sich aber nur eine Stichprobengröße von $N = 166$, die für ein umfangreiches Messmodell ggf. zu klein und stichprobenverzerrt bzgl. Praktika mit Skript und Begleitvorlesung ist.

Die Items »Der/die Betreuende eröffnet den Studierenden Möglichkeiten, sich mit interessanten Sachverhalten eingehender zu beschäftigen.« und »Das Skript macht den Stellenwert der Sachverhalte im Praktikum für das weitere Studium klar.« wiesen selbst in diesem Datensatz noch über 12% fehlende Werte auf, weswegen sie schlussendlich entfernt werden mussten. Dadurch entfiel die Skala zur *Selbsttätigkeit anregen* komplett, da nur noch ein Item verblieb. Die Modelle des PraQ-B wurden sowohl mit FIML, als auch mit multipler Imputation gerechnet, es ergaben sich keine nennenswerten Unterschiede in den Ergebnissen.

Einzelmessmodelle der Faktoren Im Unterschied zum PraQ-A lagen beim PraQ-B viele Skalen mit nur drei Items vor. Diese können nicht als Einzelmessmodelle geprüft werden (identifiziertes Modell, Eid, 2013, S. 867), sondern erst im Gesamtmodell. Dies betraf die Skalen *verständlich erklären*, *Verknüpfungen herstellen*, *Gute Lernatmosphäre herstellen*, *Relevanz verdeutlichen*, *Ermutigen und Selbstwirksamkeit stärken*, *Effizient mit Störungen umgehen*, *Diagnostik: Grundeinstellung*, *Diagnostik: situativ*, *Skriptqualität für Versuchsablauf* und *Integration der Vorlesungsinhalte*.

³³Für den Datensatz der Studie 2.2 ergaben sich für die Items der *Integration* durchschnittliche missing-Raten von 23%, für die *Skriptqualität* waren vier Items im Bereich [8%; 17%]. Sie konnten daher nicht ohne weiteres in die Imputation mitaufgenommen werden.

Die Skala *Skriptqualität* wurde mit dem Sekundärdatensatz (vgl. Umgang mit missing values) analysiert, weist daher ein geringeres N von 166 auf. Von den drei verbliebenen analysierbaren Einzelmessmodellen fällt der fit zur *Skriptqualität*, trotz notwendig befundener korrelierter Residuen, akzeptabel aus ($N = 219$, $SRMR = .04$, $RMSEA = .10$, $CFI = .94$). Die Einzelmessmodelle zu *verständlich erklären* ($N = 219$, $SRMR = .03$, $RMSEA = .13$, $CFI = .97$) und *Zusammenfassen und Wiederholen* ($N = 219$, $SRMR = .03$, $RMSEA = .10$, $CFI = .98$) fallen ebenso akzeptabel aus, da je nur ein Fitindex außerhalb des Normbereichs liegt. Der $RMSEA$ fällt allerdings für komplexe Modelle wie das vorliegende ohnehin tendenziell erhöht aus (BROWN, 2006, S. 84).

Gesamtmodell und Alternativmodelle Die Analyse des auf den Ergebnissen der EFA basierenden Gesamtmodells im Primärdatensatz ergab einen akzeptablen Fit ($N = 219$, $SRMR = .05$, $RMSEA = .06$, $CFI = .94$, vgl. Abb. 12.4), aber auch im Sekundärdatensatz unter Hinzunahme der Faktoren *Integration* und *Skriptqualität* war der Fit akkurat ($N = 166$, $SRMR = .05$, $RMSEA = .06$, $CFI = .91$).

Konkurrierende Modelle bezüglich des Gesamtmodells waren erstens das ABC-Modell, das drei Faktoren nach der Dimensionierung der *Lehrkompetenz* nach THIEL et al. (2012) plus die hinzugefügten zwei Faktoren zur *Diagnostik* und der *Skriptqualität* für die Durchführung annimmt.

Zweitens wurde auf Grund hoher Faktorkorrelationen im Gesamtmodell (.86) ein Modell vergleichend analysiert, das die Faktoren *Ermutigen und Selbstwirksamkeit stärken* und *Gute Lernatmosphäre herstellen* als gemeinsamen affektiven Faktor modelliert (Modell »SWE-förderliche Lernatmosphäre«).

Drittens wurde das plausible Alternativmodell gerechnet, dass es sich um einen einzigen Faktor der *Lehrkompetenz* handelt, statt mehrerer Einzelkompetenzen.

Alle Alternativmodelle wiesen einen schlechteren Modellfit auf und sind daher zu Gunsten der Grundmodelle abzulehnen (Tab. 12.16).

Im Sekundärdatensatz wurde das plausible Alternativmodell eines einzigen Konstrukts der *Skriptqualität* modelliert. Der Modellvergleich fiel zu Gunsten des getrennten Modells aus (vgl. Tab. 12.16).

Empirische Trennbarkeit der Faktoren Es ergaben sich bezüglich der Trennbarkeit der Faktoren zwei sehr hohe Korrelationen der Faktoren *verständlich Erklären* und *Zusammenfassen und Wiederholen* (.86) und *Gute Lernatmosphäre herstellen* und *Selbstwirksamkeitserwartung stärken* (ebenfalls .86), die in Form von Alternativmodellen untersucht wurden und empirisch trennbar waren. Die geringste

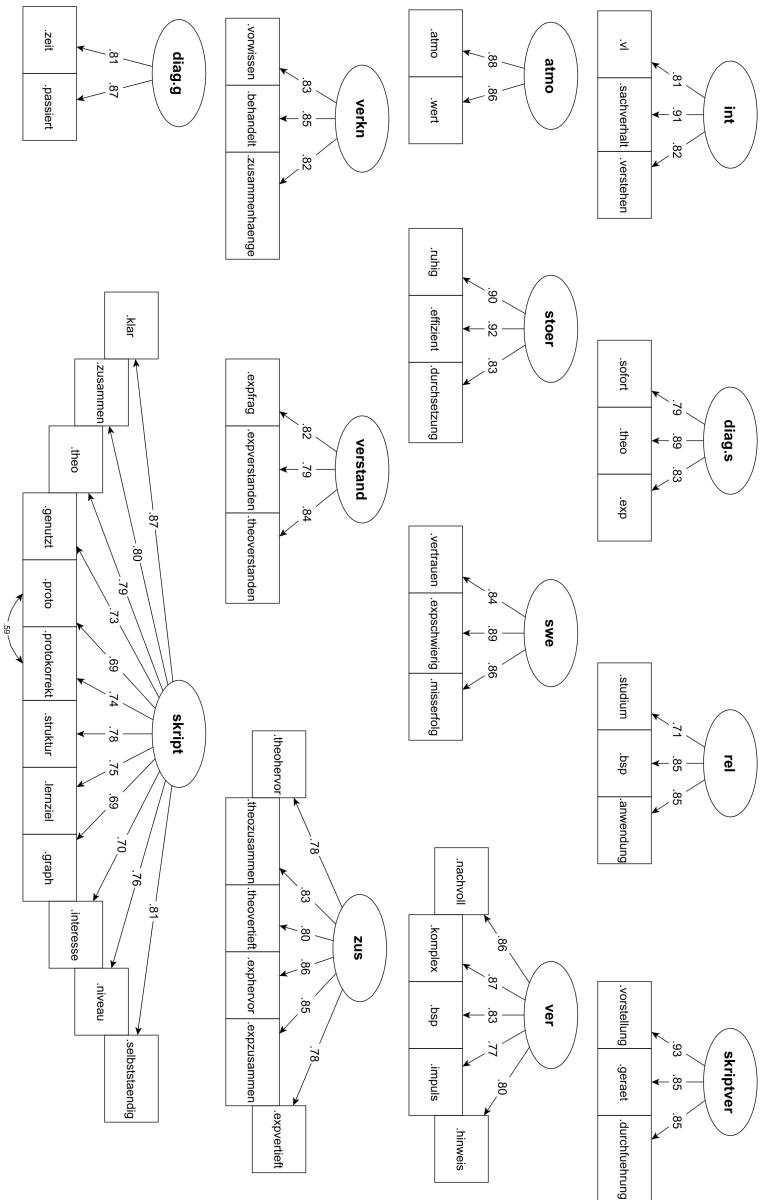


Abbildung 12.4.: Gesamtmodell zur CFA des PraQ-B. Korrelationen zwischen Faktoren werden zwecks Übersichtlichkeit nicht dargestellt.

ver: verständlich erklären; zus: Zusammenfassen & Wiederholen; verkn: Verknüpfungen herstellen; verstand: Verständnis überprüfen; atmo: Gute Lernatmosphäre herstellen; rel: Relevanz verdeutlichen; swe: Ermittigen und Selbstwirksamkeit stärken; stoer: Effizient mit Störungen umgehen; diag.g: Diagnostik: Grundeinstellung; diag.s: Diagnostik: Situativ; skriptver: Skriptqualität für Versuchsablauf; skript: Skriptqualität; int: Integration der Vorlesungsinhalte in das Praktikum
Itemformulierungen: vgl. Anhang A auf Seite 269

Tabelle 12.16.: Modellvergleiche zum Grundmodell. Sämtliche Alternativmodelle fallen signifikant schlechter im Modellfit aus und sind daher abzulehnen.

	Grundmodell	Alternativmodell 1	Alternativmodell2	Alternativmodell3
	Gesamtmodell	ABC-Modell	SWE-förderliche Lernatmosphäre	Lehrkompetenz
#	11	6	10	2
F	219	219	219	219
χ^2	935.03	1623	1021.3	2572.4
df	539	579	549	593
$p_{diff} <$		0.001	0.001	0.001
CFI	0.94	0.84	0.93	0.69
ΔCFI		0.1	0.01	0.25

	2-Faktoren-Skriptqualität ^a	Skriptqualität
N	166	166
χ^2	215.3	299.4
df	88	89
$p_{diff} <$		.001
CFI	.94	.9
ΔCFI		.04
Die fettgedruckten Modelle sind in beiden Tests signifikant besser als ihr Pendant.		
# F: Anzahl der modellierten Faktoren		
N: Stichprobengröße		
df: Freiheitsgrade		
» $p_{diff} <$ «: χ^2 -Differenzentest für genestete Modelle		

^avs. Skriptqualität für den Versuch

Korrelation wiesen die Faktoren *Zusammenfassen und Wiederholen* und *Skriptqualität für den Versuch* auf (.20). Überhaupt ist systematisch eine höhere Korrelation der Faktoren zu beobachten, die gemäß 3L-Modell in derselben Dimension (hier: *Lehrkompetenz* vs. *Lernumgebung*) verortet sind (Abb. 12.5). Die mittlere Korrelation der Faktoren des Primärdatensatzes beträgt $r = .55$ ($SD = .18$).

Reliabilitätsanalyse Die Reliabilität RAYKOV'S ρ der Faktoren des PraQ-B ist zufriedenstellend. Sie liegt im Bereich [.71; .93], mit der geringsten Reliabilität für *Verständnis überprüfen* und der höchsten für *Skriptqualität*. Tab. 12.17 stellt dies genauer dar.

12.4.2.4. Diskussion

Diskussion PraQ-A Nach der Auswertung können die statistischen Forschungsfragen für den PraQ-A beantwortet werden. Die Messmodelle der Einzelfaktoren (a) ergaben für alle beteiligten Konstrukte in den Bereichen *Fachwissen*, *Experimentelle Kompetenz*, *Kommunikationskompetenz*, *Bewertungskompetenz*, *Kooperationskompetenz* und *Personalkompetenz* für die Selbsteinschätzung der eigenen Kompetenz durch die Studierenden einen annehmbaren bis sehr guten Modellfit (Tab. 12.13), sofern dieser prüfbar war.

Für alle restlichen Faktoren und den fit der Messbeziehungen als Ganzes diente die Analyse des Gesamtmodells (b), die einen guten Fit lieferte, also das angenommene Messmodell und somit auch das theoretische 3L-Modell der Praktikumsqualität empirisch stützt, zudem sämtliche plausiblen Alternativmodelle signifikant schlechter als das Grundmodell ausfielen. Dabei ergab die Analyse der Korrelationen der Faktoren im Gesamtmodell, dass diese gering genug waren, dass die Annahme trennbarer Konstruktioninterpretationen (c) beibehalten werden darf.

Lediglich der Faktor *Rohdaten darstellen* der *Experimentellen Kompetenz* wurde mit drei Itemausschlüssen inhaltlich arg beschnitten und weist eine geringe Reliabilität auf. Dieses Konstrukt darf also vor dem Hintergrund der verbliebenen Items nur noch als ein »darstellen« der Rohdaten verstanden werden und nicht mehr als der gesamte Umgang mit Rohdaten. Die Reliabilitäten aller sonstigen Skalen fielen gut bis sehr gut aus, von Messgenauigkeit ist also für den PraQ-A auszugehen.

Diskussion PraQ-B Für den PraQ-B lässt sich bezüglich a) festhalten, dass die testbaren Messmodelle akzeptabel waren. Die Skalen zu *verständlich erklären* und *Zusammenfassen und Wiederholen* sind demnach vorsichtiger zu interpretieren. Durch die Notwendigkeit der Nutzung eines Sekundärdatensatzes sind die Ergebnisse der Skalen *Skriptqualität* und *Integration* vorsichtiger zu interpretieren, da sie

12.4. Studie 2: Strukturelle Eigenschaften (»faktorielle Validität«)

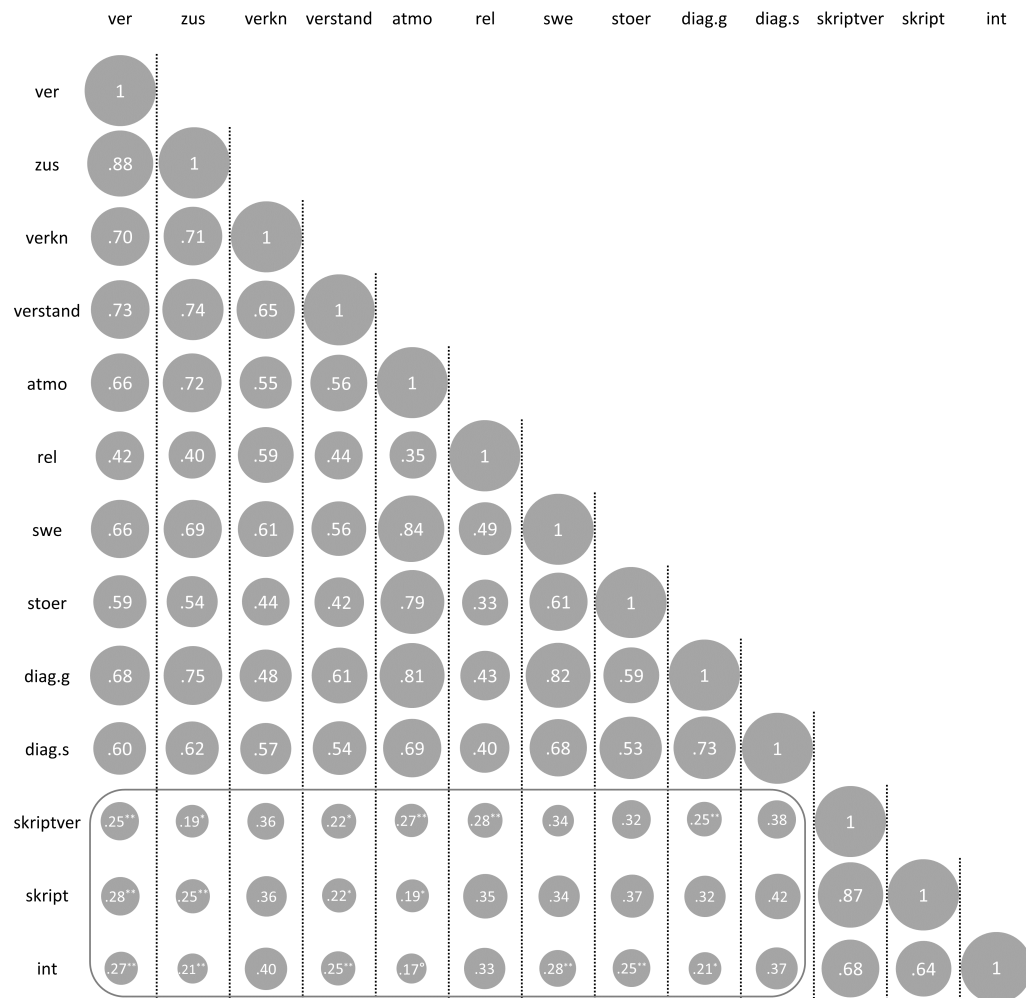


Abbildung 12.5.: Korrelationsdarstellung zu allen Faktoren des PraQ-B. Die grauen Kreisflächen sind vom Flächeninhalt proportional zum Korrelationskoeffizienten r , die Diagonalen mit $r = 1$ dienen als Referenz. Die Faktorkorrelationen zu *skript* und *int* stammen aus der Analyse des Sekundärdatensatzes und sind daher vorsichtiger zu interpretieren.

ver: verständlich erklären; zus: Zusammenfassen & Wiederholen; verkn: Verknüpfungen herstellen; verstand: Verständnis überprüfen; atmo: Gute Lernatmosphäre herstellen; rel: Relevanz verdeutlichen; swe: Ermutigen und Selbstwirksamkeit stärken; stoer: Effizient mit Störungen umgehen; diag.g: Diagnostik: Grundeinstellung; diag.s: Diagnostik: Situativ; skriptver: Skriptqualität für Versuchsablauf; skript: Skriptqualität; int: Integration der Vorlesungsinhalte in das Praktikum

* $p < .10$, ** $p < .05$, *** $p < .01$, sonst: $p < .001$

Tabelle 12.17.: Messfehlerbereinigter Reliabilitätsschätzer RAYKOVs ρ mit Standardfehler je Skala für den PraQ-B. Für Faktoren mit zugelassenen korrelierten Residuen wurde die korrigierte Formel 12.2 verwendet, das Pendant zum Standardfehler blieben die Quellen schuldig (BROWN, 2006, S. 338; KLINE, 2011, S. 242).

Konstrukt	ρ	SE_{ρ}
verständlich erklären	0.91	0.01
Zusammenfassen & Wiederholen	0.92	0.01
Verknüpfungen herstellen	0.87	0.01
Verständnis überprüfen	0.71	0.02
Gute Lernatmosphäre herstellen	0.86 ^α	0.02
Relevanz verdeutlichen	0.84	0.02
Ermutigen und Selbstwirksamkeit stärken	0.89	0.01
Effizient mit Störungen umgehen	0.92	0.01
Diagnostik: Grundeinstellung	0.83	0.02
Diagnostik: situativ	0.87	0.02
Skriptqualität für den Versuch	0.90	0.01
Skriptqualität	0.93 ^s	
Integration der Vorlesungsinhalte	0.87 ^s	0.01

^αCronbachs α (da nur 2 Items vorhanden)

^sauf Basis der Sekundärdatensatz berechnet ($N = 166$)

auf Basis eines kleineren und mit einer Stichprobenverzerrung versehenen Datensatzes analysiert wurden. Die Verzerrung betrifft die zu starke Betonung derjenigen Praktika, die ein Skript und eine Begleitvorlesung aufweisen. Bezüglich b) weist das Gesamtmodell sowohl im Primär- als auch im Sekundärdatensatz eine akzeptable Passung zu den Daten auf. Der Vergleich mit plausiblen Alternativmodellen fiel stets deutlich zu Gunsten der á priori antizipierten Struktur aus.

Die Struktur der PraQ-Items und Konstrukte, wie sie zunächst im 3L-Modell antizipiert und dann in der explorativen Analyse (Studie 2.1) spezifiziert wurde, ist also empirisch abgesichert. Die empirische Trennbarkeit der Konstrukte c) ist gemäß Korrelationsmaßen fast immer geglückt ($< .85$). Für die zwei Korrelationen zwischen .85 und .90 wurden Alternativmodelle getestet, die eine signifikant schlechtere Passung aufwiesen, daher abzulehnen waren. Die systematisch höheren Korrelationen für Faktoren innerhalb derselben Dimension des 3L-Modells beim PraQ-B stützen das 3L-Modell zudem empirisch, da dies ebenjene Dimensionen annimmt. So korrelieren Faktoren der *Lehrkompetenz* systematisch höher miteinander, als Faktoren der *Lernumgebungsqualität*. In Folge dessen kann bei der Interpretation der Messer-

gebnisse des PraQ-B konstruktgetrennt vorgegangen werden, was eine dezidiertere Sicht auf die Praktikumsqualität ermöglicht.

Dies kommt einer regulären Praktikumsevaluation zu Gute, bei der diese Trennbarkeit oft stillschweigend angenommen wird (Ziel 1). Es kommt aber ebenso der fachdidaktischen Forschung zu Gute, da nunmehr 13 faktoriell evaluierte Skalen des PraQ-B zur Verfügung stehen. Leider entfiel die Skala *zur Selbsttätigkeit anregen der Lehrkompetenz* komplett. Der PraQ-B erfasst also keinerlei Informationen mehr zu dieser aus konstruktivistischer Sicht wertvollen Skala. Eine Neukonstruktion in der Zukunft ist wünschenswert. Indes ist der Verlust des Items »Das Skript macht den Stellenwert der Sachverhalte im Praktikum für das weitere Studium klar.« geringfügig einzuschätzen. Dies liegt einerseits daran, dass die Skriptskala ohnehin mit vielen Items eine breite Konstruktabdeckung gewährleistet und andererseits daran, dass eine Betonung der Relevanz gemäß kognitiver Interviews mit Praktikumsleitenden (Studie 1.1) meist erst in F-Praktika forciert wird.

Fazit Die Antwort zur Forschungsfrage 3 lautet: Es lassen sich alle modellierten Konstrukte (vgl. Studien 2.1 und 2.2), außer dem entfernten Konstrukt *zur Selbsttätigkeit anregen* theoretisch wie empirisch getrennt voneinander interpretieren.

Einschränkungen im inhaltlichen Aspekt (vgl. auch Studie 1.1) ergaben sich für die Skala *Rohdaten darstellen*, da diese um drei Items beschnitten wurde. Für strukturelle Interpretationen ist bei den Skalen *verständlich erklären*, *Zusammenfassen und Wiederholen*, *Gute Lernatmosphäre herstellen* und *Ermutigen und Selbstwirksamkeit stärken* in der Dimension *Lehrkompetenz* Vorsicht geboten. Auf Seite der *Lernumgebung* sollten die Skalen zur *Skriptqualität* und *Integration* mit Obacht interpretiert werden.

Die Konstruktschätzungen ergaben sich als messgenau, wie die hohen Reliabilitätskoeffizienten nahelegen. Bezogen auf Ziel 1 bedeutet dies, dass Praktikumsleitende bei der Interpretation Ihrer Evaluationsergebnisse des PraQ die Konstrukte getrennt interpretieren dürfen. Auch können so gesicherter die Interventionen an eben jenen Konstrukten ansetzen, sofern ein Wirkmodell vorliegt (vgl. Abschnitt 14.4). Bezogen auf Ziel 1 lässt sich schlussfolgern, dass die PraQ-Skalen inhaltlich umfassend und strukturell adäquat ihre betreffenden Konstrukte messen, was sie bereits als Messinstrument für die fachdidaktische Forschung nutzbar macht. Einschränkend muss jedoch auf die physikstudierendenlastige Stichprobe verwiesen werden, die eine klare Absicherung von Interpretationen in Chemie- und sonstigen Praktika nicht gewährleisten konnte. Zumindest für Praktika der Chemie konnte eine Untersuchung der Messinvarianz vorgenommen werden (Studie 6). Durch Er-

hebungen in 18 Praktika im deutschsprachigen Raum ist eine Generalisierbarkeit dieser strukturellen Aussagen der Interpretation der PraQ-Messungen deutschlandweit naheliegend. Allerdings sollen die Testevaluationsstudien für den PraQ noch weitere Interpretationen absichern.

12.5. Studie 3: Zusammenhänge mit etablierten Maßen (»diskriminante und konvergente Konstruktvalidität«)

12.5.1. Studie 3.1: Leistungstest und Selbstwirksamkeitserwartung (»konvergent«)

12.5.1.1. Einleitung

In vorangegangenen Abschnitten (4.2, 4.5.2, 10.1) wurde bereits darauf hingewiesen, dass die Wahl der Erhebung von Praktikumsqualität mittels Selbst- und Fremdeinschätzungen kritisch sein kann. So gilt es nun im Zuge der Testevaluation, empirische Evidenz für oder gegen eine konstruktadäquate Interpretierbarkeit der PraQ-Messungen zu finden. Die AERA et al. fordern in ihren Standards Vergleiche mit ähnlichen Tests und Evidenz zu Zusammenhängen mit verwandten Konstrukten (AERA et al., 2014, S. 12, 16), in Studie 3.1 würde dies mit »konvergenter Evidenz« bezeichnet (ebd., S. 16). Wie in besagten Abschnitten erläutert, wird von anderen Autoren vermutet, dass Selbsteinschätzungen von Kompetenzentwicklungen eher Selbstwirksamkeitserwartungen bezüglich dieser Kompetenzen entsprechen könnten, was es zu prüfen gilt.

Zudem existieren zu einigen der PraQ-Skalen bereits analoge Leistungstests zur Erfassung von Kompetenzerwerb. So gibt es bereits Kompetenztests zur *Fachkompetenz* (RIESE & REINHOLD, 2009), *Experimentellen Kompetenz* bzw. *Erkenntnisgewinnungskompetenz* (HEIDRICH et al., 2013; STRAUBE & NORDMEIER, 2013), *Kommunikationskompetenz* (KULGEMEYER, 2010) und *Bewertungskompetenz* (EGGERT, 2008) von Schüler*innen und Studierenden. Die *Fachkompetenz* ist hierbei natürlich je inhaltsspezifisch modelliert und müsste auf seine Passung zu den diversen und je Praktikum unterschiedlichen Praktikumsthemen geprüft werden. Die *Kommunikationskompetenz* ist gemäß den Zielen nach HALLER (1999) und AAPT (1998), sowie LAGOWSKI (2002) als Nebenziel von Praktika anzusehen, ebenso wie die *Bewertungskompetenz*.

Kerngeschäft von Praktika (PSILLOS & NIEDDERER, 2002) ist die *Experimentelle Kompetenz*, weswegen diese Skalen des PraQ-A in der folgenden Studie auf Zusammenhänge sowohl mit einem etablierten Kompetenztest, als auch mit einer etablierten Skala zur *experimentenspezifischen Selbstwirksamkeitserwartung* verglichen wird.

Innerhalb der anderen beiden Dimensionen des 3L-Modells sieht es indes bezüglich Vergleichsmaßen schwierig aus. Weder existiert ein Leistungstest zur *Lehrkompetenz* experimenteller Lernumgebungen, noch ein objektiver Bewertungsmaßstab für Praktikumsdokumente. Für letzteres liefert aber die Studie 4 (Abschnitt 12.6) ein erstes Kategoriensystem.

Forschungsfragen und Hypothesen Die leitende Forschungsfrage lautet:

Forschungsfrage 4: Wie hängen die Skalen des PraQ statistisch mit Kompetenz-Leistungstests und Selbstwirksamkeitserwartung zusammen? (»Konvergente Konstruktvalidität«)

Da die PraQ-Skalen anders als andere Selbsteinschätzungsskalen kompetenzorientiert und als Veränderungseinschätzungsmaß entwickelt wurden (vgl. BRAUN et al., 2008), wird ein mittlerer korrelativer Zusammenhang zwischen den Ergebnissen der Messung mit den PraQ-Skalen zur *Experimentellen Kompetenz* und der Messung mit einem Kompetenztest erwartet.

Hypothese 1: Die Korrelation zwischen der Messung mit einem Leistungstest zur Erkenntnisgewinnungskompetenz und den PraQ-Skalen zur *Experimentellen Kompetenz* wird mittelstark ausfallen.

Der Vermutung mancher Autor*innen, dass die Selbsteinschätzung des eigenen Kompetenzzuwachses einer *Selbstwirksamkeitserwartung* gleicht, folgend, wird sich ebenso ein mittlerer Zusammenhang mit einer Skala zur *experimentenspezifischen Selbstwirksamkeitserwartung* zeigen.

Hypothese 2: Die Korrelation zwischen der Messung der *experimentbezogenen Selbstwirksamkeitserwartung* und den PraQ-Skalen zur *Experimentellen Kompetenz* wird mittelstark ausfallen.

Da der ausgewählte Leistungstest eine ähnliche Modellgrundlage aufweist wie der PraQ, kann hierfür keine gerichtete Hypothese je einzelner PraQ-Skala aufgestellt werden. Die eingesetzte Skala zur *experimentenspezifischen Selbstwirksamkeitserwartung*³⁴ enthält hauptsächlich Items zur Durchführung des Experiments, weshalb hierfür eine vergleichsweise höhere Korrelation mit PraQ-Skalen zur Durchführungsphase von Experimenten erwartet werden kann:

³⁴Die Skala ist auf Anfrage bei Stefan Schroedter (stefan.schroedter@s-h-online.de) zu erhalten.

Hypothese 3: Die Korrelation zwischen der *experimentsspezifischen Selbstwirksamkeitserwartung* und den PraQ-Skalen zur Durchführungsphase von Experimenten (*Versuchsanordnung aufbauen, Messungen durchführen, Messungen dokumentieren, mit Problemen und Fehlern umgehen*) wird vergleichsweise höher ausfallen.

12.5.1.2. Methode

Stichprobe Zur Untersuchung eines Kompetenzzuwachses sollte eine »Intervention« stattgefunden haben, die diesen Kompetenzzuwachs bewirkt hat. Im Zuge der *Experimentellen Kompetenz* boten sich hierzu Grundpraktika der Naturwissenschaften, hier Physik, an. Diese erzielten, mit Vorgriff auf Ergebnisse aus der Studie 5, in den PraQ-Skalen der *Experimentellen Kompetenz* Mittelwerte zwischen 3.5 ($SE = 0.2$) und 4.9 ($SE = 0.2$) auf der Skala von 1 bis 6, ein Zuwachs an experimenteller Kompetenz liegt also gemäß Selbsteinschätzung vor. Die Stichprobe bestand aus Teilnehmenden von zwei Grundpraktika der Physik (Universität Wien und Freie Universität Berlin) im Sommersemester 2015.

Die Erhebungen an der Freien Universität Berlin wurden kooperativ im Rahmen einer Masterarbeit (KRATSCHMAR, 2017) unter Anleitung durch den Autor vorgenommen.³⁵ Es konnte insgesamt eine Stichprobe von $N_{PRE} = 100$ und $N_{POST} = 91$ gewonnen werden, nach dem matching lag ein $N = 72$ vor (ca. 20% Dropout). Es handelte sich um Physikstudierende und Physik-Lehramtsstudierende im Grundpraktikum, 49 davon im zweiten Fachsemester studierend, 71% männlich, über 95% mit einer Schulabschlussnote in Physik besser 3.

Instrumente Für die Erhebung der Experimentellen Kompetenz wurde sich für den Ko-WADiS-Test (STRAUBE, 2016; vgl. Anhang H auf Seite 293) entschieden. Dieser zeichnet sich dadurch aus, dass er ökonomisch als Paper-Pencil-Test eingesetzt werden kann und ermöglichte daher einen analogen Befragungsmodus wie beim PraQ.³⁶ Zudem sind die theoretischen Modelle *experimenteller Kompetenz* des 3L-Modells (SCHREIBER, 2012) und der *Erkenntnisgewinnungskompetenz* hinter dem Ko-WADiS-Test zumindest in Anteilen vergleichbar.

So umfasst der Anteil des Ko-WaDiS-Kompetenzstrukturmodells (STRAUBE, 2016, S. 40) der »Naturwissenschaftlichen Untersuchungen« nach MAYER (2007, S. 181) die vier Prozeduren *Naturwissenschaftliche Fragen formulieren, Hypothesen gene-*

³⁵Dabei war das Praktikum an der Universität Wien ein regulärer Semesterkurs, an der Freien Universität handelte es sich um ein Blockpraktikum zwischen dem Sommersemester 2015 und dem Wintersemester 2015/2016 (vgl. Studie 7, Abschnitt 13.3).

³⁶Auch Erhebungen mit »hands-on-Tests« wie bei HEIDRICH et al. (2013) wären denkbar gewesen, allerdings ist speziell dieser nur auf den inhaltlichen Bereich Optik beschränkt.

rieren, Untersuchungen planen und Daten analysieren/Schlussfolgerungen ziehen. Beim Modell SCHREIBERS kann dies mit den Konstrukten *Fragestellung entwickeln, Hypothese bilden, Versuchsplan entwerfen* und *Messdaten aufbereiten/Ergebnis interpretieren* assoziiert werden. Nach MESSICK (1995) entspricht dies theoretical evidence für eine Ähnlichkeit der Modelle und damit der zu messenden Konstrukte. Aus Gründen der sprachlichen Vereinfachung wird daher der Ko-WaDiS-Test nunmehr als Test zur Erfassung experimenteller Kompetenz bezeichnet.

Nun musste aus den Items des Ko-WADiS-Tests in der Physik eine Itemauswahl getroffen werden, die vor allem jene, assoziierte Konstrukte, umfasste. Zudem sollte die von den Praktikumsleitenden vorgeschriebene Erhebungszeit von 15min eingehalten werden, was im Zuge von Messungen mittels Kompetenztest wenig Zeit entspricht. Im Zuge dessen wurde sich für das Implementieren von je Messzeitpunkt sechs Ko-WADiS-Items entschieden, wobei drei Items als Ankeritems dienten, Erinnerungseffekte an die Anker-Items wurden als gering eingestuft, da zwischen PRE- und POST-Messung drei Monate vergingen. Die Ko-WADiS-Items zum *Hypothesen prüfen* wurden nicht implementiert, da dieser Inhaltsbereich auch nicht Bestandteil der finalen PraQ-Skalen zur Experimentellen Kompetenz war. Die restlichen drei Dimensionen *Frage, Planung* und *Auswertung* des Modells MAYER (2007) wurden berücksichtigt.

Bei der Itemauswahl wurde zudem darauf geachtet, die Itemschwierigkeiten heterogen zu wählen und die POST-Items etwas schwieriger zu wählen als die PRE-Items, da durch das Praktikum eine Steigerung der Kompetenz zu erwarten war. Die größte Konzentration an Items galt der mittleren Itemschwierigkeit 0 (RASCH-Modell, z. B. BOND & FOX, 2007), da die meisten Proband*innen auf diesem Niveau erwartet werden können.

Für die Erhebung der *experimentsspezifischen Selbstwirksamkeitserwartung* wurde die Skala nach SCHROEDTER & KÖRNER (2011) verwendet. Als Begleitvariablen wurden standardmäßig das Geschlecht, Alter, Fachsemester, Hochschulzugangsberechtigungs- und Physiknote³⁷, Zweitfach der Studierenden und Leistungskursabfrage erhoben. Hinzu kam eine Skala zur *sozialen Erwünschtheit*, um die Effekte der Selbsteinschätzung genauer untersuchen zu können. Zusätzlich wurde analog zu anderen PraQ-Messungen nach der bisher absolvierten Praktikumsdauer, und dem *Vorwissen im Experimentieren* gefragt.

Auswertungsmethode Klassischerweise wird konvergente Evidenz über Korrelationsmaße empirisch untersucht (MESSICK, 1995, S. 746). Ergänzend und um für

³⁷ Indikatoren für kognitive Leistungsfähigkeit (MAYER, 2007, S. 179).]

den Einfluss von Begleitvariablen kontrollieren zu können, wurde zudem eine multiple Regression vorgenommen. Auf messfehlerbereinigte Analysen im Rahmen von Strukturgleichungsmodellen (vgl. Studien 2.2, 3.2 und 5–7) musste wegen der geringen Stichprobengröße verzichtet werden.

Gemäß Hypothese 1 und 2 wird ein »mittlerer Zusammenhang« erwartet, der hier auch über eine mittelgroße Korrelation ($r > .30$, EID, 2013, S. 508) statistisch gefasst wird. Die Korrelation der Maße der PraQ-Skalen zur *Experimentellen Kompetenz* mit Maßen der *Experimentellen Kompetenz* (Ko-WADiS) und der *experimentierspezifischen Selbstwirksamkeitserwartung* wurde zuvor in ihrer erforderlichen Stichprobengröße mit einer Poweranalyse abgesichert. Gemäß G*Power (FAUL et al., 2009) wird für das Auflösen einer mindestens mittelgroßen Korrelation von .30 eine Stichprobe von mindestens $N = 65$ benötigt, was in dieser Studie realisiert werden konnte (vgl. Stichprobe).

Design & Durchführung Das Design der Studie ist Abb. 12.6 zu entnehmen. Darin ist zu erkennen, dass die Instrumente zur Erfassung der *Experimentellen Kompetenz* (Ko-WADiS) und der *experimentierspezifischen Selbstwirksamkeitserwartung* sowohl zu Beginn (PRE), als auch zu Ende des Praktikums (POST) erhoben wurden. Diese beiden Messzeitpunkte dienten in der Auswertung der Berechnung eines Zuwachsmäßes für beide Konstrukte durch Subtraktion der PRE-Werte von den POST-Werten. Das Praktikum steht als für diesen Zuwachs verantwortliche Intervention zwischen den Messzeitpunkten. Die Messung mit PraQ erfolgt am Ende des Semesters.

Die Durchführung der Studie erfolgte an den genannten Universitäten. Die Erhebungen an der Freien Universität wurden PRE in den Einführungsveranstaltungen zur Pause vor der Sicherheitsbelehrung und POST im Rahmen der Abschlussveranstaltungen durchgeführt. Die Erhebungen an der Universität Wien fanden ebenso in den ersten respektive letzten Veranstaltungswochen statt.

12.5.1.3. Ergebnisse

Die Ergebnisse fasst Tab. 12.18 zusammen. Aus dieser ist zunächst abzulesen, dass es schwache bis mittlere Korrelationen zwischen den Maßen der PraQ-Skalen und dem Zuwachs an *experimenterspezifischer Selbstwirksamkeitserwartung* gab. Diese fielen für die Skalen *Daten interpretieren*, *Erwartungen formulieren* und *Mit Problemen und Fehlern umgehen* mittelstark aus (gerundet um oder über .30) und für die Skalen *Versuchsanordnung aufbauen*, *Messungen durchführen* und *Messungen dokumentieren* gering aus.

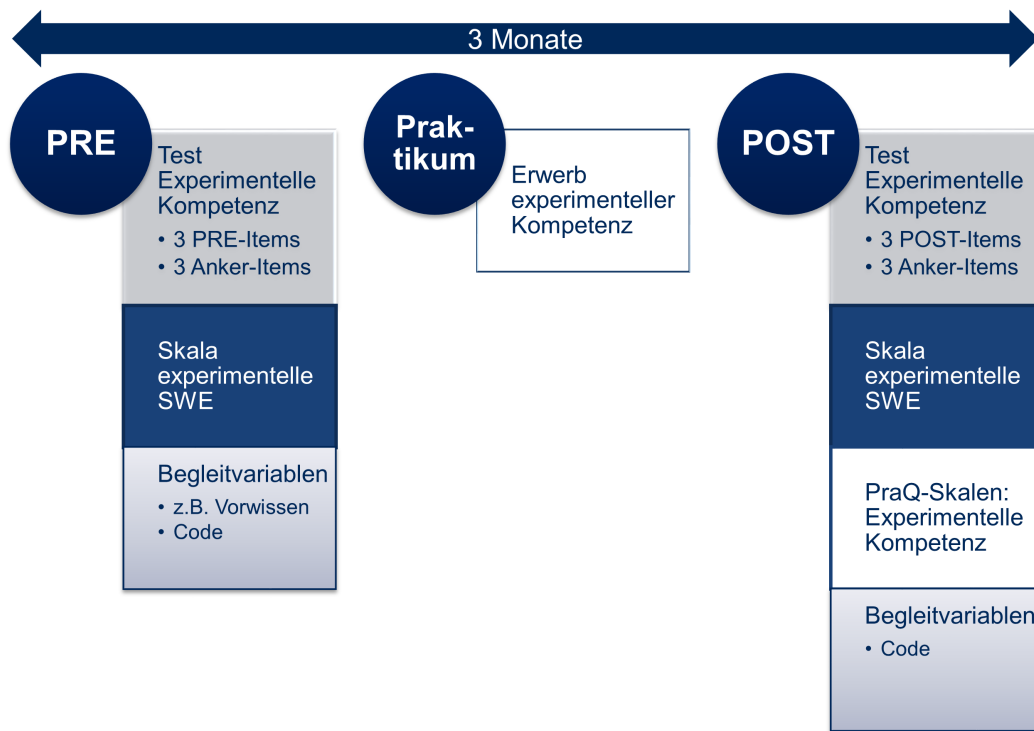


Abbildung 12.6.: Erhebungsdesign der Studie 3.1: Die *Experimentelle Kompetenz* (Ko-WADiS, STRAUBE, 2016) und die *experimentierspezifische Selbstwirksamkeitserwartung* (SCHROEDTER & KÖRNER, 2011) werden im PRE-POST-Design zur Veränderungsmessung erhoben. PraQ muss nur POST erhoben werden, da es nativ eine Veränderungsmessung beinhaltet. Die »Intervention« stellt hier das Grundpraktikum dar.

Die Korrelationen mit dem Zuwachs an *Experimenteller Kompetenz* gemäß Ko-WADiS fielen fast alle nichtsignifikant aus. Es ergab sich lediglich eine schwache, negative Korrelation mit der Skala *Rohdaten darstellen*.

Ergänzende Analysen mit multipler Regression ergaben auch unter Berücksichtigung der Einflüsse der genannten Begleitvariablen keine weiteren signifikanten Zusammenhänge zwischen der Ko-WADiS-Messung und den PraQ-Messungen.

12.5.1.4. Diskussion

Die Ergebnisse der Studie fallen zu einem guten Teil nicht konform zu den Hypothesen 1 und 2 aus, aber recht konform zu Hypothese 3. Für den Zuwachs an *Selbstwirksamkeitserwartung* bezüglich des Experimentierens durch die Intervention ergaben sich nur bei sechs von neun Skalen des PraQ signifikante Zusammenhänge. Diese fielen meist hypothesenkonform, also im Bereich von mittleren Korrelationen, aus. Wie in Hypothese 3 vermutet ergaben sich signifikante Korrelationen hauptsächlich zu den PraQ-Skalen zur Durchführungsphase (vier von sechs Korrelationen).

Hypothesenkonforme Zusammenhänge mit dem Zuwachs der *Experimentellen Kompetenz* nach Ko-WADiS konnten indes für keine der PraQ-Skalen gefunden werden, was zwei Schlüsse zulässt.

Entweder liegt tatsächlich kein Zusammenhang zwischen der Messung des Zuwachses *Experimenteller Kompetenzen* durch den Kompetenztest und der Messung von Selbsteinschätzungen zum Zuwachs *experimenteller Kompetenzen* durch PraQ vor. Oder eines der beiden Maße weist eine sehr niedrige Reliabilität auf, d. h. die den Proband*innen zugewiesenen Ausprägungen sind zum großen Teil zufälliger Natur. Letzteres würde das Vorliegen von fast keinen Korrelationen erklären, da Zufallszahlen per se nicht korrelieren.

Das Ergebnis und die letztere Überlegung führten zu einer Rückschau bezüglich der eingesetzten Instrumente. Rein argumentativ scheinen sowohl die PraQ-Maße als auch die *Selbstwirksamkeitserwartungs*-Maße solide auszufallen, da sie hypothesenkonform miteinander korrelieren.³⁸

Dies legt die Vermutung nahe, dass in der Erhebung mit Ko-WADiS Messfehler unterlaufen sein könnten. Und tatsächlich ergaben Diskussionen mit dem statistischen Berater des Ko-WADiS-Projektes³⁹, dass der Einsatz von zweimal sechs Items mit enthaltenen drei Ankeritems zu nicht optimalen Schätzungen der Personfähigkeiten führt. Allein die geringe Anzahl der Items sei in Kombination mit

³⁸ Allerdings kann dies auch ein Effekt der gemeinsamen Erhebungsmethode, hier Rating-Skala, sein.

³⁹ Persönliche Kommunikation des Autors mit Stefan Hartmann. Berlin, April, 2016.

Tabelle 12.18.: Korrelationen der Ausprägungen der PraQ-Skalen zur *Experimentellen Kompetenz* mit dem Kompetenzzuwachs laut Ko-WADiS-Test und dem Zuwachs an *experimentierspezifischer Selbstwirksamkeitserwartung*.

	$\Delta Komp_{EXP}$	ΔSWE_{EXP}
EXP: Daten interpretieren		.30*
EXP: Erwartungen formulieren		.28*
EXP: Versuchsplanung/Geräte		
EXP: Gerätekenntnisse deklarativ		
EXP: Versuch aufbauen		.22°
EXP: Messungen durchführen		.24°
EXP: Messungen dokumentieren		.21°
EXP: Probleme lösen		.34**
EXP: Rohdaten darstellen	-.26*	

EXP: *Experimentelle Kompetenz* (PraQ)

$\Delta Komp_{EXP}$: Kompetenzzuwachs *Erkenntnisgewinnung*

ΔSWE_{EXP} : Zuwachs an *experimentierspezifischer Selbstwirksamkeitserwartung*

** $p < .01$, * $p < .05$, ° $p < .10$, Leerstellen: nichtsignifikant

der verhältnismäßig kleinen Stichprobe dafür verantwortlich, dass keine verlässliche Schätzung vorgenommen werden könne. Die geringe Itemanzahl war auf Grund der Einschränkungen der Erhebungszeit durch die Praktikumsleitenden zu Stande gekommen.

Empfohlen wird zukünftig eine neue Studie unter Erhöhung der Itemanzahl und das »ankern« sämtlicher Items, um eine direkte Vergleichbarkeit der Leistungen vornehmen zu können.

Zudem werden auch aktuell weitere Testevaluationsstudien zum Ko-WADiS-Test durchgeführt. Hierzu ergaben sich nach Itemanalysen Vermutungen, dass dieser eher das Konstrukt »naturwissenschaftliches Denken« tangiert (Projekt »ValiDiS«, KRÜGER et al., 2017).

12.5.2. Studie 3.2: Praktikumssozialklima (»diskriminant«)

12.5.2.1. Einleitung

Die Untersuchung auf diskriminante Evidenz (AERA et al., 2014, S. 16) beinhaltet den Vergleich mit Konstruktmaßen, die unerwünschten Einfluss auf die zu evaluierende Messung haben können. Im Zusammenhang mit Lehrevaluationen wurde in Abschnitt 12.1 argumentiert, dass das *Studiensozialklima* in der betreffenden Lehrveranstaltung einen großen Einfluss auf die Evaluation haben kann, der der adäquaten Interpretation der Messwerte entgegenwirkt.

12. Testevaluation (Validität und Reliabilität)

Es geht darum, möglichst objektiv überlegte, statt subjektiv im Affekt getätigte Aussagen zur Evaluation zu erhalten. Dies legt einen Vergleich der PraQ-Skalen mit einer Skala zum *Studiensozialklima* im Praktikum, kurz *Praktikumsklima*, nahe.

Forschungsfragen und Hypothesen Die Forschungsfrage dieser Studie lautet:

Forschungsfrage 5: Wie hängen die Skalen des PraQ statistisch mit der Messung von allgemeinem Praktikums(sozial)klima zusammen? (»diskriminante Konstruktvalidität«)

Durch die kleinschrittige und handlungsbezogene Operationalisierung von Praktikumsqualität im PraQ sind die Einflüsse affektiver Variablen, hier des Praktikums-klimas, lediglich gering bis mittelstark anzunehmen:

Hypothese 1: Die Korrelationen der Messung des Praktikumssozialklimas und der PraQ-Skalen fallen gering bis mittelstark aus.

Für stärker affektiv geprägte Konstrukte wird eine vergleichsweise höhere Korrelation angenommen. Dies betrifft die Konstrukte *Interesse* des PraQ-A und *gute Lernatmosphäre herstellen*, *Ermutigen und Selbstwirksamkeit stärken* und *Effizient mit Störungen umgehen* des PraQ-B.

Hypothese 2: Für stärker affektiv geprägte Konstrukte wird eine vergleichsweise höhere Korrelation angenommen. Dies betrifft die Konstrukte *Interesse* des PraQ-A und *gute Lernatmosphäre herstellen*, *Ermutigen und Selbstwirksamkeit stärken* und *Effizient mit Störungen umgehen* des PraQ-B.

12.5.2.2. Methode

Stichprobe Die Stichprobe entspricht der Stichprobe aus der Studie 2.2. Die Skala zum *Praktikumsklima* wurde dort je in den regulären PraQ-Fragebogen integriert.

Instrumente Als Maß für das *Praktikumsklima* wurde die Skala *Studienklima* nach THIEL ET AL. (2010) für Praktika leicht adaptiert, indem je Item die Formulierung »im Praktikum« ergänzt wurde.⁴⁰ Es handelt sich um eine Operationalisierung als »soziales Klima im Praktikum«. Darin findet sich der zentrale Aspekt der Zufriedenheit mit der Integration in die peergroup mit Studierenden/Lehrenden und die Integration in die Institution wider.

⁴⁰Das erste Item der Skala wurde, wie bei THIEL, 2010, geschehen, ausgeschlossen.

Beispielitem 1: Es ist schwierig, Kontakte zu Kommilitonen im Praktikum zu finden.

Beispielitem 2: Es herrscht ein großer Wettbewerb zwischen den Studierenden im Praktikum.

Parallel wurde je der in Studie 2.2 beschriebene PraQ-A, respektive PraQ-B eingesetzt.

Auswertungsmethode Analog zur Studie 3.1 wird klassisch über Korrelationsmaße bestimmt, wie stark der Zusammenhang zwischen den Ergebnissen der PraQ-Messungen und der *Praktikumsklima*-Messungen ist. Damit Modellprüfungen möglich werden, wird dies hier über eine zu Studie 2.2 identische CFA vorgenommen, in der das *Praktikumsklima* als weiterer Faktor in das Modell miteinbezogen wird. Die latenten Korrelationsschätzer ergeben sich dann aus der Auswertung dieses Modells.

Erhebung Die Erhebung entspricht der Erhebung in Studie 2.2 unter Ergänzung der Skala zum *Praktikumsklima*.

12.5.2.3. Ergebnisse

PraQ-A Die CFA unter Einbezug des *Praktikumsklimas* ergab für das Messmodell der Faktoren des PraQ-A einen akzeptablen Modellfit ($N = 489$, $SRMR = .04$, $RMSEA = .04$, $CFI = .93$). Die latenten Korrelationen der PraQ-A-Faktoren mit dem *Praktikumsklima* fallen niedrig ($|r| < .08$ für alle nichtsignifikanten Korrelationen) bis mittel ($r = .20$ für *Fachwissen: Praxis*) aus (vgl. Tab. 12.19).

PraQ-B Gemäß Auswertungsmethode wurde sowohl mit dem Primär- als auch dem Sekundärdatensatz (Studie 2.2) eine CFA unter Einbezug des *Praktikumsklimas* vollzogen. Der Modellfit für die Auswertung mit Primärdatensatz war zufriedenstellend ($N = 166$, $SRMR = .05$, $RMSEA = .06$, $CFI = .93$), mit Sekundärdatensatz nicht zufriedenstellend ($N = 130$, $SRMR = .06$, $RMSEA = .07$, $CFI = .88$). Die latenten Korrelationen fallen niedrig ($|r| < .17$ für nichtsignifikante Korrelationen) bis mittelstark ($r = .43$ für *verständlich erklären*) aus (vgl. Tab. 12.20).

12.5.2.4. Diskussion

Die Studie 3.2 konnte zeigen, dass der statistische Zusammenhang der Maße des *Praktikumssozialklimas* mit den PraQ-Faktoren hypothesenkonform nur gering bis mittelstark ausfällt. Dies kann so gedeutet werden, dass die Interpretation von Maßen der PraQ-Skalen nur wenig von dem affektiv geprägten *Praktikumssozialklima* beeinflusst wird. Es ist daher als empirisch gestützt anzusehen, dass Interpretatio-

12. Testevaluation (Validität und Reliabilität)

Tabelle 12.19.: Latente Korrelationen r des Praktikumssozialklimas mit den Faktoren des PraQ-A.

	Praktikumssozialklima
EXP: Daten interpretieren	.13*
EXP: Erwartungen formulieren	.19***
EXP: Versuchsplanung/Geräte	
EXP: Gerätekenntnisse deklarativ	.09*
EXP: Versuch aufbauen	
EXP: Messungen durchführen	.10*
EXP: Messungen dokumentieren	.16*
EXP: Probleme lösen	.16*
EXP: Rohdaten darstellen	
Fachwissen (theoretisch)	.18***
Fachwissen (praktisch)	.20***
Kommunikation (mündlich)	
Kommunikation (schriftlich)	
Bewertung innerfachlich	
Kooperation	
Interesse	
Zeitmanagement	

*** $p < .001$, * $p < .05$, Leerstellen: nichtsignifikant

EXP: *Experimentelle Kompetenz* (PraQ)

Tabelle 12.20.: Korrelation des Praktikumssozialklimas mit den PraQ-Faktoren des PraQ-B. Die Korrelationen zur *Skriptqualität allgemein* und *Integration* basieren auf Analysen des Sekundärdatensatzes.

	Praktikumssozialklima
Verständlich erklären	.43***
Zusammenfassen/Wiederholen	.36***
Verknüpfungen herstellen	.37***
Verständnis überprüfen	.25**
Lernatmosphäre herstellen	.28**
zur Selbsttätigkeit anregen	
Relevanz verdeutlichen	.24**
Ermutigen/SWE stärken	.29***
Effizient mit Störungen umgehen	.38***
Diagnostik: Grundeinstellung	.33***
Diagnostik: Situativ	.33***
Skriptqualität für den Versuch	.17*
Skriptqualität allgemein	
Integration der Vorlesung	.23*

*** $p < .001$, * $p < .05$, Leerstellen: nichtsignifikant

12.6. Studie 4: Zusammenhang mit Expert*innenratings (»Multiperspektivität«)

nen auf Basis einer Praktikumsevaluation mittels PraQ arm an affektiv-subjektiven Einflüssen sind.

Bezüglich Hypothese 2 ergaben sich keine erhöhten Zusammenhänge des *Praktikumssozialklimas* mit affektiv geprägten PraQ-Skalen. Allerdings ergaben sich deutlich höhere und mehr signifikante Zusammenhänge mit Skalen der *Lehrkompetenz* im Vergleich zu Skalen des *Lernzuwachses* bzw. der *Lernumgebung*. Dies kann derart gedeutet werden, dass das *Sozialklima* im Praktikum stärker die Evaluation des/der Betreuenden konfundiert als die Selbstevaluation des eigenen *Lernzuwachses* bzw. die Bewertung der *Lernumgebung*.

Kritisch gesehen werden können die für die Methode der CFA recht geringen Stichproben $N < 250$.

12.6. Studie 4: Zusammenhang mit Expert*innenratings (»Multiperspektivität«)

12.6.1. Einleitung

JUSSIM (2012, S. 202) plädiert für eine Interpretierbarkeit von Evaluationsergebnissen dafür, die beteiligten Gruppen mit in den Prozess einzubeziehen und deren Einschätzungen zu erfassen. Dies entspricht einer Multiperspektivität der Evaluation. Für Interpretationen der PraQ-Messungen bei der Praktikumsevaluation (Ziel 1) wäre es demzufolge wünschenswert, dass die Messung mittels Einschätzung der Studierenden keinen zu starken »Perspektiveneffekt« aufweisen, sondern ähnlich wie Einschätzungen von Expert*innen ausfallen. Für die Skalen der Dimension *Lernzuwachs* müsste hierzu eine Expert*inneneinschätzung des *Lernzuwachses* jedes Studierenden und jeder Studierender vorgenommen werden.

Die Betreuenden erscheinen allerdings als mögliche Expert*innen eher ungeeignet, da diese in der Regel jede Woche eine neue Studierendengruppe betreuen und daher keine Diagnostik über den Gesamtfortschritt vornehmen können, zudem selten pädagogisch und didaktisch ausgebildet sind. Es müsste sich also um externe Expert*innen handeln, die eine Beobachtung des Praktikumsgeschehens vornehmen und den *Lernzuwachs* der beobachteten Studierenden einschätzen. Für die Skalen der Dimension *Lehrkompetenz* müsste selbiges für jeden Betreuenden und jede Betreuende vorgenommen werden. Der Aufwand für diese beiden Dimensionen wurde im Rahmen dieser Arbeit nicht geleistet (vgl. Ausblick).

Das Praktikumsskript in der dritten Dimension des 3L-Modells stellt allerdings einen nützlichen Anker für die multiperspektivische Erfassung von *Skriptqualität* dar, da dieses Expert*innen einfach zugesandt werden kann und keine Beobachtung

in situ notwendig ist. Dadurch wird die Forschungsfrage 6 zunächst nur auf diesen Bereich konzentriert.

Forschungsfrage

Forschungsfrage 6: In welchen Facetten der Skriptqualität unterscheiden sich Einschätzungen von Studierenden gegenüber Einschätzungen von Expert*innen? (»Multiperspektivität«)

12.6.2. Methode

Stichprobe Die Stichprobe auf Seiten der Expert*innen entspricht exakt der Stichprobe aus Studie 1.2, da die Befragungen nacheinander innerhalb derselben Erhebung abliefen. Die Stichprobe auf Seiten der Studierenden bildete eine Subgruppe aus dem Datensatz der Studie 2.2. Die Subgruppe bestand aus Physik-Studierenden der Humboldt-Universität zu Berlin, die das Grundpraktikum 2 der Physik besuchten. Im Rahmen der Befragungen der Studie 2.2 nahmen diese Einschätzungen der Skriptqualität für die beiden PraQ-Skalen Skriptqualität und Skriptqualität für den Versuch vor.

Instrumente Sowohl auf Studierenden-, als auch auf Expert*innenseite wurden dieselben beiden PraQ-Skalen zur *Skriptqualität* und *Skriptqualität für den Versuch* eingesetzt. Die Expert*innen erhielten zusätzlich einen exemplarisch gewählten Auszug des Praktikumsskripts des Grundpraktikums 2 der Humboldt-Universität zu Berlin (Anhang F auf Seite 289), den sie im Rahmen der Skalen bewerten sollten.

Auswertungsmethode Die benötigte Stichprobengröße wurde vor der Proband*innenauswahl der Expert*innen bestimmt. Die $N = 42$ Studierenden der Humboldt-Universität zu Berlin standen bereits aus Studie 2.2 fest. Die Analyse mittels G*Power ergab, dass für einen Vergleich der Gruppenmittelwerte der Skalen (t-Test für unabhängige Stichproben) zusätzlich mindestens $N = 30$ Expert*innen befragt werden mussten, um auf dem α -Level von .05 einen mittleren Effekt ($d = .50$) auflösen zu können.

Auf messgenauere Analysen im Rahmen von Strukturgleichungsmodellen (vgl. Studien 2.2, 3.2 und 5–7) musste wegen der geringen Stichprobengröße verzichtet werden. Die Gesamtbeurteilungen des Skriptes werden dann je Proband*in ermittelt. Dies geschah durch die manifeste Skalenwertberechnung mittels nach Faktorladungen gewichteter, arithmetisch gemittelter Itemausprägungen. Die so gewonnen Gesamtbeurteilungen der beiden Proband*innengruppen wurden anschließend

12.6. Studie 4: Zusammenhang mit Expert*innenratings («Multiperspektivität«)

durch t-Tests für unabhängige Stichproben unter BONFERRONI-HOLM-Korrektur (HOLM, 1979) verglichen.

Durchführung Die Studie wurde Ende des Sommersemesters 2015 auf Studierendenseite analog zur Erhebung in Studie 2.2 durchgeführt. Die Erhebung der Expert*innenratings fand im Juli und August 2015 online statt. Der exemplarische Auszug aus dem Praktikumsskript war in die Online-Befragung integriert.

12.6.3. Ergebnisse

Die Erfüllung der notwendigen Vorbedingungen wurde für die Normalverteilungsannahme durch den SHAPIRO-WILK-Test sowie für die Varianzhomogenität durch den LEVENE-Test geprüft und bestätigt. Für Skala 1 (*Skriptqualität für den Versuch*, Item 1–3) ergab der t-Test keine signifikanten Unterschiede zwischen Studierenden ($M = 4.1, SE = 0.2$) und Didaktik-Expert*innen ($M = 3.7, SE = 0.2$). ($t(70) = -1.36, p_{adj} = .36$). Für Skala 2 (*allgemeine Skriptqualität*, Item 4–16) ergab der t-Test, dass das Skript im Durchschnitt von den Studierenden ($M = 4.2, SE = 0.2$) besser beurteilt wurde als von den Expert*innen ($M = 3.7, SE = 0.2$). Dieser Unterschied ist tendenziell signifikant: $t(70) = -2.33, p_{adj} = .06$. Es handelt sich mit $d = 0.41$ um einen schwachen bis mittleren Effekt.

Des Weiteren wurde eine deskriptive Auswertung auf Itemebene vorgenommen. Dafür wurden die Mittelwerte der Antworten bezüglich der einzelnen Items zwischen den beiden Proband*innengruppen visuell verglichen (Abb. 12.7).

Betrachtet man das Antwortmuster der beiden Gruppen, so fällt auf: (a) Die mittlere Einschätzung durch die Studierenden lag bei allen Items höher oder auf dem gleichen Niveau wie die durch die Expert*innen. Dies stützt die Befunde aus dem Gruppenvergleich der Skalenmittelwerte auch auf Einzelitemebene. (b) Die Antwortmuster beider Proband*innengruppen ähneln sich in weiten Teilen. (c) Große Abweichungen im Muster ergeben sich nur für Items 7 und 8 (Skala *Skriptqualität*). Zudem ist die Expert*innenbeurteilung bei Items 11 und 16 eher schlechter ausgeprägt als die der Studierenden.

12.6.4. Diskussion

Die Analyse der Multiperspektivität der *Skriptqualität*-Items lieferte die Erkenntnis, dass Studierende tendenziell ihr Skript besser bewerten als Expert*innen der Fachdidaktik. Dies konnte für die Skala *Skriptqualität* festgestellt werden, da sich Expert*innenurteile tendenziell signifikant und mit mittlerer Effektstärke von Studierendenurteilen unterscheiden. Gemäß explorativer Analyse der Einzelitems ist

12. Testevaluation (Validität und Reliabilität)

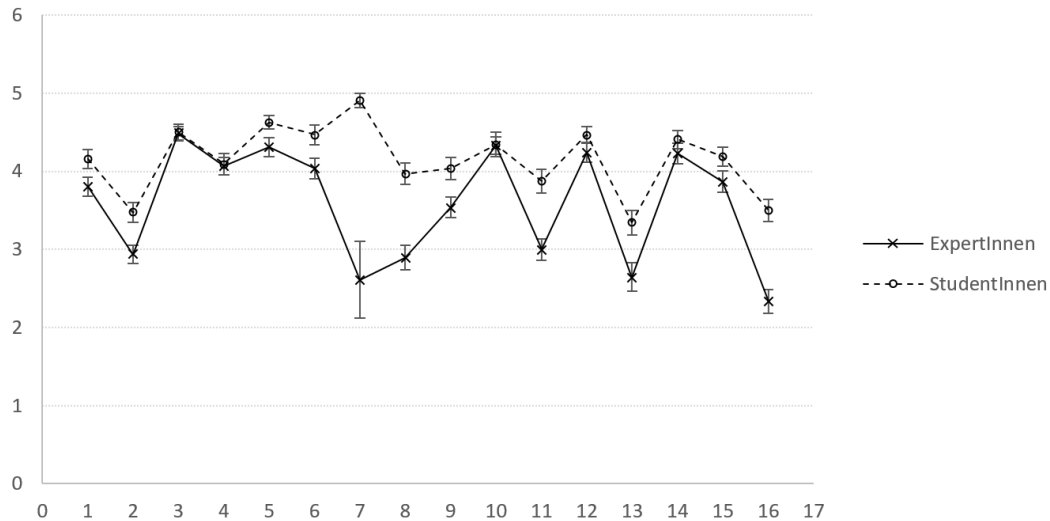


Abbildung 12.7.: Mittelwerte der Einzelitems der Skalen *Skriptqualität für den Versuch* (1–3) und *Skriptqualität* (3–16) der Studierenden und Expert*innen im Vergleich. Fehlerbalken entsprechen Standardfehlern, die Verbindungslinien dienen der Übersichtlichkeit und stellen keinen statistischen Trend dar.

diese Einschätzung aber stark unter Vorbehalt zu betrachten: Die großen Abweichungen in dieser Skala basieren mehrheitlich auf den Items 7, 8, 11 und 13.

Betrachtet man die Inhalte dieser Items genauer, fällt auf, dass es gute Argumente dafür gibt, warum Expert*innenratings hier geringer ausfallen. Diese Items sprechen nämlich funktionelle Ziele des Skripts an: Die Formulierungen von Item 7 und 16 beziehen sich sehr stark auf den tatsächlichen Ablauf des Praktikums, der von den Expert*innen durch die Art der Befragung nicht beurteilt werden konnte.⁴¹ Item 8 (»...ordnungsgemäße Protokolle/Versuchsberichte anzufertigen.«) bezieht sich auf einen Abschnitt des Skriptes, der aufgrund der Eingrenzung des Skriptauszuges in der Expert*innenbefragung nicht vorgelegt werden konnte. Bei Item 11 (»Die Lernziele des jeweiligen Versuches werden im Skript verdeutlicht.«) schließlich ist zu vermuten, dass die Verwendung von Begriffen aus der Fachsprache der Didaktik (»Lernziele«) zu verschiedenen Assoziationen in beiden Proband*innengruppen führte, wodurch in den beiden Gruppen letztendlich nicht das gleiche Merkmal erfasst wurde.

Die zweite Skala *Skriptqualität für den Versuch* zeigt indes weder auf Einzelitemebene noch auf Skalenebene einen signifikanten Unterschied zwischen Studierenden-

⁴¹ »Ich habe die Theorie aus dem Skript auch bei der Versuchsdurchführung genutzt.«, »Das Skript macht den Stellenwert der Sachverhalte im Praktikum für das weitere Studium klar.«

und Expert*innenurteilen. Für diese Skala scheint also die Annahme, dass Studierendenurteile auf Expert*innenniveau ausfallen, gerechtfertigt, was die Aussagekraft von Interpretationen der Maße dieser Skala stärkt.

Als Fazit der Studie 4 bleibt festzuhalten, dass für die PraQ-B-Skalen zur *Skriptqualität* eine zufriedenstellende Gleichheit zwischen Studierenden- und Expert*innenurteilen herrscht. Die Studierendenurteile bei einer Praktikumsevaluation mit PraQ (Ziel 1) sind also für die Skriptqualität »auf Expert*innenniveau«.

12.7. Zusammenfassung: Testevaluation

Im Folgenden werden zunächst die Evidenzen zur Validität und Reliabilität der Interpretationen von PraQ-Messungen zusammengefasst und eine Methodenkritik geliefert, was die Ergebnisse der Studien 1 bis 4 zusammenfasst. Danach folgen Argumente dazu, wie generalisierbar (vgl. Anforderung 4: Standardisierung, Abschnitt 4.4 auf Seite 43) die Ergebnisse der Testevaluationsstudie für alle naturwissenschaftlichen Grundpraktika im deutschsprachigen Raum sind. Schließlich wird im Ausblick auf weitere mögliche Testevaluationsschritte eingegangen.

12.7.1. Evidenz zur Validität und Reliabilität

Die vorangegangenen Abschnitte lieferten bezüglich der validen Interpretierbarkeit der PraQ-Messung sowohl theoretische wie auch empirische Evidenz. Gemäß argument-based approach lassen sich folgende Aussagen nach einer Messung mittels PraQ valide treffen:

- a) Aussagen über die inhaltlichen Aspekte von *Praktikumsqualität* in all seinen für relevant befundenen Facetten (Ausnahmen: Konstrukt *Experimentelle Kompetenz: Eigenständiges Planen von Experimenten*, Studie 1.1)
- b) Aussagen über die einzelnen Konstrukte von *Praktikumsqualität* in Abgrenzbarkeit voneinander und Bezug zueinander (Ausnahmen: Konstrukte *Experimentelle Kompetenz: Fragestellung klären*, *Lehrkompetenz: Lerndienliches Feedback geben* (je Studie 2.1), *Lehrkompetenz: Zur Selbsttätigkeit anregen* (Studie 2.2))
- c) Aussagen, die einen starken Einfluss affektiver *Praktikumszufriedenheit* auf die PraQ-Messung verneinen und damit die angegebenen Konstrukte als ursächlich für die Ausprägungen der Messergebnisse sehen (Studie 3.2).
- d) Aussagen, die bezüglich der *Skriptqualität* das Niveau der Skriptbeurteilung durch Studierende als annähernd gleichwertig zur Skriptbeurteilung durch Expert*innen ansehen.

12. Testevaluation (Validität und Reliabilität)

Folgende Aussagen lassen sich nur unter starkem Vorbehalt treffen:

- e) Aussagen, die die Entwicklung der *experimentierbezogenen Selbstwirksamkeitserwartung* über Messungen mit den PraQ-Skalen zur *Experimentellen Kompetenz* bewerten (Studie 3.1).

Folgende Aussage lässt sich auf Basis der bisherigen Ergebnisse *nicht* treffen:

- f) Die Aussage, die die Entwicklung der *Experimentellen Kompetenz* der Studierenden über Messungen mit den PraQ-Skalen zur Experimentellen Kompetenz bewertet (Studie 3.1).

Offensichtlich ist an dieser Stelle einzuräumen, dass bezüglich der Aussagen e) und f) noch Nachholbedarf bezüglich der Testevaluation besteht. Bezüglich f) ist bereits eine Folgestudie angelaufen.

Reliabilität Die Reliabilität sämtlicher Skalen des PraQ fällt zufriedenstellend bis sehr gut aus. Orientiert man sich an den Ergebnissen der Studie 2.2, wo eine messfehlerbereinigte Reliabilitätsschätzung erfolgen konnte, so liegen die Reliabilitäten RAYKOVs ρ aller Skalen im Bereich [.66, .93] mit einem Mittelwert von .87 ($SD = .06$).

12.7.2. Methodenkritik

Die dargestellten Testevaluationsstudien sind nicht frei von Verzerrungseffekten, methodischen Schwächen und Einflüssen von Subjektivität. Dies soll hier kompakt dargestellt werden, um einen Überblick über die Art und Stärke der Aussagekraft der Testevaluationsstudien zu gewinnen.

Studie 1.1 Die im kognitiven Interview erhaltenen Expert*innenurteile wurden direkt durch den Interviewer dahingehend interpretiert, ob sie ein Passungsproblem anzeigen oder nicht. Objektiver wäre hier eine Transkription der Interviewdaten und eine inhaltsanalytische Auswertung unter Hinzuziehen multipler Rater gewesen (Interrater-Reliabilität). Zudem liegt je Fachgebiet nur ein Expert*inneninterview vor, was die Generalisierbarkeit je Fach in Frage stellt. Besser wären hier je Fach mehrere Expert*innen gewesen.

Studie 1.2 Die Frage an die Expert*innen, Aspekte der *Skriptqualität* zu ergänzen, wurde über offene Fragen im Onlinefragebogen realisiert. Dieser Fragetyp wird aus Erfahrung selten genutzt. Die Abwesenheit von Ergänzungen kann daher nur sehr bedingt als Vollständigkeitsargument für den *Skriptqualität*-Itempool angesehen werden. Besser wären hier vorangestellte, qualitative Befragungen (Offener

Fragebogen oder Interview) gewesen, die nach einer Definition und Aspekten der *Skriptqualität* fragen und dann inhaltsanalytisch ausgewertet werden.

Studien 2.1 und 2.2 Die Studie 2.1 weist zunächst eine Stichprobenverzerrung hin zu Praktika mit engagierteren Praktikumsleitenden auf, da die Teilnahme freiwillig war und der Kontakt zu den Praktika über Kommunikation mit Praktikumsleitenden aufgenommen wurde. Diese Verzerrung ist allerdings als moderat einzuschätzen, da von den initiativ vom Autor angeschriebenen Praktika immerhin 71% Zusagen erfolgten. Bei über die Praktikumsleitertagung DPG-Schule hergestellten Kontakten waren es 100%.

Die zweite Verzerrung der Stichprobe eine Häufung von Physikpraktika in der Stichprobe. Auf Grund der fachlichen Nähe des Autors zur Physik konnten dort deutlich mehr Kontakte zu Praktika im deutschsprachigen Raum aufgebaut werden. Die enthaltenen Chemie-Praktika weisen also dahingehend eine Verzerrung auf, als das sie vornehmlich den Berlin-Brandenburger Raum abdecken, wo Kontakte leichter geknüpft werden konnten. Biologisch geprägte Praktika wurden indes nur in sehr geringem Maße berücksichtigt. Die Aussagen aus den Studien 2.1 und 2.2 sind also vor allem auf andere Physik- und ggf. Chemiepraktika potentiell übertragbar.

Eine Verzerrung der Studierendenkohorten jedes einzelnen Praktikums ist auf Basis der freiwilligen Teilnahme an der Evaluationsbefragung natürlich möglich, allerdings lassen die im Schnitt über 90% Rücklaufquoten bei Präsenzerhebungen Gegenteiliges vermuten.

Bezüglich der Konstrukte der *Lehrkompetenz* der Betreuenden im PraQ-B kann Kritik geäußert werden, dass nur wenige Studierende je Einzelbetreuer*in eine Bewertung abgaben (je $N < 10$). Dies liegt daran, dass die PraQ-B-Erhebungen stets den/die »zuletzt erlebte/n« Betreuende/n zum Gegenstand hatten und sich diese/r je Praktikumskleingruppe unterscheidet. Besser wäre hier eine Bewertung aller Betreuenden durch je alle Studierenden gewesen, was aber aus erhebungsökonomischen Gründen schwierig zu realisieren war. Für strukturelle Untersuchungen (Studie 2) ist dies allerdings wenig relevant, da die Zusammenhänge zwischen den PraQ-B-Items der Lehrkompetenz über die Lehrenden hinweg als konstant angenommen wurden.

Studie 3.1 Für Studie 3.1 lässt sich zunächst kritisieren, dass ein Instrument zur Kompetenzmessung (Ko-WADiS) verwendet wurde, das selbst noch einige Testevaluationsschritte zu durchlaufen hat (KRÜGER et al., 2017), sich daher nur bedingt als »Referenz« eignet. Allerdings war dies zum Zeitpunkt der Erhebungen eines der etabliertesten Instrumente. Besser wäre natürlich ein Vergleich mit einem ausreichend testevaluierten Instrument.

Darüber hinaus lässt sich kritisieren, dass die Maße für *Experimentelle Kompetenz* und *Selbstwirksamkeitserwartung* unterschiedliche Erhebungsmethoden (Leistungstest bzw. Ratingskala) aufweisen, für die nicht kontrolliert wurde. Besser wäre hier entweder eine Gleichheit der Methoden oder der Einsatz vierer Maße, der *Experimentellen Kompetenz* als Leistungstest und als Ratingskala, sowie der *Selbstwirksamkeitserwartung* als Leistungstest und Ratingskala. So könnte für den Erhebungsmodus kontrolliert werden. Derzeit fehlen dafür allerdings die benötigten Instrumente.

Studie 3.2 Die Studie 3.2 untersucht nur den Zusammenhang mit einer affektiv geprägten Variablen, dem Praktikumsclima. Besser wäre hier eine breite Auswahl an möglicherweise das Evaluationsergebnis verzerrenden affektiven Variablen.

Studie 4 Für Studie 4 wurde bereits in der Diskussion erörtert, dass vier der Items, die im Vergleich zwischen Expert*innen- und Studierendenurteilen enthalten sind, nicht zur Bewertung durch die Expert*innen geeignet waren. Eine Multiperspektivität der Skriptskalen ließ sich dadurch nur für die restlichen Items untersuchen.

12.7.3. Generalisierbarkeit der Testevaluation

Die Testevaluationsstudie soll empirische Evidenz dafür liefern, dass viele Arten von Interpretationen der Testergebnisse des PraQ abgesichert sind und Aussagen getroffen werden dürfen. Die Faktoren für eine Generalisierbarkeit der Validierungsergebnisse nach AERA et al. (2014, S. 18) sind Gleichheit der Konstruktmessung, Gleichheit des Curriculums dahinter, Vergleiche mit gemessenen Kriterien und geringe Unterschiede in der Stichprobe:

Gleichheit der Konstruktmessung Wird in anderen Praktika als den bisherigen erhoben, so bleiben die Konstrukte durch die standardisierte PraQ-Erhebung gleich und die neue Stichprobe wird mit hoher Wahrscheinlichkeit nicht andersartig auf die Itemformulierungen reagieren, da für verschiedene Studierendengruppen testevaluiert wurde.⁴² Dies gilt insbesondere für Physikpraktika und eingeschränkt für Chemiepraktika.

Gleichheit des Curriculums dahinter Der Abschnitt 3 auf Seite 25 zur Lehre in den naturwissenschaftlichen Studiengängen und auch der Forschungsstand zu Praktika (vgl. Abschnitt 5 auf Seite 55) haben gezeigt, dass im Wesentlichen ein curricularer, zielspezifischer Konsens in der Zielsetzung und Ausgestaltung von naturwissenschaftlichen Praktika im deutschsprachigen Raum besteht.

⁴²Dies kann auch mittels DIF statistisch untersucht werden, vgl. Studien 6 und 7.

Gemessene Kriterien Eine Testevaluation mittels Vergleich mit einem relevanten Kriterium wurde nicht vorgenommen, da aus Sicht des Autors für Praktika kein derartiges Kriterium vorliegt.⁴³

Unterschiede in der Stichprobe Für Physik- und mit Abstrichen auch Chemiepraktika (vgl. Studie 6) kann wie erörtert eine Übertragbarkeit angenommen werden. Nachholbedarf besteht beim Einsatz des PraQ in biologischen und anderen naturwissenschaftlichen Praktika.

Zeitäquivalenz Die Zeitpunkte der Erhebungen werden optimalerweise zu denen in den Testevaluationsstudien angeglichen, erfolgen also zu Ende des Semesters, meist zwei Wochen vor Abschluss, da dort die Anwesenheit der Studierenden noch nicht durch den Prüfungszeitraum dezimiert ist.

Festzuhalten ist also eine recht sichere Übertragbarkeit aller Testevaluationsergebnisse auf andere physikalische Grundpraktika und eine moderate Übertragbarkeit auf andere Chemiepraktika im deutschsprachigen Raum.

12.7.4. Ausblick: Weitere Testevaluationsschritte

Die Untersuchung des Folgeaspekts (MESSICK, 1995) der Konstruktvalidität kann dann geschehen, wenn Praktika auf Basis der Messinterpretation des PraQ modifiziert werden und dann eine erneute Messung erfolgt, ob die Modifizierung zu wünschenswerten und fairen Folgen geführt hat. Die Grenzen der Aussagekraft von Ergebnissen des PraQ zeigt auch der Bedeutungsaspekt der Validität (ebd.) auf. Es bleibt zunächst unklar, welche kognitiven Prozesse bei den Studierenden tatsächlich bei der Instrumentbearbeitung ablaufen. Dies sollte in Zukunft über Interviews abgesichert werden.

Der Einsatz für verschiedene Gruppen wie Studierende unterschiedlicher Studiengänge sollte zudem über Analysen zur Messinvarianz abgesichert werden. Für die Gruppen Chemie- und Physikpraktika, sowie Block- und Semesterpraktika wird dies in den Studien 6 und 7 realisiert.

Da Gruppenarbeiten und damit gegenseitige Beeinflussungen des Lernprozesses in Praktika zum Standard gehören, könnten zudem zukünftig hierarchische Modelle in der CFA betrachtet werden, was überdies auch Möglichkeiten der Hierarchisierung nach Standorten der Praktika zuließe.

⁴³ Gegebenenfalls könnte das Kriterium »Protokollnote« relevant für das Konstrukt Kommunikation (schriftlich) sein.

Kapitel 13

Studien: Einsatz von PraQ im deutschsprachigen Raum

Die Erhebungen im Zuge der Studien 2.1 und 2.2 hatten erstmalig überhaupt einen Datensatz zur Folge, der deutschlandweit über 18 Praktika in zwölf Standorten hinweg erhobene Daten beinhaltet und in einer einzigen deutschlandweiten Studie über 1300 Proband*innen beinhaltet. Gemäß der positiven Ergebnisse der Testevaluationsstudien (Studien 1 bis 4) ist zudem nun ein inhaltlich geprägtes Forschen mit den PraQ-Ergebnissen möglich (Ziel 1).

Für die Ergebnisse der PraQ-Messung wurde argumentiert, dass diese eine gute Generalisierbarkeit über naturwissenschaftliche Praktika in Physik und mit Abstrichen Chemie in ganz Deutschland aufweisen. Die Verallgemeinerbarkeit ist nahelegend, da der Datensatz groß, die Messsituationen authentisch und übertragbar ist und die Stichproben relativ verzerrungsfrei die Lehrrealität in vielen Praktika in Deutschland abbilden (AERA et al., 2014, S. 18).

Es existieren 52 Physikpraktikums-Standorte gemäß Liste der AG Physikalische Praktika der DPG (RUNGE, 2017). Die allgemeine Übertragbarkeit der nachfolgenden Ergebnisse auf den gesamten deutschen Sprachraum kann damit für die Physik berechtigt angenommen werden.

Die nun folgenden drei Studien haben zum Gegenstand, generalisierbare Erkenntnisse zu Praktika im deutschsprachigen Raum hervorzubringen.

In der Studie 5 wird zunächst deskriptiv ein Stärken-und-Schwächen-Profil der Praktika erstellt und auf Unterschiede zwischen Praktika mit besonders hoher und besonders gering ausgeprägter Praktikumsqualität exemplarisch eingegangen. Dies dient einem ersten Überblick.

Die Studie 6 hat zum Gegenstand, die Unterschiede zwischen Chemie und Physikpraktika herauszustellen, um Hinweise auf Bereiche zu erhalten, in denen diese voneinander lernen können. Dort wird vorab eine notwendige Analyse auf Messinvarianz zwischen diesen beiden Gruppen erfolgen.

Bei der Studie 7 schließlich wird die hochschulorganisatorisch relevante Frage untersucht, ob Blockpraktika sich von Semesterpraktika in ihrer Qualität unterscheiden. Auch hier wird vorab eine notwendige Analyse auf Messinvarianz zwischen diesen beiden Gruppen erfolgen.

13.1. Studie 5: Stärken und Schwächen von Praktika in Deutschland

13.1.1. Einleitung

Forschungsstand Der Forschungsstand zu Praktika ist in Abschnitt 5 auf Seite 55 zu finden. Zusammenfassend zeichnen sich die bisher untersuchten Praktika im deutschsprachigen Raum dadurch aus, dass sie geschlossene Lernumgebungen mit einem hohen Anleitungsgrad darstellen (»expository style«, vgl. Tab. 3.1 auf Seite 35).

Die Vorbereitung geschieht meist außerhalb der Präsenzzeit und meist theoriebent. Sie wird oftmals durch ein Skript gesteuert, das die versuchsbezogene Theorie enthält. Das Skript ist gleichzeitig auch Versuchsanleitung für die Präsenzzeit und wird oft als rezeptartig in seiner anleitenden Art beschrieben und gibt den Studierenden dadurch wenig Einblick in das tatsächliche Experiment. Es lassen sich auch einige Überlegungen zu PraQ-relevanten Konstrukten ableiten:

Bezüglich des Kompetenzzuwachses der Studierenden kann gemäß den Ergebnissen von HUCKE (1999) geschlussfolgert werden, dass *Fachwissen* durchaus im Praktikum erworben wird, wenn auch motivationale Einschränkungen vorliegen können.

Die *Experimentelle Kompetenz* muss dagegen je Subkonstrukt differenziert betrachtet werden. Das *Formulieren von Erwartungen* an den Versuch steht in einem nennenswerten Bezug zum Bilden bzw. *Verstehen von Hypothesen*. Da letzteres gemäß Forschungs- und Argumentationslage in Praktika wenig gefordert ist, dürfte auch der Kompetenzzuwachs hier gering sein. Die *deklarativen Gerätekenntnisse* werden durch die Skripte mutmaßlich nur wenig gefördert (vgl. *Skriptqualität für den Versuch*; ZASTROW 2001), die Rolle des/der Betreuenden hierfür ist unklar, weswegen keine Hypothese aufgestellt wird. Das *Geräte Auswählen und Versuchsanordnung Aufbauen* wird gemäß Forschungslage eher gering ausfallen, da diese Phasen auf Grund des gerade in der Physik oftmals bereits voraufgebauten Versuches vernachlässigt werden. Das *Durchführen und Dokumentieren von Messungen* bildet gemäß HUCKE (1999) einen zeitlich starken Schwerpunkt im Praktikumsgeschehen, sollte daher dort auch gut gefördert werden. Das *Darstellen von Rohdaten* ist vor allem in Physikpraktika bei der Erarbeitung der Versuchsprotokolle relevant und dürfte dort entsprechend mittelstark bis hoch ausfallen.

Gleiches gilt für die *schriftliche Kommunikationskompetenz*, die sich ausschließlich auf das Versuchsprotokoll bezieht, dem ein hoher Zeitaufwand zugeschrieben wird (METZGER & SCHULMEISTER, 2011).

Mündliche Kommunikationskompetenz könnte sich sowohl bei der Interaktion mit Betreuenden als auch in der Tandemarbeit ausbilden, Evidenz hierfür ist erneut rar.

Bezüglich des *Umgangs mit Problemen und Fehlern und der Dateninterpretation zum Experimentieren* lässt sich wenig Evidenz heranziehen, die eine Hypothese möglich machte.

Für *Bewertungskompetenz* lässt sich für Physikpraktika der starke Fokus auf Fehlerrechnung anführen, allerdings ist dies nur ein kleiner Teil des Konstrukts. Es kann daher keine plausible Hypothese aufgestellt werden.

Ähnliches gilt für das *Zeitmanagement* und die *Kooperationskompetenz*, die allerdings in Physikpraktika durch die vermehrte Tandemarbeit durchaus höher ausfallen dürfte (Studie 6). Auch hier existieren keine leitenden Vorbefunde.

Bezüglich der *intrinsischen Motivation* ist zu erwarten, dass diese für fachfremde Praktikumsbesuchende geringer ausfällt als für fachnahe (Studie steht noch aus), eine Annahme zur Gesamtausprägung ist vor dem Hintergrund der Studie von NAGEL (2009, vgl. Abschnitt 5.1) eher als gering einzuschätzen.

Da die *Lehrkompetenz* der Betreuenden in Praktika noch nicht systematisch analysiert wurde, können hierfür nur theoretische Argumente vorgebracht werden. Da die Mehrzahl der Betreuenden in den untersuchten Praktika Studierende oder wissenschaftliche Mitarbeiter des Fachstudienganges sind, liegt in der Regel keine didaktische oder pädagogische Vorbildung bei den Betreuenden vor. Die didaktisch und pädagogisch geprägten Konstrukte der *Lehrkompetenz* werden also mutmaßlich mittelmäßig bis gering ausfallen. Für die *Skriptqualität bezogen auf den Versuch* sind gemäß obigen Ausführungen geringe Ausprägungen zu erwarten. Für die *allgemeine Skriptqualität* lässt sich auf Basis der häufigen Nennung von veralteten, »hausgemachten« Skripten und dem Wunsch nach multimedialen Alternativen seitens vieler intervenierender Fachdidaktiker*innen in der Literatur vermuten, dass diese keine hohen Ausprägungen im deutschlandweiten Schnitt aufweisen werden.

Die *Integration der Begleitvorlesung* impliziert zunächst, dass überhaupt eine solche existiert. Laut Erhebungsdaten ist dies in 52% der Praktika der Gesamtstichprobe des PraQ-B der Fall. Für diese dürfte die *Integration* mutmaßlich nicht sehr hoch ausfallen, da viele Befunde zum mangelhaften Erlernen der Theorie durch die Studierenden dem widersprechen, wird doch die *Integration* als Schlüsselmoment dafür gesehen, das theoretische Wissen aus der Vorlesung durch das praktische Arbeiten im Praktikum zu vertiefen.

Forschungsfragen und Hypothesen Die Ausführungen zum Forschungsstand lassen sich direkt auf die Forschungsfrage 7 beziehen und Hypothesen diesbezüglich

aufstellen:

Forschungsfrage 7: Was sind Stärken, Schwächen und die qualitative Bandbreite von Praktika in Deutschland in Bezug auf die Praktikumsqualität?

Hypothesen 1: Folgende Konstrukte der Praktikumsqualität fallen gemäß Forschungsstand gering aus (Kriterium $M < 3.5$):

- a) Experimentelle Kompetenz: Erwartungen formulieren
- b) Experimentelle Kompetenz: Versuchsplanung & Geräte zusammenstellen
- c) Experimentelle Kompetenz: Versuchsanordnung aufbauen¹
- d) Personalkompetenz: Interesse
- e) Lehrkompetenz in sämtlichen didaktisch und pädagogisch geprägten Konstrukten
- f) Skriptqualität
- g) Skriptqualität für Versuchsablauf
- h) Integration

Hypothesen 2: Folgende Konstrukte der Praktikumsqualität fallen gemäß Forschungsstand hoch aus ($M > 4.5$):

- i) Fachwissen Theorie
- j) Fachwissen Praxis
- k) Experimentelle Kompetenz: Messungen durchführen
- l) Experimentelle Kompetenz: Messungen dokumentieren
- m) Experimentelle Kompetenz: Rohdaten graphisch, tabellarisch darstellen²
- n) Kommunikationskompetenz schriftlich

Keine Hypothese konnte bezüglich folgender Konstrukte aufgestellt werden:

¹ v. a. Physikpraktika, vgl. Studie 6

² v. a. in Physikpraktika, vgl. Studie 6

- o) Experimentelle Kompetenz: Gerätekenntnisse deklarativ
- p) Experimentelle Kompetenz: mit Problemen und Fehlern umgehen
- q) Experimentelle Kompetenz: Daten interpretieren
- r) Kommunikationskompetenz mündlich
- s) Bewertungskompetenz
- t) Personalkompetenz: Zeitmanagement
- u) Kooperationskompetenz

Explorative Frage: Wie fällt ein Vergleich der Qualitätskonstrukte zwischen einem von der Praktikumsqualität her hoch ausgeprägten mit einem gering ausgeprägten Praktikum aus?

13.1.2. Methode

Stichprobe Nach der erfolgreichen strukturellen Validierungsstudie (Studie 2) kann der Gesamtdatensatz zur Analyse weiterer Forschungsfragen genutzt werden, wobei natürlich nur diejenigen Items berücksichtigt wurden, die nach der zweistufigen strukturellen Validierung mittels explorativer und konfirmatorischer Faktorenanalyse erhalten blieben.

Es ergaben sich $N = 813$ Fälle für den PraQ-A und $N = 501$ Fälle für den PraQ-B. Erneut wurde auf Grund auch designbedingt hoher Anteile fehlender Werte bei den Skalen *Skriptqualität* und *Integration* mit einem Sekundärdatensatz von $N = 305$ für den PraQ-B gearbeitet. Insgesamt ergibt sich so ein Bild über 18 Praktika im deutschsprachigen Raum an insgesamt zwölf Standorten. Die Studierendenstichprobe beschreibt Tab. 13.1 genauer.

Instrumente Die eingesetzten Instrumente waren je Praktikum entweder sämtliche Skalen des PraQ-A oder des PraQ-B nebst je den Kontroll- und Gruppenvariablen, die in der Testkonstruktion (Kapitel 10 auf Seite 107) beschrieben wurden.

Auswertungsmethode Wie in den Hypothesen bereits angedeutet wird der 1-Stichproben-t-Test mit BONFERRONI-HOLM-Korrektur genutzt, um die Höhe der Konstruktausprägungen gegen das jeweilige Kriterium (z. B. Prüfwert 3.5) zu testen (Skala reicht von 1 – stimme gar nicht zu bis 6 – stimme völlig zu). Die Skalenscores

Tabelle 13.1.: Stichprobenbeschreibung für die Studie 2.2.

Studien- gänge	Sem	Fachpr	fachfremd Physik	fachfremd Chemie	Block	weibl	Tan- dem
	43% Phy 30% Chem 7% Bio 20% sonst.	75% Phy 13% Chem 3% VetMed 9% sonst.					
PraQ-A	3.1 (1.4)		47%	0%	23%	34%	92%
	42% Phy 20% Bio 15% VetMed 11% Chem 12% sonst.	78% Phy 14% VetMed 7% Chem 1% sonst.	39%	0%	13%	43%	91%
PraQ-B							

Studiengänge TN: Studiengänge der teilnehmenden Studierenden
Sem: Fachsemester: $M(SD)$
Fachpr: Fachpraktikum
fachfremd Phy: Anteil fachfremder Studierender in Physikpraktika
fachfremd Chem: Anteil fachfremder Studierender in Chemiepraktika
Block: War es ein Blockpraktikum?
weibl: Anteil weiblicher Studierender
Tandem: Wurde in Tandem gearbeitet?
VetMed: Veterinärmedizin

werden hierzu ladungsgewichtet manifest berechnet, wie es für spätere Auswertungen mit PraQ üblich sein wird.³

Die jeweiligen Items sind in ihrer Itemschwierigkeit teils unterschiedlich, allerdings wird durch die Operationalisierung der PraQ-Items als Veränderungsmaß bei der Ablehnung eines Items keine Entwicklung und bei der Zustimmung eine Entwicklung im jeweiligen Konstrukt ausgedrückt. Dadurch sind die Skalenausprägungen zwar nur bedingt miteinander vergleichbar, aber durchaus gegen ein Kriterium (hier beispielsweise 3.5, respektive 4.5) zu testen, um einen Eindruck vom Zustimmungsverhalten zu gewinnen.

Darüber wird in erster Instanz operationalisiert werden, ob es sich um eine »Stärke« ($M > 4.5$) oder eine »Schwäche« ($M < 3.5$) von Praktika handelt. Der AERA et al. (2014, S. 29, Standard 1.20) folgend werden je signifikantem Resultat Effektstärken berichtet, um ein standardisiertes Maß der Bedeutsamkeit des Unterschiedes zum Prüfwert zu erhalten.

Durchführung Die Daten stammen aus den Studien zur strukturellen Validierung (Studie 2), weshalb die Durchführungsbeschreibung dort zu finden ist.

13.1.3. Ergebnisse

Die Ergebnisse sind in Abb. 13.1 für den PraQ-A⁴ und in Abb. 13.2 auf Seite 203 für den PraQ-B dargestellt. Es handelt sich wie gewohnt um eine Darstellung auf der 6-stufigen Likertskala von 1 – stimme gar nicht zu, bis 6 – stimme völlig zu. Dargestellt sind die Ausprägungen der Skalenscores in drei verschiedenen Gruppen.

Zur Gruppenauswahl wurde sich am Gesamtskalenscore orientiert. Dieser stellt den mittleren Skalenscore über alle gemessenen Skalen dar und gibt einen Hinweis auf eine hohe oder niedrige Gesamtqualität des jeweiligen Praktikums.⁵

Hellgrau dargestellt sind die Skalenscores für das Praktikum mit dem höchsten Gesamtskalenscore der Stichprobe (»bestes Praktikum«), weiß dargestellt sind die Skalenscores für das Praktikum mit dem geringsten Gesamtskalenscore (»schlechtestes Praktikum«). Dies dient der Darstellung der qualitätsbezogenen Bandbreite von Praktika. Darüber hinaus werden die Skalenscores für die Gesamtstichprobe

³Es kann von erhebenden Praktikumsleitenden oder Hochschulevaluierenden (Ziel 1) in der Regel nicht abverlangt werden, latente Modelle zur Auswertung heranzuziehen, es empfiehlt sich eine gewisse Pragmatik.

⁴Für den PraQ-A konnte die Skala Experimentelle Kompetenz: Versuchsplanung & Geräte zusammenstellen auf Grund zu hohem Anteil fehlender Werte nicht sinnvoll ausgewertet werden. Hypothese b) kann daher nicht geprüft werden.

⁵Eine Gewichtung einzelner Skalen wurde nicht vorgenommen. Es handelt sich lediglich um einen einfachen Indikator zur Kontrastierung der Praktika.

des PraQ dargestellt, also je der mittlere Skalenscore aller erfassten Praktika zusammengekommen.

Ergebnisse PraQ-A Für die Konstrukte des PraQ-A werden nun die Hypothesen a–d, i–n, sowie die explorativ zu untersuchenden Konstrukte o–u untersucht. Tab. 13.2 stellt, sofern erforderlich, die Ergebnisse der inferenzstatistischen Tests dar.

Es ist ersichtlich, dass nur zwei gerichtete Hypothesen statistisch gestützt werden. Die Verbesserung der *intrinsischen Motivation* als Teil der *Personalkompetenz* fällt erwartungsgemäß tendenziell signifikant kleiner als der Prüfwert 3.5 aus, allerdings nur mit äußerst kleiner Effektstärke ($p_{adj\&onetailed} < .09^\circ$, $d = 0.05$). Das *Fachwissen zur Praxis* verbessert sich hypothesenkonform stark (Prüfwert 4.5 hochsignifikant überschritten). Es handelt sich hierbei allerdings in Bezug zum Prüfwert 4.5 um einen kleinen Effekt ($p_{adj\&onetailed} < .001$, $d = 0.19$).

Die explorativen Analysen zeigen weder besonders stark, noch besonders schwach ausgeprägte Skalen. Bezüglich des Vergleiches von bestem und schlechtestem Praktikum fällt zunächst auf, dass eine gewisse Spannweite der Skalenscores besteht. Sie decken im Mittel den Bereich [2.7; 5.0] ab, weisen also eine leichte Deckentendenz auf. Unterschiede werden hier für alle Konstrukte außer *EXP: Mit Problemen und Fehlern umgehen* und *Fachwissen: Praxis* sichtbar. Dies gibt Evidenz dafür, dass eine Profilierung von Praktika auf Basis ihrer Praktikumsqualität möglich ist.

Ergebnisse PraQ-B Für die Konstrukte des PraQ-B werden nun die übrigen Hypothesen e–h untersucht. Abb. 13.2 und Tab. 13.3 stellen die Ergebnisse dar.

Es ist ersichtlich, dass keine der aufgestellten gerichteten Hypothesen statistisch gestützt wird. Allerdings zeigen einige Konstrukte der als niedrig ausgeprägt hypothetisierten *Lehrkompetenz* von Praktikumsbetreuenden ein gegenteiliges Bild: Die Konstrukte *verständlich erklären*, *Gute Lernatmosphäre herstellen*, *Effizient mit Störungen umgehen* und *Diagnostik: Grundeinstellung* übertreffen bei zweiseitiger Testung den Prüfwert 4.5 signifikant bei geringer Effektstärke ($d \in [0.21; 0.40]$).

Die explorativen Analysen zeigen erneut weder besonders stark, noch besonders schwach ausgeprägte Skalen.

Bezüglich des Vergleiches von bestem und schlechtestem Praktikum fällt erneut auf, dass eine gewisse Varianz der Skalenscores besteht. Sie decken im Mittel den Bereich [3.1; 5.5] ab, weisen also auch eine leichte Deckentendenz auf. Prägnante Unterschiede werden hier für die Konstrukte *Verknüpfungen herstellen*, *Verständnis überprüfen*, *Diagnostik situativ*, *Skriptqualität für den Versuch* und *Skriptqualität* sichtbar. Dies gibt Evidenz dafür, dass eine Profilierung von Praktika auf Basis ihrer Praktikumsqualität möglich ist.

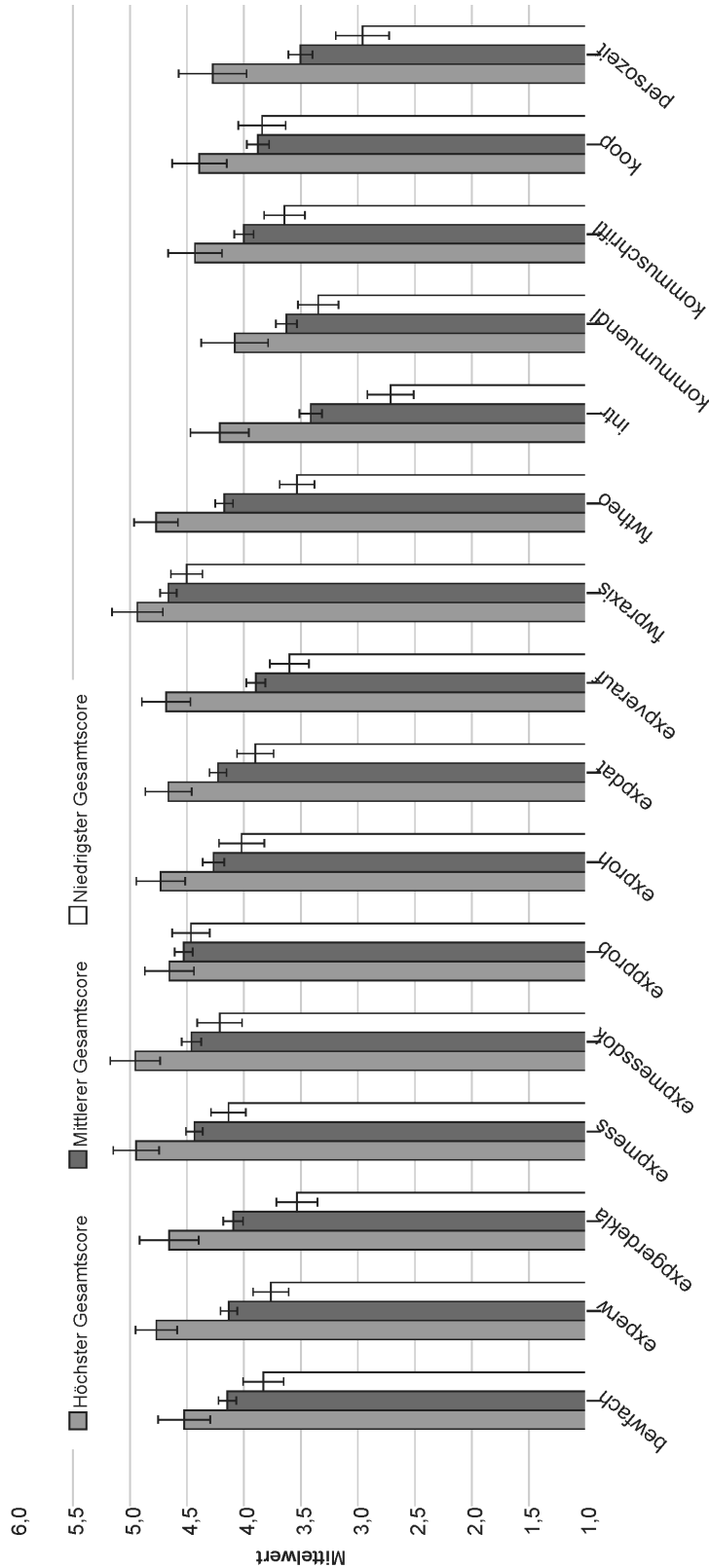


Abbildung 13.1.: Gegenüberstellung der manifesten Skalenwerte für das Praktikum mit dem höchsten ($N = 60$) und dem geringsten ($N = 148$) Gesamtscore im PraQ-A. Zusätzlich sind die mittleren Ausprägungen über den Gesamtdatensatz ($N = 813$) dargestellt.

fwtheo: Fachwissen Theorie; fwpraxis: Fachwissen Praxis; experw: Experimentelle Kompetenz: Erwartungen formulieren; expverger: Experimentelle Kompetenz Versuchsplanung & Geräte zusammenstellen; expgerdeka: Experimentelle Kompetenz: Gerätekenntnisse deklarativ; expverauf: Experimentelle Kompetenz: Versuchsanordnung aufbauen; expmess: Experimentelle Kompetenz: Messungen durchführen; expmessdok: Experimentelle Kompetenz: Messungen dokumentieren; expprob: Experimentelle Kompetenz: mit Problemen und Fehlern umgehen; exprohr: Experimentelle Kompetenz: Rohdaten graphisch tabellarisch darstellen; expdat: Experimentelle Kompetenz: Daten interpretieren; kommunschrift: Kommunikationskompetenz schriftlich; kommunuendl: Kommunikationskompetenz mündlich; bewfach: Bewertungskompetenz fachlich; koop: Kooperationskompetenz; intr: Interesse; persozzeit: Personalkompetenz; Zeitmanagement

Tabelle 13.2.: Ergebnisse der 1-Stichproben-t-Tests zur Konstruktausprägung in der PraQ-A-Gesamtstichprobe ($N = 813$). *padj&onetailed* wurde mittels BONFERRONI-HOLM-Korrektur (HOLM, 1979) und einseitig für gerichtete Hypothesen berechnet. Wichtige Ergebnisse sind fett gedruckt.

Test auf $M < 3.5$						
Hypothese	Konstrukt	M	SE	<i>padj&onetailed</i>	d	
a	EXP: Erwartungen formulieren	4.15	0.03	n.s.	-	
b	EXP: Versuchsplanung & Geräte	-	-	-	-	
c	EXP: Versuchsanordnung	3.93	0.04	n.s.	-	
d	Interesse	3.43	0.05	.09	0.05	
Test auf $M > 4.5$						
i	Fachwissen Theorie	4.17	0.04	n.s.	-	
j	Fachwissen Praxis	4.67	0.03	< .001	0.19	
k	EXP: Messungen durchführen	4.44	0.04	n.s.	-	
l	EXP: Messungen dokumentieren	4.45	0.04	n.s.	-	
m	EXP: Rohdaten darstellen	4.25	0.05	n.s.	-	
n	Kommunikation schriftlich	3.98	0.04	n.s.	-	
explorativ						
o	EXP: Gerätekenntnisse	4.12	0.04	n.s.	-	
p	EXP: Probleme lösen	4.52	0.04	n.s.	-	
q	EXP: Daten interpretieren	4.23	0.03	n.s.	-	
r	Kommunikation mündlich	3.64	0.04	n.s.	-	
s	Bewertungskompetenz	4.10	0.04	n.s.	-	
t	Zeitmanagement	3.52	0.05	n.s.	-	
u	Kooperationskompetenz	3.89	0.05	n.s.	-	

padj&onetailed: BONFERRONI-HOLM-korrigierter p -Wert bei gerichteter Hypothese
d: Effektstärke COHENS d
EXP: Experimentelle Kompetenz (PraQ)

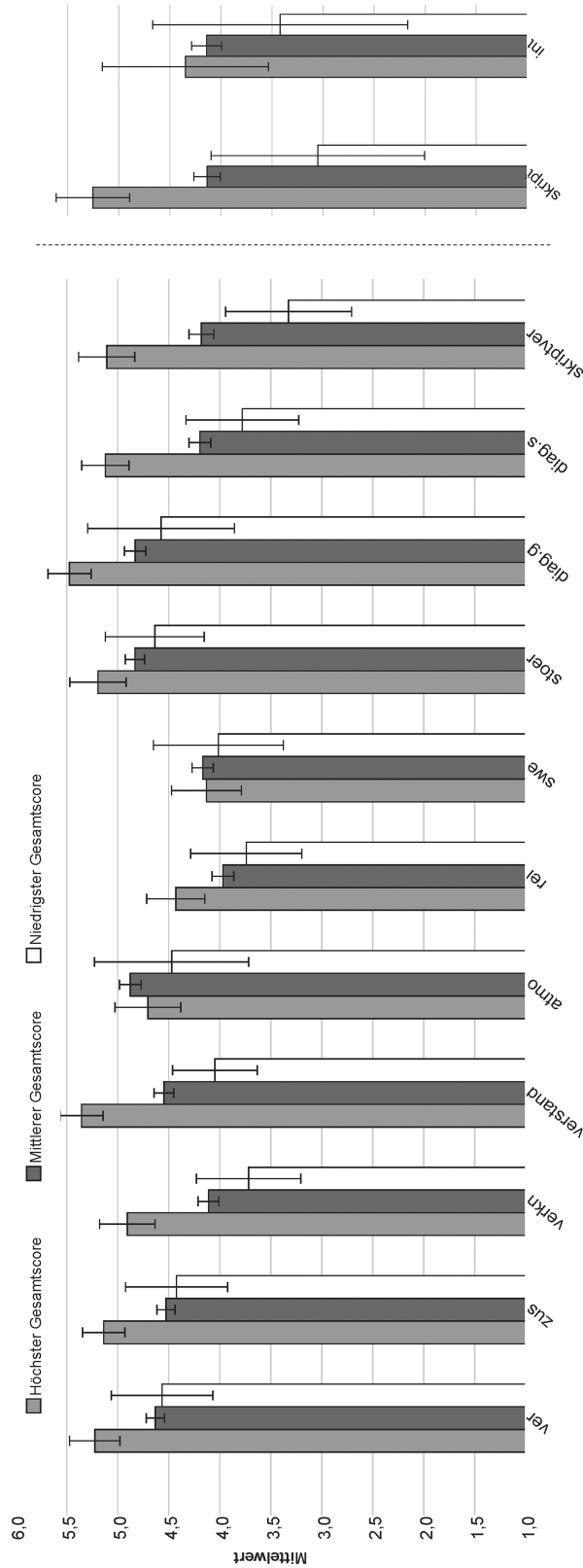


Abbildung 13.2.: Gegenüberstellung der manifesten Konstruktausprägungen für das Praktikum mit dem höchsten ($N = 49$) und dem geringsten ($N = 19$) Gesamtscore des PraQ-B. Zusätzlich sind die mittleren Ausprägungen über den Gesamtdatensatz ($N_{\text{primär}} = 501$) dargestellt. Die rechts abgetrennten Konstrukte betreffen den Sekundärdatensatz, sind daher mit höherer Vorsicht zu interpretieren ($N_{\text{sekundär}} = 305$).

ver: verständlich erklären; zus: Zusammenfassen & Wiederholen; verkn: Verknüpfungen herstellen; verstand: Verständnis überprüfen;
atmo: Gute Lernatmosphäre herstellen; rel: Relevanz verdeutlichen; swe: Ermutigen und Selbstwirksamkeit stärken; stoer: Effizient mit
Störungen umgehen; diag.g: Diagnostik: Grundeinstellung; diag.s: Diagnostik: Situation; skriptver: Skriptqualität für Versuchsaufbau;
skript: Skriptqualität; int: Integration der Vorlesungsinhalte in das Praktikum

Tabelle 13.3.: Ergebnisse der 1-Stichproben-t-Tests Tests zur Skalenausprägung in der PraQ-B-Gesamtstichprobe ($N = 305$). *padj&onetailed* wurde mittels BONFERRONI-HOLM-Korrektur (HOLM, 1979) und einseitig für gerichtete Hypothesen berechnet. Wichtige Ergebnisse sind fett gedruckt.

Test auf $M < 3.5$						
Hypothese	Konstrukt	M		SE	<i>padj&onetailed</i>	d
e.1	Konstrukte der <i>Lehrkompetenz</i> (außer e.2)–e.5), s. u.)	in [4.00; 4.58]	in [0.04; 0.05]		n.s.	-
f	Skriptqualität	4.15	0.06		n.s.	-
g	Skriptqualität für Versuchsablauf	4.24	0.06		n.s.	-
h	Integration	4.13	0.07		n.s.	-
explorativ (Hier: Test auf $M > 4.5$)						
e.2	verständlich erklären	4.69	0.04		< .001	0.21
e.3	Gute Lernatmosphäre herstellen	4.84	0.05		< .001	0.31
e.4	Effizient mit Störungen umgehen	4.86	0.04		< .001	0.40
e.5	Diagnostik: Grundeinstellung	4.88	0.05		< .001	0.34

padj&onetailed: BONFERRONI-HOLM-korrigierter p -Wert bei gerichteter Hypothese
 d : Effektstärke COHENS d

13.1.4. Diskussion

Die Studie 5 hat gezeigt, dass auf Basis von PraQ-Messungen ein Bild von Stärken und Schwächen von Praktika im deutschsprachigen Raum erhalten werden kann. Das Bild der Gesamtstichprobe von 18 Praktika mit 813 Teilnehmenden zum PraQ-A und 501 Teilnehmenden zum PraQ-B ergibt Stärken der Praktika in der Förderung von nur einem Kompetenzbereich der Studierenden, nämlich dem *Fachwissen* bezogen auf *praktische* Praktikumsinhalte, was sich gut mit den Ergebnissen von HUCKE (1999, S. 126, 132, 149) und dem Ziel 1 nach HALLER (1999) deckt (vgl. Abschnitt 3.4).

Dagegen fällt die *Interesse* der Studierenden als Schwäche auf, es besteht gemäß Analyse mindestens im affektiven Bereich also Handlungsbedarf, wie auch bereits im Abschnitt 5.1 von HUCKE (1999, S. 150) angesprochen.

Dies wird gestützt durch den Vergleich zum Benchmarking des BEvaKomp (BRAUN et al., 2008) im Bereich *intrinsischer Motivation*, die eine fast identische Skala verwendeten (vgl. Abschnitt 10.3.3 auf Seite 113). Dort sind in einer großen Studie ($N = 2507$) Mittelwerte der Skala im Bereich [3.5; 3.6] berichtet, die hier erfassten Praktika liegen im Mittel knapp darunter (eigene Berechnungen).

Auf Seite der *Lehrkompetenz* der Betreuenden ist dagegen positives zu berichten. Entgegen der Erwartungen können Betreuer*innen von Praktika im Mittel recht gut *verständlich erklären*, es gelingt ihnen, eine *gute Lernatmosphäre herzustellen*, sie zeigen sich effizient im *Umgang mit Störungen* und weisen eine *positive Grundeinstellung gegenüber der Diagnostik* des Lernfortschritts auf. Kein Bereich der *Lehrkompetenz* fällt als Schwäche auf.

Die *Lernumgebungskonstrukte* zeigen sich dagegen eher mittelmäßig.

Als Ergebnis für alle untersuchten Praktika lässt sich also feststellen, dass am ehesten Handlungsbedarf bei der Kompetenzförderung besteht, vor allem im affektiven Bereich.

Es konnte ebenso gezeigt werden, dass ein Praktikumsqualitätsprofil gebildet werden kann, dass Vergleiche zwischen einzelnen Praktika und in Bezug zur Gesamtstichprobe ermöglicht. Diese baseline der Gesamtstichprobe von 18 Praktika im deutschsprachigen Raum kann als Benchmarking (Anforderung 7, vgl. Abschnitt 10.3.3) für alle zukünftigen Qualitätsanalysen von Praktika dienen und stellt daher eine Errungenschaft für alle Praktika dar. Beispielsweise handelt es sich für den PraQ-A beim Praktikum mit dem niedrigsten Gesamtscore (weiß, Abb. 13.1 auf Seite 201) um ein Praktikum, in dem anscheinend die basale Beschäftigung mit den Geräten zu kurz kommt, der Erwerb *deklarativer Gerätekenntnisse* und der Fä-

higkeit, *Versuchsanordnungen aufzubauen* wird gering gefördert, wobei der sonstige Experimentierprozess wenig beeinträchtigt wird, die sonstigen Teilkompetenzen der *experimentellen Kompetenz* fallen durchschnittlich aus. Hier würden eine weitere Analyse und Intervention (z. B. mit einer Tiefenanalyse wie bei REHFELDT et al., 2015) also auf den Umgang mit Gerätschaften fokussieren.

Das Praktikum mit dem höchsten Gesamtscore zum PraQ-B (hellgrau, selbe Abb.) weist in seinem Qualitätsprofil ein überdurchschnittlich gutes Skript auf, bei gleichzeitig starker *Lehrkompetenz* der Betreuenden. Lediglich im affektiven (*Selbstwirksamkeitserwartung stärken*) und lebensweltlichen (*Relevanz verdeutlichen*) Bereich liegen hierfür eher durchschnittliche Ausprägungen (dunkelgrau, selbe Abb.) vor. Am ehesten könnte also dort eine Analyse und Intervention bezüglich der Betreuenden vollführt werden.

Einschränkend muss hier allerdings auf die Zusammensetzung der Stichproben eingegangen werden. Beide Stichproben weisen einen starken Überhang an Physikpraktika auf (75% bzw. 78% der Studierenden), weswegen die hier dargestellten Ergebnisse vor allem für diese Fachrichtung der Praktika gelten können, weitere Untersuchungen in Chemie- und sonstigen Praktika könnten langfristig auch dort das Bild schärfen und eine gruppengetrennte Auswertung ermöglichen (siehe auch Studie 6).

Darüber hinaus ist der Vergleich mit einem Skalennormwert (3.5 bzw. 4.5) zur Festlegung von Stärken und Schwächen der Praktika zwar praktikabel, aber kritisiert. Die eingesetzten Items der Skalen weisen nicht unbedingt dieselbe Itemschwierigkeit auf, weswegen ein Vergleich mit kriterial gesetzten Normwerten die Interpretation verzerren kann. Da sich in der Itemkonstruktion aber beim PraQ-A für Items, die nach Veränderungen fragen, entschieden wurde, kann zumindest dort das Argument teils entkräftet werden. Es handelt sich per theoriegeleiteter Itemkonstruktion um für Praktika relevante Veränderungen (Studie 1), weshalb eine Verneinung dieser normativ als schlechte Qualität gedeutet werden kann.

Für Konstrukte der *Lehrkompetenz* ist ein Vergleich mit Normwerten des LeKo (THIEL et al., 2012) fruchtbar. Die Normwerte des LeKo wurden zu den Dimensionen A, B und C (vgl. Abschnitt 4.5.1 auf Seite 49) und nicht zu den Einzelskalen numerisch veröffentlicht und stellen Benchmarks der Freien Universität Berlin mit einem $N = 735$ aus diversen Lehrveranstaltungen diverser Fachrichtungen dar (THIEL et al., 2012; BLÜTHMANN et al., 2017). Der Dimension A) Vermittlung von Wissen und Unterstützen von Verstehen sind die auch im PraQ-B fast unverändert verwendeten Skalen *verständlich erklären* und *Verknüpfungen herstellen* zugeordnet. Mit

13.2. Studie 6: Unterschiede zwischen Praktika der Chemie und der Physik

einem der PraQ-B-Metrik entsprechenden Mittelwert von $M = 4.86$ ($SE = 0.05$)⁶ für A) lässt sich für diese Skalen ein grober, da nicht skalenweiser Vergleich anstellen. Die Ergebnisse in Praktika zu *verständlich erklären* ($M = 4.69$ (0.04)) und Verknüpfungen herstellen ($M = 4.17$ (0.05)) liegen hierbei leicht bzw. deutlich unterhalb des LeKo-Benchmarks. Die Betreuer*innen in den untersuchten Praktika wären demnach nur durchschnittlich bzw. leicht unterdurchschnittlich in ihrer *Lehrkompetenz* zu bewerten.⁷

Innerhalb der Dimension B) (Benchmark $M = 4.98$ (0.05)) liegt die PraQ-B-Skala *Gute Lernatmosphäre herstellen*, die in Praktika einen Mittelwert von $M = 4.84$ (0.11) erreichte. Sie liegt damit im Rahmen des Standardfehlers innerhalb des Benchmarks. Interpretiert werden kann dies also als durchschnittliche Ausprägung.

Der Vergleich mit dem Benchmark liefert also die Erkenntnis, dass die auf den ersten Blick recht hohen Ausprägungen der *Lehrkompetenz* bei Betreuer*innen der Praktika durchaus eher durchschnittlich ausfallen, also noch Verbesserungspotential besteht.

In Bezug zur Darstellung besonders gelungener und besonders schwacher Praktikumsqualität (Abb. 13.1 und 13.2) lässt sich kritisieren, dass die Stichprobe mancher Praktika gering ausfällt ($N < 20$) und damit quantitative Vergleiche verzerrt ausfallen, was sich auch in großen Standardfehlern widerspiegelt. All dies lässt obige Ergebnisse mit der notwendigen Vorsicht interpretieren und stützt erneut die Notwendigkeit von auf den PraQ-Ergebnissen basierenden, tiefergreifenden Praktikumsanalysen.

13.2. Studie 6: Unterschiede zwischen Praktika der Chemie und der Physik

13.2.1. Einleitung

Zu Unterschieden der Lehrqualität zwischen Chemie- und Physikpraktika liegt bisher keine Studie, weder national noch international, vor, weswegen hier empirisches Neuland betreten wird. Aus den Eigenschaften von Physik- und Chemiepraktika (vgl. Abschnitt 3.4), sowie den empirisch erhobenen Charakteristika der Fachpraktika lassen sich allerdings einige Hypothesen bezüglich der Ausprägung der Qualitätskonstrukte des PraQ aufstellen. Die Hypothesen werden hier teils mit Vorgriff auf die Stichprobenbeschreibung aufgestellt.

⁶Berechnungsgrundlage der Skalierung von 7-stufiger auf 6-stufige Metrik: $f(x) = \frac{5}{6} \cdot x + \frac{1}{6}$.

⁷Die Skala *zur Selbsttätigkeit anregen* wurde bereits entfernt (vgl. Abschnitt 12.4.2.3 auf Seite 156).

Forschungsfragen und Hypothesen Die Studie 6 wird von folgender Forschungsfrage gerahmt:

Forschungsfrage 8: Was sind Unterschiede zwischen Chemie- und Physikpraktika in Bezug auf die Praktikumsqualität?

Hypothesen

- a) Die *Kooperationskompetenz* wird in Physikpraktika stärker gefördert als in Chemiepraktika, da erstere deutlich mehr Tandemarbeit etablieren (vgl. Stichprobe).
- b) Das *Geräte auswählen und Versuchsanordnung aufbauen* wird gemäß Forschungslage eher gering ausfallen, da diese Phasen auf Grund des gerade in der Physik oftmals bereits aufgebauten Versuches vernachlässigt werden (vgl. Argumentation Studie 5).
- c) Für Physikpraktika wird die *Integration der Vorlesung* geringer ausgeprägt sein, als für Chemiepraktika, da letztere deutlich häufiger eine geplante Begleitvorlesung etablieren (vgl. Stichprobe).
- d) Für Physikpraktika wird die Teilkompetenz des *Rohdaten graphisch und tabellarisch Darstellens* höher ausfallen als für Chemiepraktika, da erstere häufiger quantitative Ansätze der Auswertung in Grundpraktika nutzen.

13.2.2. Methode

Stichprobe Erneut wurde auf Erhebungsdaten der Studie 2 zurückgegriffen. Für diese Studie wurden solche Proband*innen ausgewählt, die entweder ein Physik- oder ein Chemiepraktikum besuchten.

PraQ-A: Für den PraQ-A bestand die Stichprobe aus $N = 563$ Studierenden in Physik- und $N = 93$ Studierenden in Chemiepraktika. Den Unterschied der Charakteristika zwischen Physik- und Chemiepraktika zeigt Tab. 13.4. Darin wird ersichtlich, dass die Chemiepraktika häufiger Blockpraktika anboten und die Physikpraktika nahezu ausschließlich Tandemarbeit an baugleichen Experimenten propagierten. Das durchschnittliche Fachsemester unterschied sich, ebenso wie die Geschlechterverteilung, gering. 47% der Teilnehmer*innen besuchten ein Physikpraktikum ohne selbst Physik zu studieren.

Tabelle 13.4.: Gegenüberstellung von Physik- und Chemiepraktika für die Stichprobe des PraQ-A. 75% aller Studierenden besuchte ein Physikpraktikum ($N = 563$), 13% ein Chemiepraktikum ($N = 93$).

Fachpraktikum	Block	Tandem	baugleich	fachfremd	weiblich	Semester
Physik	16%	99%	90%	47%	33%	3.0 (1.4)
Chemie	61%	50%	49%	0%	35%	3.2 (1.6)

Block: War es ein Blockpraktikum?

Tandem: Wurde in Tandems gearbeitet?

baugleich: Waren die Experimente baugleich mehrfach angeboten?

fachfremd: Anteil fachfremder Studierender im Fachpraktikum

weiblich: Anteil weiblicher Studierender

Semester: Fachsemester: M (SD)

PraQ-B: Für den PraQ-B ergaben sich $N = 368$ Studierende in Physik- und $N = 31$ Studierende in Chemiepraktika. Den Unterschied der Charakteristika zwischen Physik- und Chemiepraktika zeigt Tab. 13.5. Darin wird im Unterschied zur Stichprobe des PraQ-A deutlich, dass Chemiestudierende hier fast nie in Blockpraktika arbeiteten⁸. Ansonsten entsprechen die Charakteristika von Physik- und Chemiepraktika im Wesentlichen denen der Stichprobe des PraQ-A.

Speziell im PraQ-B relevant sind die Existenzfragen nach einem Praktikumsskript und einer Begleitvorlesung, da diese dort erforscht werden. Ein Skript scheint in beiden Fachpraktikagruppen verbreitet zu sein, dagegen ist für Chemiepraktika deutlich häufiger (83%) eine Begleitvorlesung vorgesehen als für Physikpraktika (44%, vgl. Hypothese c). Einschränkend ist hier auf die geringe Stichprobe für Chemiepraktika zu verweisen.

Instrumente Erneut wurde je Praktikum entweder der PraQ-A oder PraQ-B eingesetzt (vgl. Studie 2).

Auswertungsmethode Um die Forschungsfrage 2 adäquat beantworten zu können, sollte laut AERA et al. (2014, S. 13, 16) eine Messinvarianz zwischen Studierenden in Chemie- und Physikpraktika bezüglich der PraQ-Konstrukte nachgewiesen werden: Die Items des PraQ müssen in beiden Gruppen je die gleiche Rolle in Bezug auf das zu messende Konstrukt einnehmen, also in Bezug auf das jeweilige Konstrukt strukturell gleich verstanden werden. Statistisch spricht man von Messinvarianz bzw. bei unterschiedlichem »Funktionieren« von Items von DIF (differential item functioning).

⁸Die 3% sind anzuzweifeln, da diese Studierenden sich in denselben zwei Praktika befanden wie die verbliebenen 97%

Tabelle 13.5.: Gegenüberstellung von Physik- und Chemiepraktika für die Stichprobe des PraQ-B. 74% der Studierenden besuchten ein Physikpraktikum ($N = 368$), 6% ein Chemiepraktikum ($N = 31$). Fehlende Werte wurden in die prozentuale Berechnung miteinbezogen.

Fachpr	Block	Tandem	baugleich	fachfremd	Skript	B-VL	weibl	Sem
Physik	15%	93%	96%	43%	84%	42%	33%	3.0 (1.4)
Chemie	4%	47%	39%	14%	65%	78%	35%	3.2 (1.6)

Fachpr: Fachpraktikum

Block: War es ein Blockpraktikum?

Tandem: Wurde in Tandems gearbeitet?

baugleich: Waren die Experimente baugleich mehrfach angeboten?

Skript: Gab es ein Skript?

B-VL: Existierte eine Begleitvorlesung?

fachfremd: Anteil fachfremder Studierender im Fachpraktikum

weibl: Anteil weiblicher Studierender

Sem: Fachsemester: M (SD)

Auf Grund der gerade in Chemiepraktika mittelgroßen bis geringen Stichprobe wurde sich für das MIMIC-Verfahren entschieden (benötigt $N_{ges} = 100\text{--}150$, BROWN, 2006, S. 305), mit dem DIF untersucht werden kann und nicht für das klassische Verfahren der stufenweisen Analyse der Messinvarianzen (benötigt $N = 100\text{--}150$ je Gruppe, BROWN, 2006, S. 305). Nachteil an diesem Verfahren ist die automatische Annahme, dass Faktorladungen (Item-Konstrukt-Beziehung), Fehler(ko)varianzen (Item-Item-Beziehung) und Faktor(ko)varianzen (Faktor-Faktor-Beziehung) über Gruppen hinweg gleich sind (BROWN, 2006, S. 305ff), was es in multi-group-Verfahren eigentlich erst nachzuweisen gilt. Das MIMIC-Verfahren bietet darüber hinaus aber den Vorteil, dass es für ungleiche Gruppengrößen bessere Schätzungen liefert als das klassische multi-group-Verfahren (BROWN, 2006, S. 305ff).

Kern der MIMIC-Analyse sind Regressionen der Gruppenvariablen (hier: dichotom Chemie/Physik) auf die Itemausprägungen (DIF-Analyse) und auf die Faktoren (Gruppenunterschiede). Für die DIF-Analyse wird bei plausiblen inhaltlichen Argumenten der Einfluss der Gruppenvariablen auf das betreffende Item zugelassen (»befreit«), es handelt sich dann bei der betreffenden Skala nur noch um partielle Messinvarianz. Erweitert werden kann die Analyse um relevante Kovariaten, für deren Einfluss dann in der Bestimmung der Gruppenunterschiede kontrolliert werden kann. Untersucht wurden Einflüsse des Fachsemesters der Studierenden, ihres selbst berichteten *theoretischen* und *praktischen Vorwissens*, des Geschlechts der Studierenden und des Fachsemesters.

Durchführung Details zur Durchführung sind den Studien 2.1 und 2.2 zu entnehmen.

13.2.3. Ergebnisse

PraQ-A: Wird die MIMIC-Analyse lediglich unter Einbezug des Einflusses der Gruppenvariablen vorgenommen, ergibt sich für die Gruppenunterschiede das Ergebnis in Abb. 13.3, die DIF-Analyse fasst Tab. 13.6 zusammen. Das finale Modell unter Freisetzen der unten beschriebenen DIF-Parameter weist einen guten Modellfit ($CFI = .93$, $TLI = .92$, $RMSEA = .04$, $SRMR = .04$) ohne inhaltlich begründbare standardisierte $EPCs > .20$ auf.

Die MIMIC-Ergebnisse zeigen fünf Faktoren, die systematische Gruppenunterschiede zwischen Chemie- und Physikpraktika aufweisen. In Chemiepraktika fällt das Konstrukt *EXP: Gerätekenntnisse deklarativ* mit kleinem Effekt ($d = 0.27$) höher aus. In Physikpraktika fallen die Konstrukte der *Experimentellen Kompetenz, Probleme lösen* ($d = -0.37$), Rohdaten aufbereiten ($d = -0.36$) und Messungen dokumentieren ($d = -0.25$) höher aus, mit kleiner bis mittlerer Effektstärke. Des Weiteren ist die Ausprägung des Konstrukts *Kommunikationskompetenz schriftlich* in Physik- höher ausgeprägt als in Chemiepraktika ($d = -0.37$).

Die DIF-Analyse ergab für insgesamt vier Items einen DIF, der in das Modell implementiert wurde, also dort die Fixierung des Gruppen-Regressionsparameters auf 0 aufgehoben und der Parameter geschätzt wurde.

Die Schätzergebnisse zeigt Tab. 13.6 in der Spalte »DIF«, die Metrik des DIFs entspricht der Metrik der Likert-Skala, da es sich um unstandardisierte Regressionsparameter handelt. Der höchste und zweithöchste DIF ergaben sich für die Items »... erkenne ich besser mögliche Fehlerquellen des Experiments.« und »... beachte ich experimentelle Fehlerquellen adäquater.«, beide aus dem Faktor *Experimentelle Kompetenz: mit Problemen und Fehlern umgehen* mit ca. einem halben (.47) bzw. einem drittel (.31) Skalenpunkt Unterschied zugunsten von Physikpraktika. Ebenfalls mit ca. einem drittel bis einem viertel Skalenpunkt weisen die Items »... kann ich meine experimentelle Arbeit besser zeitlich organisieren.« (.31) des Faktors *Zeitmanagement* und »... erkenne ich besser, welches Gerät ich vor mir habe.« (-.27) des Faktors *Experimentelle Kompetenz: Gerätekenntnisse deklarativ* einen DIF auf.

Für die betreffenden Faktoren liegt daher durch das Freisetzen der entsprechenden Parameter lediglich eine partielle Messinvarianz vor. Für weitere zwei Items lag in der modificationindex-Analyse ein DIF-relevanter standardisierter EPC von $> .20$ (SARIS et al., 2009) vor. Für diese Items konnten inhaltlich allerdings keine

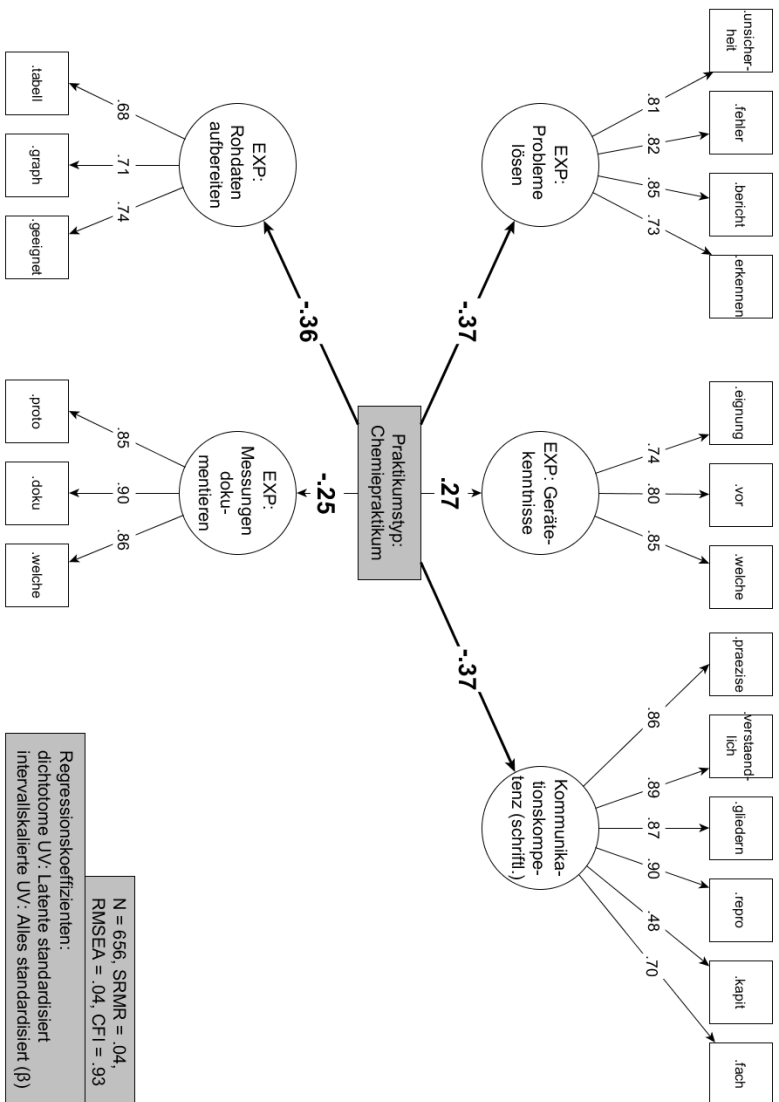


Abbildung 13.3.: Signifikante Ergebnisse der MIMIC-Analyse zu Gruppenunterschieden Physik-/Chemiepraktika zum PraQ-A. Regressionskoeffizienten > 0 indizieren eine höhere Ausprägung des betroffenen Faktors für Chemiepraktika. Die für die Latente standardisierten Regressionskoeffizienten können als Effektstärke COHENS d interpretiert werden (BROWN, 2006, S. 314ff). Ladungen werden standardisiert dargestellt. Korrelationen zwischen Faktoren werden zwecks Übersichtlichkeit nicht dargestellt.

Tabelle 13.6.: Ergebnisse zur DIF-Analyse der Items des PraQ-A bezüglich der Gruppen Chemie-/Physikpraktika. Von insgesamt 70 untersuchten Items in 17 Faktoren traten sechs substanzielle DIFs auf. Die unstandardisierten DIF-Werte können in der Metrik der Faktoren, also Punkten auf einer Likertskala von 1 bis 6, interpretiert werden (BROWN, 2006, S. 314ff). Positive Vorzeichen entsprechen einer systematisch höheren Itemausprägung in Chemiepraktika.

MIMIC-Analyse PraQ-A: DIF			
Konstrukt	Itemwortlaut	DIF	SE
EXP: mit Problemen und Fehlern umgehen	...erkenne ich besser mögliche Fehlerquellen des Experiments.	0.47***	0.10
	...beachte ich experimentelle Fehlerquellen adäquater.	0.31***	0.09
Zeitmanagement	...kann ich meine experimentelle Arbeit besser zeitlich organisieren.	0.31***	0.10
	EXP: Gerätekenntnisse deklarativ	-0.27**	0.10

*** $p < .001$, ** $p < .01$

EXP: *Experimentelle Kompetenz* (PraQ)

plausiblen Argumente gefunden werden, warum diese in Chemie oder Physik stärker ausfallen sollten.⁹

MIMIC-Analyse mit Kovariaten: Bezieht man nun die oben beschriebenen Kovariaten mit in die Analyse ein und kontrolliert für diese (vgl. Abb. 13.4), so ergeben sich nur für die Kovariate *Vorwissen (Theorie)* Effekte (Fit: $CFI = .93$, $TLI = .92$, $RMSEA = .04$, $SRMR = .04$).¹⁰

Der in der MIMIC-Analyse signifikante und mit kleiner Effektstärke auftretende Einfluss der Gruppenzugehörigkeit auf den Faktor *EXP: Gerätekenntnisse deklarativ* wird nach Hinzufügen der Kovariaten *Vorwissen (Theorie)* nichtsignifikant. Eine Analyse des direkten Effekts des Praktikumsfaches auf die Kovariate *Vorwissen (Theorie)* zeigt einen signifikanten, großen Effekt. Es handelt sich also um eine Mediation des Effekts über die Kovariate *Vorwissen (Theorie)*.¹¹

Zudem ergibt sich ein neuer Regressionspfad des Gruppenunterschieds, sowie des *Vorwissens* auf den Faktor *Bewertungskompetenz*.¹²

Für den Einfluss des Praktikumsfaches ergeben sich ansonsten ähnlich große und gleich gerichtete Effekte wie im MIMIC-Modell ohne Kovariaten. Diese Faktoren fallen allesamt in Physik systematisch höher aus, mit je kleinem bis mittlerem Effekt. Das *theoretische Vorwissen* wirkt sich als Kovariate systematisch auf alle beschriebenen Faktoren des PraQ-A aus und bewirkt stets einen positiven Effekt auf die Faktorenausprägung mit etwa einer fünftel (.19) bis etwa einer drittel (.29) Standardabweichung je Erhöhung des Vorwissens um eine Standardabweichung.

PraQ-B: Für die Analyse der Faktoren und Items des PraQ-B ergibt sich für die Gruppenunterschiede zwischen Physik- und Chemiepraktika das Ergebnis in Abb. 13.5, die DIF-Analyse fasst Tab. 13.7 zusammen. Das Modell unter Freisetzen der unten beschriebenen DIF-Parameter weist einen guten Modellfit ($CFI = .93$, $TLI = .92$, $RMSEA = .06$, $SRMR = .04$) ohne inhaltlich begründbare standardisierte *EPCs* $> .20$ auf.¹³

⁹Dies waren die Items #fwtheo.typisch »...sehe ich mich nun in der Lage, eine typische Fragestellung des behandelten Gegenstandsbereiches zu bearbeiten.« und #intr.ausserhalb »...beschäftige ich mich aus Spaß außerhalb der Hochschule mit dem Gegenstandsbereich.«

¹⁰Das nur je einmalige Auftreten von kleinen Effekten des Geschlechts und des Alters wird hier nicht näher betrachtet. Interaktionseffekte zwischen der Gruppenzugehörigkeit zu Chemie- oder Physikpraktika und dem *theoretischen Vorwissen* waren nichtsignifikant.

¹¹Der unstandardisierte indirekte Effekt beträgt $0.51 \cdot 0.18 = 0.09$. Betrachtet man in diesem Modell trotz Nichtsignifikanz den unstandardisierten direkten Effekt mit 0.17, so ergibt sich wieder der aus dem einfachen MIMIC-Modell bekannte unstandardisierte Effekt von $0.09 + 0.17 = 0.26$.

¹²Es kann sich hier um einen Suppressionseffekt (EID, 2013, S. 594ff) handeln.

¹³Die reinen MIMIC-Analysen des PraQ-B der Studien 6 und 7 entstanden nach und hoch analog zu den Analysen des PraQ-A in Kooperation mit Daniel Becker in 2016 im Rahmen einer Masterarbeit.

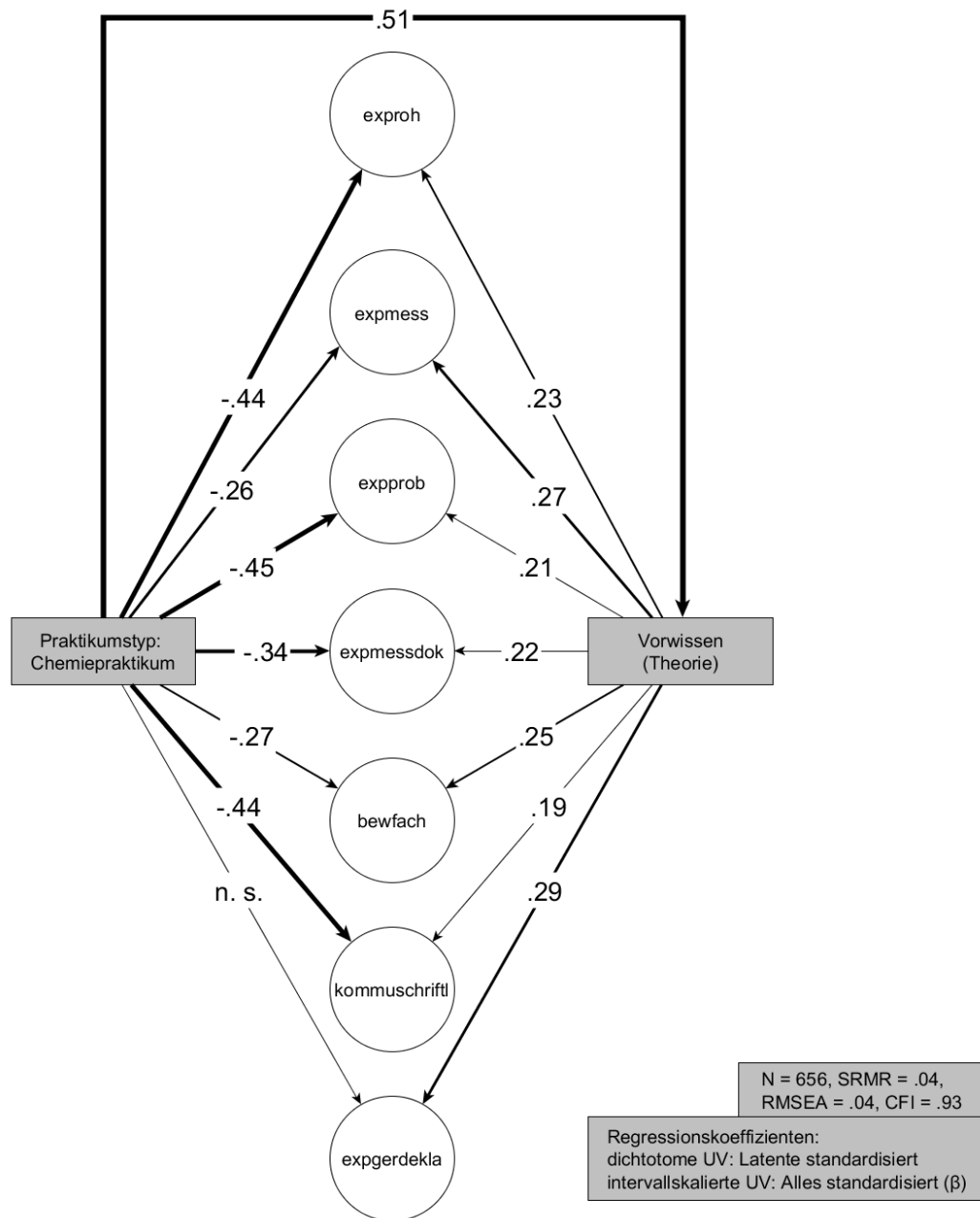


Abbildung 13.4.: MIMIC-Modell zu Gruppeneffekten Chemie-/Physikpraktika des PraQ-A mit Kovariaten. Die von dichotomen Variablen ausgehenden Regressionspfade sind berechnet für die standardisierte latente Variable und daher als Effektstärke d zu interpretieren. Regressionskoeffizienten > 0 indizieren eine höhere Ausprägung des betroffenen Faktors für Chemiepraktika. Die von der Variablen *Vorwissen (Theorie)* ausgehenden Regressionspfade sind für beide Seiten standardisiert berechnet und daher wie z -Werte in Standardabweichungen zu interpretieren. Korrelationen zwischen Faktoren werden zwecks Übersichtlichkeit nicht dargestellt.

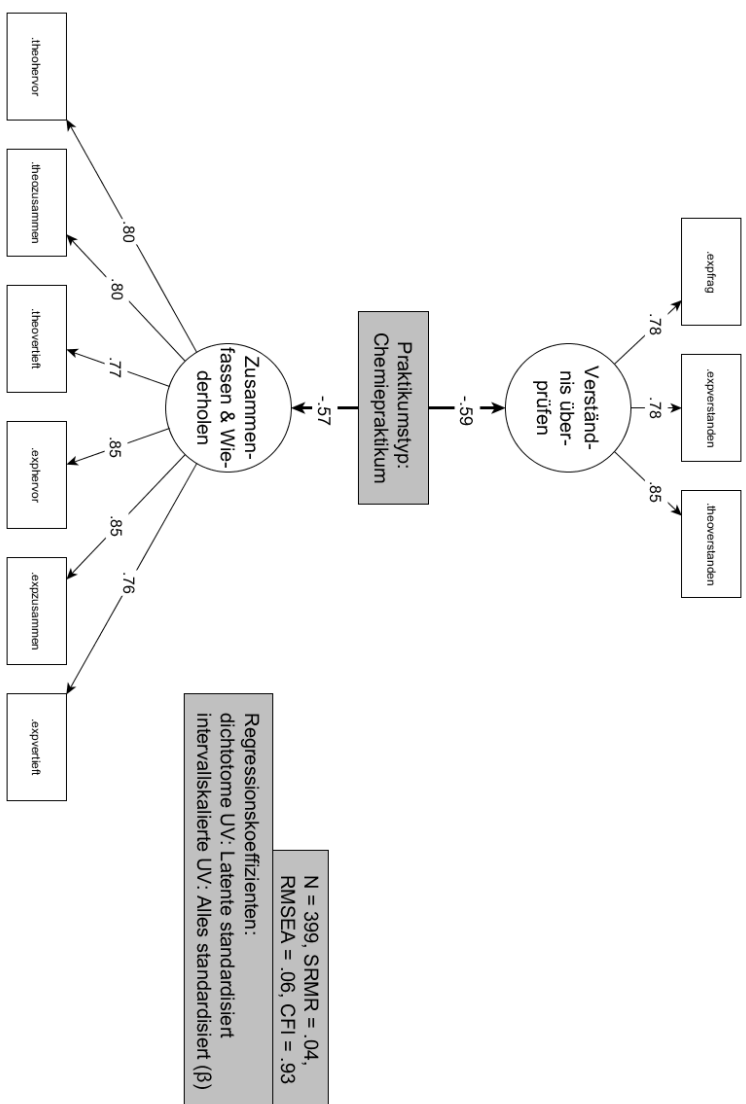


Abbildung 13.5.: Signifikante Ergebnisse der MIMIC-Analyse zu Gruppenunterschieden Physik-/Chemiepraktika zum PraQ-B. Regressionskoeffizienten > 0 indizieren eine höhere Ausprägung des betroffenen Faktors für Chemiepraktika. Die für die Latente standardisierten Regressionskoeffizienten können als Effektstärke COHENS d interpretiert werden (BROWN, 2006, S. 314ff). Ladungen werden standardisiert dargestellt. Korrelationen zwischen Faktoren werden zwecks Übersichtlichkeit nicht dargestellt.

Die MIMIC-Ergebnisse zeigen lediglich zwei Faktoren, die signifikante systematische Gruppenunterschiede zwischen Chemie- und Physikpraktika aufweisen. Diese fallen je zu Gunsten von Physikpraktika aus. So sind die Faktoren *Verständnis überprüfen* und *Zusammenfassen und Wiederholen* je mit mittlerer Effektstärke ($d = -.57$ bzw. $-.59$) in Physikpraktika signifikant höher ausgeprägt als in Chemiepraktika.

Die DIF-Analyse ergab für insgesamt vier Items einen DIF, der in das Modell implementiert wurde, also dort die Fixierung des Gruppen-Regressionsparameters auf 0 aufgehoben und der Parameter geschätzt wurde. Die Schätzergebnisse zeigt Tab. 13.7. Alle DIFs entsprechen ca. einem Drittel Skalenpunkt Unterschied zu Gunsten von Chemiepraktika. Dabei fallen die Items dadurch auf, dass sie denselben Faktor, *Lehrkompetenz: Zusammenfassen & Wiederholen* adressieren.¹⁴

MIMIC-Analyse mit Kovariaten: Bezieht man die bereits beim PraQ-A vorgestellten Kovariaten mit in die Analyse mit ein, so ergeben sich erneut nur für die Kovariate *Vorwissen (theoretisch)* je sehr kleine Effekte. Für die in der MIMIC-Analyse gefundenen statistischen Beziehungen bedeutet dies hier keine Veränderung (vgl. Abb. 13.6). Mediationseffekte und Interaktionen traten nicht auf.

13.2.4. Diskussion

PraQ-A:

Diskussion des Messmodells über Gruppen Für den PraQ-A lag nur für wenige Items ein systematischer Unterschied in ihrer Messrelevanz für ihren Faktor zwischen den Gruppen der Chemie- und Physikpraktika vor, was für eine verlässliche Messung spricht. Für die Skala der *Experimentellen Kompetenz: mit Problemen und Fehlern umgehen* ergaben sich zwei DIF-Items. Erklärt werden kann das unterschiedliche Funktionieren dieser Items dadurch, dass erstens der Begriff »Fehler« in der Physik stärker auch mit numerischen Messunsicherheiten assoziiert wird.¹⁵

Da zweitens das Auffinden numerisch relevanter Fehlerquellen im Experiment potentiell abstrakter ausfällt als das Auffinden praktischer Fehlerquellen, könnte dies ein Grund für die vergleichsweise höhere Ausprägung dieser Items in Chemiepraktika sein. Beide Items sollten für eine zukünftige PraQ-Version revidiert werden, die

¹⁴Für das Item »... geht angemessen und effizient mit Unruhe und Störungen um.« des Faktors *Effizient mit Störungen umgehen* ergab sich statistisch ein DIF, der aber klein und nicht erklärbar ausfällt. Es könnte sich hier um ein Artefakt der geringen Stichprobe der Chemiepraktika oder der hohen Standardfehler handeln (vgl. Methodenkritik).

¹⁵In OC-Praktika ist dies nur sehr selten der Fall, selbst bezüglich PC-Praktika erläuterte der Praktikumsleiter in Studie 1.1, dass dort eine numerische Fehlerbetrachtung meist nur angebahnt werde.

Tabelle 13.7.: Ergebnisse zur DIF-Analyse der Items des PraQ-B bezüglich der Gruppen Physik- und Chemiepraktika. Von insgesamt 36 untersuchten Items in 11 Faktoren traten vier substanziale DIFs auf. Die umstandardisierten DIF-Werte können in der Metrik der Faktoren, also Punkten auf der angesprochenen Likertskala von 1 bis 6, interpretiert werden (BROWN, 2006, S. 314ff). Positive Vorzeichen entsprechen einer höheren Itemausprägung in Chemiepraktika. Die Faktoren *Skriptqualität* und *Integration* konnten wegen hoher Anteile fehlender Werte nicht in die Analyse miteinbezogen werden (vgl. Studien 2.2 und 5).

MIMIC-Analyse PraQ-B: DIF				
Konstrukt	Itemwortlaut	DIF		SE
Zusammenfassen & Wiederholen	... erkenne ich besser mögliche Fehlerquellen des Experiments.	0.37*		0.18
Zusammenfassen & Wiederholen	... beachte ich experimentelle Fehlerquellen adäquater.	0.37*		0.16
Zusammenfassen & Wiederholen	... kann ich meine experimentelle Arbeit besser zeitlich organisieren.	0.32*		0.16

* $p < .05$

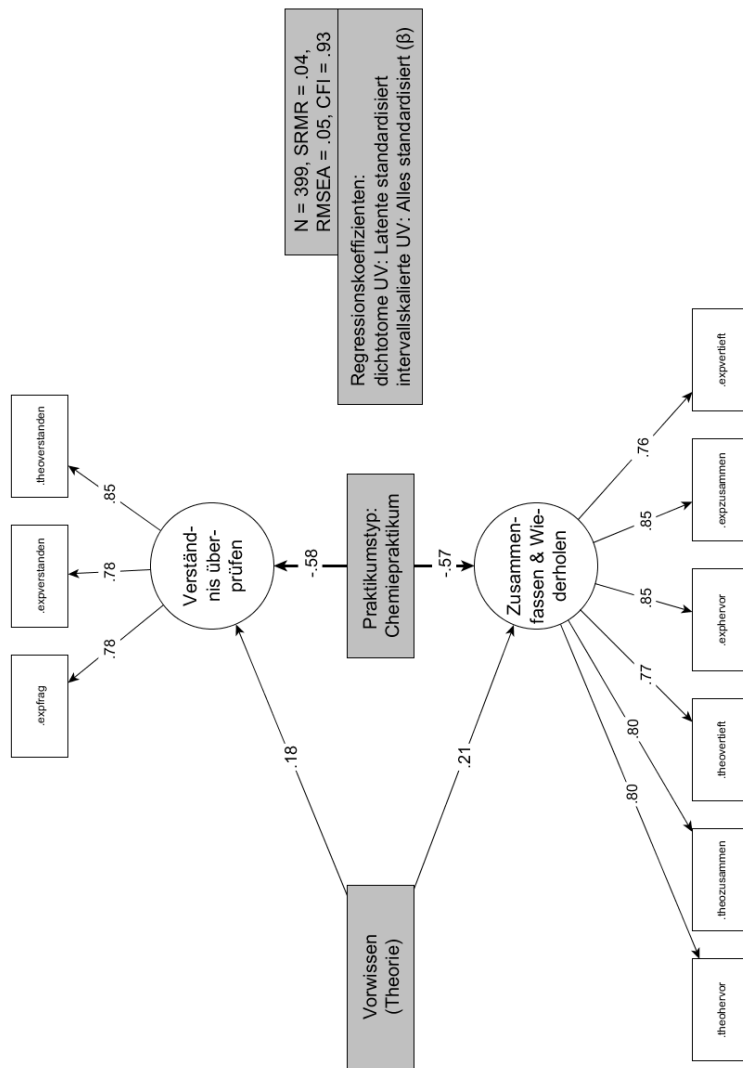


Abbildung 13.6.: MIMIC-Modell zu Gruppeneffekten Chemie-/Physikpraktika des PraQ-B mit Kovariaten. Die von dichotomen Variablen ausgehenden Regressionspfade sind berechnet für die standardisierte latente Variable und daher als Effektstärke d zu interpretieren. Regressionskoeffizienten > 0 indizieren eine höhere Ausprägung des betroffenen Faktors für Chemiepraktika. Die von der Variablen *Vorwissen (Theorie)* ausgehenden Regressionspfade sind für beide Seiten standardisiert berechnet und daher wie z -Werte in Standardabweichungen zu interpretieren. Korrelationen zwischen Faktoren werden zwecks Übersichtlichkeit nicht dargestellt.

Bezeichnung »Unsicherheit« bietet sich hier statt »Fehler« als fächerunabhängiger an.

Zudem wiesen die Items »... kann ich meine experimentelle Arbeit besser zeitlich organisieren.« der Skala *Zeitmanagement* und »... erkenne ich besser, welches Gerät ich vor mir habe.« der Skala *Experimentelle Kompetenz: Gerätekenntnisse deklarativ* einen DIF auf. Ersteres könnte in Chemiepraktika daher höher ausfallen, da dort häufiger freie Zeitfenster für das Experimentieren vergeben werden, die eine experimentelle Arbeitsorganisation notwendiger machen. Letzteres könnte deshalb in Physikpraktika höher ausfallen, da dort potentiell weniger ähnlich aussehende Geräte verwendet werden, die eine Verwechslungsgefahr bergen.

Für obige Faktoren liegt daher durch das Freisetzen der entsprechenden Parameter lediglich eine partielle Messinvarianz vor, gemäß BROWN (2006, S. 314) ist dies allerdings für eine adäquate Interpretierbarkeit akzeptabel, wenn wenigstens noch ein verbleibendes Item ohne DIF je Skala vorliegt, was der Fall ist. Es bedeutet aber auch, dass diese Skalen bei der Interpretation von Unterschieden zwischen Chemie- und Physikpraktika vorsichtiger interpretiert werden sollten.

Diskussion von Gruppenunterschieden Folgende Hypothesen bezüglich der Konstrukte des PraQ-A haben sich im Rahmen des Ergebnisteils bestätigt:

- d) Für Physikpraktika wird die Teilkompetenz des *Rohdaten graphisch/tabellarisch Darstellens* höher ausfallen als für Chemiepraktika, da erstere häufiger quantitative Ansätze der Auswertung in Grundpraktika nutzen.

In der Analyse zeigte sich ein mittlerer Effekt zugunsten von Physikpraktika. Interpretiert werden kann dies so, dass es Physikpraktika besser als Chemiepraktika gelingt, diese Teilkompetenz in der Wahrnehmung der Studierenden zu fördern. Hier kann von den Physikpraktika gelernt werden, indem die Lernprozesse in Praktika hinsichtlich dieses Aspekts genauer untersucht werden.

Folgende Hypothesen sind dagegen auf Basis der Ergebnisse zunächst abzulehnen:

- a) Die *Kooperationskompetenz* wird in Physikpraktika stärker gefördert als in Chemiepraktika, da erstere deutlich mehr Tandemarbeit etablieren.
- b) Das *Geräte auswählen und Versuchsanordnung aufbauen* wird gemäß Forschungslage eher gering ausfallen, da diese Phasen auf Grund des gerade in der Physik oftmals bereits voraufgebauten Versuches vernachlässigt werden.

Für beide Hypothesen ergaben sich in der Analyse keinerlei signifikante Gruppenunterschiede. Chemie- und Physikpraktika unterscheiden sich also im Rahmen dieser Erhebung nicht voneinander, was die Förderung kooperativen Verhaltens oder

der Fähigkeit zur Geräteauswahl und zum Versuchsaufbau betreffen. Dies kann als Argument dafür gedeutet werden, dass auch die nur zu etwa 50% in Tandems arbeitenden Studierenden in Chemiepraktika kooperative Fähigkeiten, vielleicht in der Interaktion mit anderen Studierenden oder den Betreuenden, erwerben.

Darüber hinaus fördern Chemiepraktika scheinbar als Ergebnis der MIMIC-Analyse besser den Aufbau von *deklarativem Wissen über Versuchsgeräte*. Das Entfallen dieses Effekts nach Einbezug des *theoretischen Vorwissens* in das Modell lässt allerdings eher den Schluss zu, dass die tatsächlichen Unterschiede geringer bis nichtig ausfallen. Studierende mit höherer Einschätzung ihres *theoretischen Vorwissens* bewerten auch ihren Fortschritt im Erwerb *deklarativer Gerätekenntnisse* tendenziell besser – und dies ohne einen messbaren zusätzlichen Effekt des Praktikumsfaches.

Der in der MIMIC-Analyse festgestellte Gruppenunterschied bezüglich *deklarativer Gerätekenntnisse* kann also vollständig auf natürlich gegebene Vorwissensunterschiede der Stichproben zurückgeführt werden. Dieses Ergebnis ist allerdings auf Basis der nur partiellen Messinvarianz des Konstrukts *Experimentelle Kompetenz: Gerätekenntnisse deklarativ* unter Vorbehalt zu verstehen. Für zukünftige Evaluationen dieses Konstrukts sollte aber unbedingt für das *theoretische Vorwissen* kontrolliert werden.

Physikpraktika dagegen besitzen zusätzlich vergleichsweise ihre Stärken in der Förderung von *experimentellem Problemlösen*¹⁶, *schriftlichem Kommunizieren* und eine leichte Stärke im Erlernen einer guten *Dokumentation von Messergebnissen*. Gerade letztere beide spiegeln den starken Fokus der Physikpraktika auf schriftlicher Versuchsdokumentation in Form von Protokollen wider. Für alle Konstrukte, die je Praktikumsfach unterschiedlich ausfallen, gibt es ebenso Effekte des selbst-eingeschätzten *theoretischen Vorwissens*.

Dies bedeutet hier, dass Studierende mit höher eingeschätztem *theoretischem Vorwissen* ihren Kompetenzzuwachs in den beschriebenen Konstrukten systematisch höher einschätzen. Diese Effekte fallen teils bedeutsam aus, weswegen für zukünftige Evaluationserhebungen diese Kovariate stets mit betrachtet werden sollte.

PraQ-B

Diskussion des Messmodells über Gruppen Bezüglich der Item-DIFs sind zunächst die äußerst hohen Standardfehler hervorzuheben, die eine Interpretation der DIF-Ausprägung erschweren. Teilweise handelt es sich um Standardfehler in der Größenordnung von 50% des DIF-Wertes selbst.

¹⁶ Auch hier ist das Ergebnis auf Grund partieller Messinvarianz aber vorsichtig zu werten.

Beim PraQ-B fallen alle Items mit systematischen Gruppenunterschieden dadurch auf, dass sie dasselbe Konstrukt, *Lehrkompetenz: Zusammenfassen & Wiederholen* adressieren und in Chemiepraktika höher ausfallen. Die höhere Ausprägung dieser Items in Chemiepraktika kann dadurch begründet werden, dass sie im Vergleich zu den sonstigen Items dieses Konstrukts sich auf experimentelle, nicht auf fachtheoretische, Aspekte beziehen.

Da in Chemiegrundpraktika oft eine offenere Experimentierumgebung in Form von Großraumlabor vorliegt, werden Betreuende ggf. mehr Akzentuierungen und Zusammenfassungen anbieten, da der Betreuungsschlüssel ungünstiger ausfällt. Die Skala *Lehrkompetenz: Zusammenfassen & Wiederholen* kann auf Grund des dreifachen DIFs weder inhaltlich noch statistisch plausibel zwischen Chemie- und Physikpraktika verglichen werden.

Diskussion von Gruppenunterschieden Folgende Hypothese ist auf Basis der Ergebnisse abzulehnen:

- c) Für Physikpraktika wird die Integration der Vorlesungsinhalte geringer ausgeprägt sein, als für Chemiepraktika, da letztere deutlich häufiger eine geplante Begleitvorlesung etablieren.

Für die *Integration* ergaben sich keinerlei Unterschiede zwischen Chemie- und Physikpraktika. Darüber hinaus traten aber Unterschiede in anderen Faktoren auf: In Physikpraktika *überprüfen* die Betreuenden auswertungsgemäß regelmäßiger das *Verständnis* der Studierenden als in Chemiepraktika.¹⁷

Der kleine Effekt des *theoretischen Vorwissens* auf dieses Konstrukt indiziert eine wie schon beim PraQ-A systematisch höhere, wenngleich nur wenig bedeutsam höhere Einschätzung der betreffenden Fähigkeit der/des Praktikumsbetreuenden, wenn höheres selbsteingeschätztes *theoretisches Vorwissen* der evaluierenden Studierenden vorlag. Zukünftig ist eine Kontrolle bezüglich dieses Faktors beim PraQ-B also optional. Es fällt auf, dass Physikpraktika in einigen Konstrukten sowohl der *Lernzuwachs*-, als auch der *Lehrkompetenz*-Dimension besser abschneiden als Chemiepraktika. Es böte sich also auch hier an, die Lernprozesse in Physikpraktika bezüglich der günstiger ausfallenden Konstrukte genauer zu untersuchen und die Erkenntnisse auch Chemiepraktikabetreibenden zukommen zu lassen.

Methodenkritik Obschon dies die erste standortübergreifende Studie zu Unterschieden zwischen Chemie- und Physikpraktika ist, gibt es einige Einschränkungen

¹⁷ Der Unterschied beim *Zusammenfassen und Wiederholen* wird der Argumentation in der Auswertung gemäß nicht interpretiert.

13.3. Studie 7: Unterschiede zwischen Block- und Semesterpraktika

obiger Interpretationen der Ergebnisse. So zeigen sich die Stichproben zu den jeweiligen Erhebungsinstrumenten PraQ-A und PraQ-B heterogen. Dies beginnt mit der Stichprobengröße der Chemiepraktika, die beim PraQ-A mit $N = 93$ ungleich höher ausfällt als beim PraQ-B mit $N = 31$.

Bei letzterem sind es nur zwei Lehrveranstaltungen, die in die Stichprobe eingingen. Es ergeben sich in diesem Rahmen zudem deutliche Stichprobenunterschiede bei den Studierenden der Chemiepraktika, da sie beim PraQ-B fast nie in Blockpraktika arbeiteten, was aber gemäß anschließender Studie 7 durchaus ein potenter Prädiktor für Praktikumsqualität sein kann.¹⁸

Zudem ist die geringe Stichprobe des PraQ-B für eine MIMIC-Analyse zwar ausreichend, da verfahrensgemäß die gesamte Stichprobe von $N = 368$ für das Verfahren zählt, aber dennoch mit hohen Standardfehlern (siehe Auswertungsmethode), also geringer Messgenauigkeit, belegt. Dies erschwert die adäquate Interpretation von DIF und Gruppenunterschieden.

Zuletzt lässt sich erneut das MIMIC-Verfahren zur Analyse von Gruppenunterschieden selbst kritisieren, da es automatisch Faktorladungen (Item-Konstrukt-Beziehung), Fehler(ko)varianzen (Item-Item-Beziehung) und Faktor(ko)varianzen (Faktor-Faktor-Beziehung) über Gruppen hinweg gleich setzt (BROWN, 2006, S. 305ff). Eine »echte« Messinvarianz über die Gruppen der Studierenden von Chemie- und Physikpraktika gilt es also noch nachzuweisen. Auch die statistisch bedeutsame Kovariate *Vorwissen (Theorie)* ist als Selbsteinschätzungsitem kritisch zu sehen, u.a. da hier keine Reliabilität geschätzt werden konnte.

13.3. Studie 7: Unterschiede zwischen Block- und Semesterpraktika

13.3.1. Einleitung

Die Lehrveranstaltungsform »Praktikum« wird klassischerweise auch manchmal als Blockveranstaltung angeboten. Dies wird meist in der vorlesungsfreien Zeit realisiert und dieselben Versuche implementiert, wie im Semester-Pendant. Es liegen allerdings noch keinerlei Hinweise oder Befunde vor, die einen Einfluss dieser anderen Organisationsform auf die Praktikumsqualität erörtern.

Es ist bisher unklar, ob die Vorteile einer starken Fokussierung auf die Praktikumsarbeit über mehrere ganze Tage hinweg überwiegen oder eher die Nachteile, beispielsweise einer Verdichtung des Präsenzzeit-Heimarbeits-Verhältnisses.

¹⁸Eine gemeinsame Betrachtung der Gruppenvariablen *Praktikumsfach* und *Praktikumstyp* war statistisch auf Grund daraus folgender noch weiter reduzierter Stichproben je Subgruppe nicht ratsam.

Forschungsfragen und Hypothesen Die Forschungsfrage dieser Studie lautet daher:

Forschungsfrage 9: Was sind Unterschiede zwischen Block- und Semesterpraktika in Bezug auf die Praktikumsqualität?

Hypothesen Da in Semesterpraktika mehr Heimarbeitszeit für die theoretische Einarbeitung vor dem Praktikumsversuch und die Nachbereitung mit Auswertung, Protokollerstellung und Reflexion besteht, werden diejenigen Konstrukte, die diese Phasen tangieren, vermutlich dort höher ausfallen.

- a) Das Konstrukt *Experimentelle Kompetenz: Rohdaten graphisch, tabellarisch darstellen* wird in Semesterpraktika stärker gefördert als in Blockpraktika.
- b) Das Konstrukt *Experimentelle Kompetenz: Daten interpretieren* wird in Semesterpraktika stärker gefördert als in Blockpraktika.
- c) Das *theoretische Fachwissen* wird in Semesterpraktika stärker gefördert als in Blockpraktika.
- d) Die *schriftliche Kommunikation* verbessert sich mehr in Semesterpraktika als in Blockpraktika.

Durch den höheren Zeitdruck für die oft notenrelevanten Protokolle in Blockpraktika könnte sich dort zudem ein negativer Effekt auf die *Interesse* ergeben, da diese durch *Arbeitsbelastung* und *Erfolgsdruck* negativ beeinflusst ist (DECI, 1993, S. 233):

- e) Die Förderung *intrinsischer Motivation* gelingt in Semesterpraktika besser als in Blockpraktika.

Auf der anderen Seite spricht für Blockpraktika, dass durch die starke zeitliche Verdichtung der Lehrveranstaltung ein gutes *Zeitmanagement* gefragt ist und dadurch ggf. auch erlernt werden kann:

- f) Die Fähigkeit zum *Zeitmanagement* entwickelt sich in besser in Blockpraktika als in Semesterpraktika.

Hingegen ist für die Konstrukte des PraQ-B, vornehmlich *Lehrkompetenz*konstrukte und *Skriptqualität*, zu vermuten, dass diese keine systematischen Unterschiede zwischen Semester- und Blockpraktika aufweisen, da die Interaktionszeiten der Studierenden mit dem/der Betreuenden und mit dem Skript etwa gleich ausfallen dürften:

- g) Weder für die Konstrukte der *Lehrkompetenz*, noch für die Konstrukte der *Lernumgebungsqualität* gibt es Unterschiede zwischen Semester- und Blockpraktika.

13.3.2. Methode

Stichprobe Erneut wurde auf Erhebungsdaten der Studie 2 zurückgegriffen. Für diese Studie wurden solche Proband*innen ausgewählt, die entweder ein Semester- oder ein Blockpraktikum besuchten. Da nur für Physikpraktika eine genügend große Stichprobe für die MIMIC-Analyse vorlag, wird sich hier aus Gründen der Analysegenauigkeit auf die Stichprobe der Physikstudierenden (Mono und Lehramt) in Physikpraktika beschränkt. Facheinflüsse der Praktika wurden bereits in Studie 6 erforscht.

Für den PraQ-A bestand die Stichprobe aus $N = 290$ Physikstudierenden in Physikpraktika, davon besuchten $N = 77$ Studierende ein Blockpraktikum, $N = 207$ ein reguläres Semesterpraktikum. Genauer beschreibt Tab. 13.8 die Stichprobe.

Für den PraQ-B ergaben sich $N = 185$ Physikstudierende in Physikpraktika. Tab. 13.9 verdeutlicht weitere Eigenschaften der Praktika und Studierenden.

Instrumente Erneut wurde je Praktikum entweder der PraQ-A oder PraQ-B eingesetzt (vgl. Studie 2).

Auswertungsmethode Um die Forschungsfrage zu beantworten wird analog zum Vorgehen in Studie 6 die Messinvarianz der PraQ-Messungen über MIMIC-Modelle geprüft, diesmal ist die dichotome Einflussvariable der Praktikumstyp. Dies ist zulässig, da die Stichprobengröße je Analyse (PraQ-A bzw. PraQ-B) $N = 150$ überschreitet (BROWN, 2006, S. 305).

Tabelle 13.8.: Gegenüberstellung von Semester- und Blockpraktika für die Stichprobe des PraQ-A. $N = 207$ Studierende besuchten ein Semesterpraktikum, $N = 77$ besuchten ein Blockpraktikum. Es handelt sich um Physikstudierende (Mono und Lehramt) in Physikpraktika.

Praktikumstyp	Tandem	baugleich	weiblich	Semester
Semester	99%	84%	25%	3.3 (1.6)
Block	100%	99%	30%	2.8 (1.1)

Tandem: Wurde in Tandems gearbeitet?

baugleich: Waren die Experimente baugleich mehrfach angeboten?

weiblich: Anteil weiblicher Studierender

Semester: Fachsemester: $M(SD)$

Tabelle 13.9.: Gegenüberstellung von Semester- und Blockpraktika für die Stichprobe des PraQ-B. $N = 156$ Studierende besuchten ein Semesterpraktikum, $N = 29$ ein Blockpraktikum. Es handelt sich um Physikstudierende (Mono und Lehramt) in Physikpraktika.

Praktikumstyp	Tandem	baugleich	Skript	Begleit-VL	weiblich	Semester
Semester	90%	98%	81%	29%	25%	2.6 (1.0)
Block	96%	92%	97%	21%	36%	3.7 (1.1)

Tandem: Wurde in Tandems gearbeitet?

baugleich: Waren die Experimente baugleich mehrfach angeboten?

Skript: Gab es ein Skript?

Begleit-VL: Existierte eine Begleitvorlesung?

weiblich: Anteil weiblicher Studierender

Semester: Fachsemester: M (SD)

Erneut sind die Einzelgruppen zu klein für eine echte multi-group-Analyse (nur $N = 29$ für PraQ-B-Proband*innen in Blockpraktika).

Für die DIF-Analyse (vgl. Studie 6) wird bei plausiblen inhaltlichen Argumenten der erneut Einfluss der Gruppenvariablen auf das betreffende Item zugelassen, es handelt sich dann bei der betreffenden Skala nur noch um partielle Messinvarianz. Erweitert werden kann die Analyse erneut um Kovariaten, für deren Einfluss dann in der Bestimmung der Gruppenunterschiede kontrolliert werden kann. Untersucht wurden Einflüsse des Fachsemesters, des *theoretischen* und *praktischen Vorwissens*, Geschlechts, Fachsemesters und des Alters.

Durchführung Details zur Erhebung sind den Studien 2.1 und 2.2 zu entnehmen.

13.3.3. Ergebnisse

PraQ-A: Für das reine MIMIC-Modell mit der dichotomen Einflussvariablen Semesterpraktikum/Blockpraktikum ergibt sich nach Zulassen der inhaltlich begründbaren Item-DIFs (s.u.) ein gerade eben akzeptabler Modellfit ($CFI = .91$, $TLI = .90$, $RMSEA = .05$, $SRMR = .06$) ohne inhaltlich begründbare standardisierte $EPCs > .25$.

Alle signifikanten Gruppenunterschiede der Faktoren des PraQ-A fallen zu Gunsten der Blockpraktika aus. Für den Faktor *EXP: Versuch aufbauen* ergibt sich eine mittlere Effektstärke zu Gunsten von Blockpraktika ($d = 0.55$), für die Faktoren *EXP: Erwartungen formulieren* ($d = 0.32$), *Zeitmanagement* ($d = 0.40$) und *Kooperationskompetenz* ($d = 0.34$) und ergeben sich kleine Effekte zu Gunsten von Blockpraktika (Abb. 13.7).

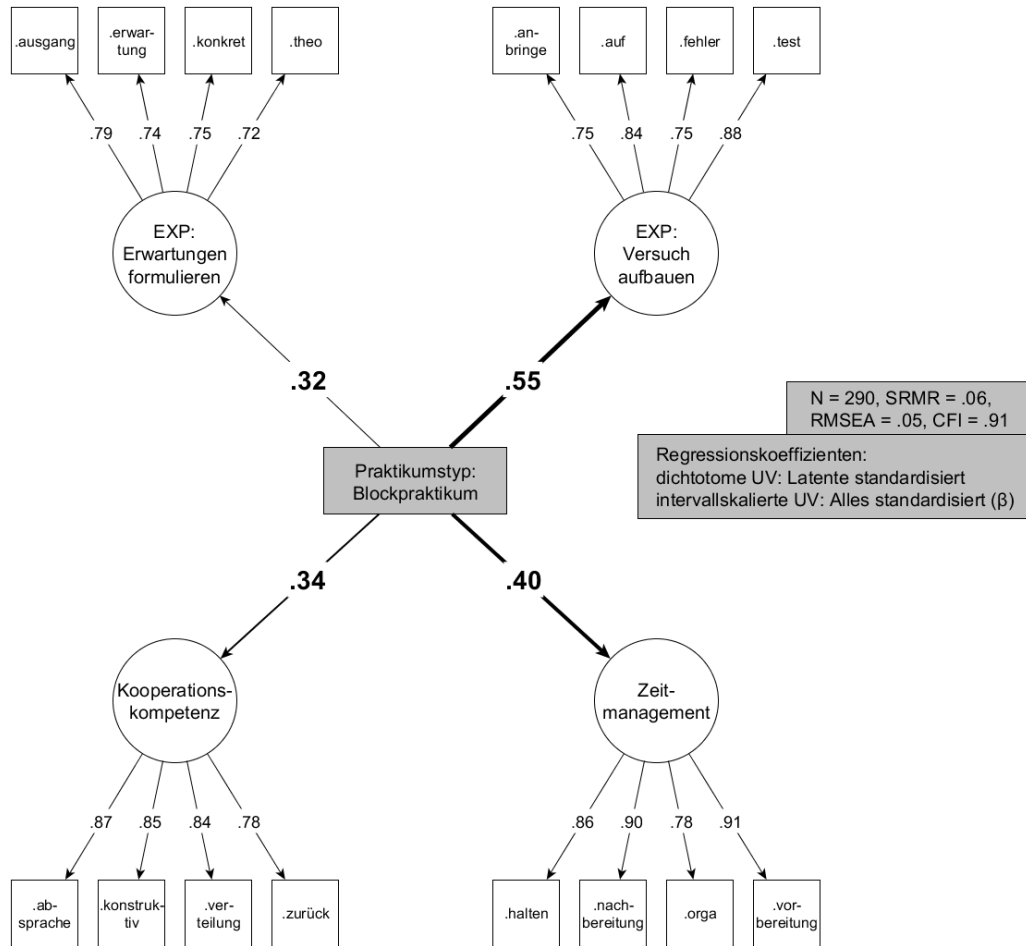


Abbildung 13.7.: Signifikante Ergebnisse der MIMIC-Analyse zu Gruppenunterschieden Semester-/Blockpraktika zum PraQ-A. Regressionskoeffizienten > 0 indizieren eine höhere Ausprägung des betroffenen Faktors für Blockpraktika. Die für die Latente standardisierten Regressionskoeffizienten können als Effektstärke COHENS d interpretiert werden (BROWN, 2006, S. 314ff). Ladungen werden standardisiert dargestellt. Korrelationen zwischen Faktoren werden zwecks Übersichtlichkeit nicht dargestellt.

In der Analyse wurden insgesamt drei inhaltlich plausible Item-DIFs zugelassen bzw. ein direkter Effekt der MIMIC-Variablen auf diese Itemausprägungen geschätzt. Dies fasst Tab. 13.10 zusammen.

Es ist zu entnehmen, dass das Item »... weiß ich besser, was in welches Kapitel des Protokolls/Versuchsberichtes gehört.« des Faktors *Kommunikationskompetenz schriftlich* einen substanziellen DIF von 0.84 Skalenpunkten ($SE = 0.2$) zu Gunsten von Blockpraktika aufweist.

Ebenso systematisch größer in Blockpraktika fällt die Ausprägung des Items »... kann ich die einzelnen Geräte besser montieren.« des Faktors *Experimentelle Kompetenz: Versuchsplanung & Geräte zusammenstellen* aus ($M = 0.39$, $SE = 0.14$).

In Semesterpraktika ergibt sich ein systematisch höherer Wert für das Item »... kann ich die Eignung von Geräten für das Experimentieren besser einschätzen.« des Faktors *Experimentelle Kompetenz: Gerätekenntnisse deklarativ* mit 0.48 Skalenpunkten ($SE = 0.13$). Die inhaltliche Begründung der Freisetzung erfolgt in der Diskussion.

MIMIC-Analyse mit Kovariaten Bezieht man nun die in der vorherigen Studie beschriebenen Kovariaten mit in die Analyse ein, so ergibt sich erneut nur für das *theoretische Vorwissen* ein signifikanter Modellbeitrag. Dies fasst Abb. 13.8 zusammen.

Die MIMIC-Modellierung weist hier erneut einen akzeptablen Modellfit auf ($CFI = .91$, $TLI = .89$, $RMSEA = .05$, $SRMR = .05$). Der Einschluss der Variablen *Vorwissen (Theorie)* liefert hier die Erkenntnis, dass diese Variable den Einfluss des Praktikumstyps auf *EXP: Erwartungen formulieren* vollständig mediert¹⁹

Sämtliche übrige im MIMIC-Modell beschriebenen Einflüsse des Praktikumstyps werden zudem durch Gruppenunterschiede des *theoretischen Vorwissens* partiell schwach mediert.²⁰ Neu hinzu kommt ein kleiner direkter Gruppeneffekt auf den Faktor *EXP: Probleme lösen*.²¹

PraQ-B: In der MIMIC-Analyse für die Faktoren des PraQ-B ergibt sich nach Einschluss unten aufgeführter DIFs ein akzeptabler Modellfit ($CFI = .92$, $TLI = .90$, $RMSEA = .07$, $SRMR = .05$) ohne inhaltlich begründbare standardisierte *EPCs* $> .24$. Abb. 13.9 stellt dies wie gewohnt graphisch dar.

¹⁹ Erneut ergibt der unstandardisierte indirekte Effekt $0.41 \cdot 0.28 = 0.11$ unter Addition des nichtsignifikanten direkten Effekts von 0.17 den ursprünglichen direkten Effekt aus dem reinen MIMIC-Modell von 0.28.

²⁰ Die Effektstärken reichen hier von 0.17 bis 0.40.

²¹ Auch hier könnte es sich um ein statistisches Artefakt handeln.

Tabelle 13.10.: Ergebnisse zur DIF-Analyse der Items des PraQ-A bezüglich der Gruppen Semester-/Blockpraktika. Von insgesamt 70 untersuchten Items in 17 Faktoren traten drei substanzielle DIFs auf. Die hier dargestellten, unstandardisierten DIF-Werte können in der Metrik der Faktoren, also Punkten auf einer Likertskala von 1 bis 6, interpretiert werden (BROWN, 2006, S. 314ff). Positive Vorzeichen entsprechen einer systematisch höheren Itemausprägung in Blockpraktika.

MIMIC-Analyse PraQ-A: DIF	
Konstrukt	Itemwortlaut
DIF	
SE	
Kommunikation schriftlich	... weiß ich besser, was in welches Kapitel des Protokolls/ Versuchsberichtes gehört.
Experimentelle Kompetenz: Gerätekenntnisse deklarativ	... kann ich die Eignung von Geräten für das Experimentieren besser einschätzen.
Experimentelle Kompetenz: Versuchsplanung & Geräte zusammenstellen	... kann ich die einzelnen Geräte besser montieren.

*** $p < .001$, ** $p < .01$

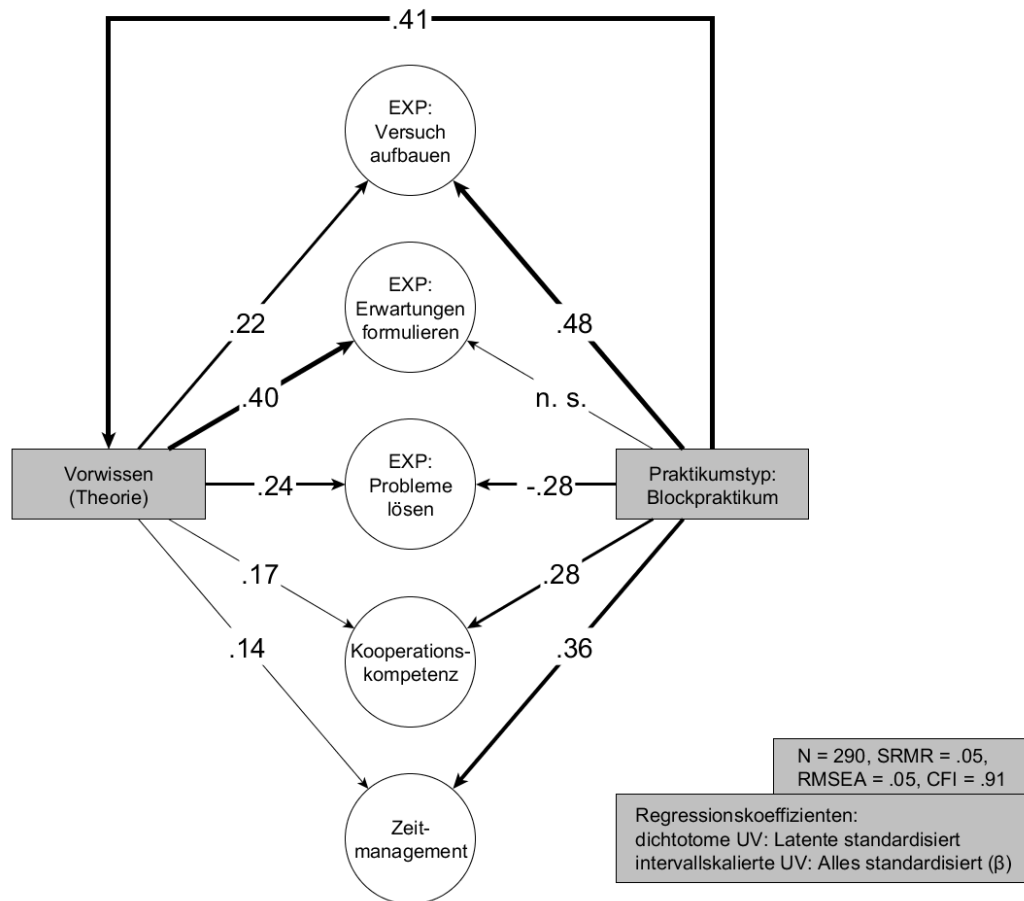


Abbildung 13.8.: MIMIC-Modell zu Gruppeneffekten Semester-/Blockpraktika PraQ-A mit Kovariate *Vorwissen (Theorie)*. Regressionskoeffizienten > 0 indizieren eine höhere Ausprägung des betroffenen Faktors für Blockpraktika. Die von dichotomen Variablen ausgehenden Regressionspfade sind berechnet für die standardisierte latente Variable und daher als Effektstärke d zu interpretieren. Die von der Variablen *Vorwissen (Theorie)* ausgehenden Regressionspfade sind auf beiden Seiten standardisiert und daher wie z -Werte in Standardabweichungen zu interpretieren. Korrelationen zwischen Faktoren werden zwecks Übersichtlichkeit nicht dargestellt.

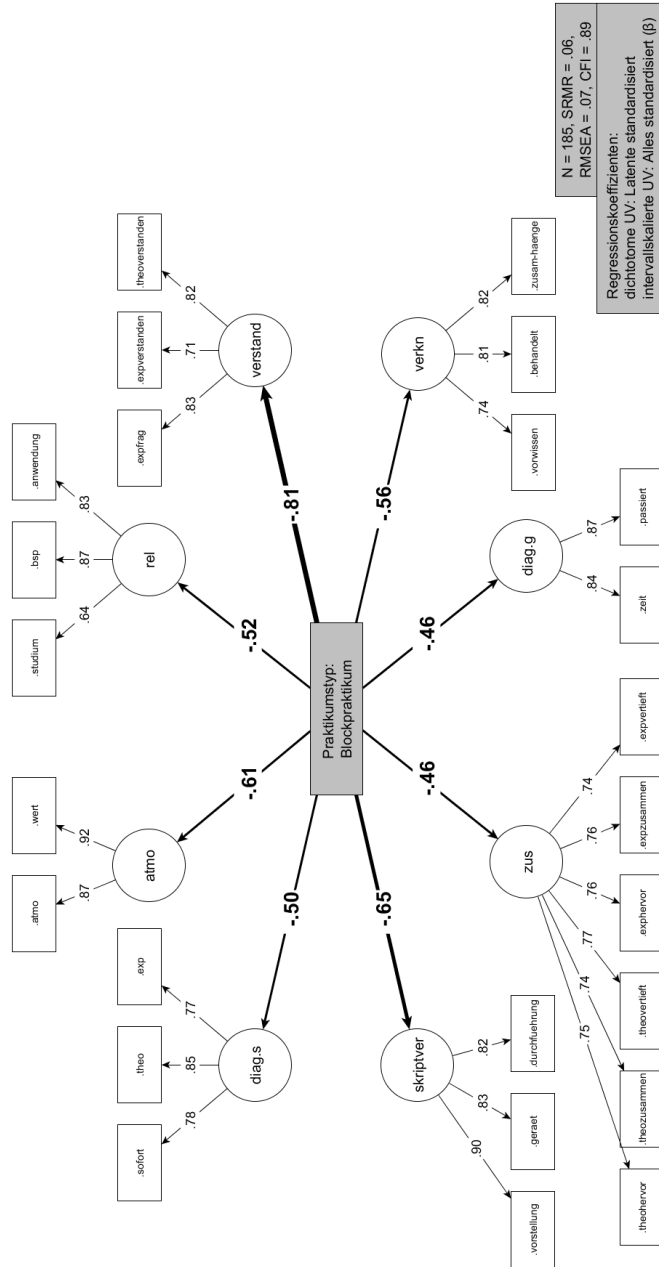


Abbildung 13.9.: Signifikante Ergebnisse der MIMIC-Analyse zu Gruppenunterschieden Semester-/Blockpraktika zum PraQ-B. Regressionskoeffizienten > 0 indizieren eine höhere Ausprägung des betroffenen Faktors für Blockpraktika. Die für die Latente standardisierten Regressionskoeffizienten können als Effektstärke COHENs d interpretiert werden (BROWN, 2006, S. 314ff). Ladungen werden standardisiert dargestellt. Korrelationen zwischen Faktoren werden zwecks Übersichtlichkeit nicht dargestellt.

Ersichtlich sind mittlere bis große direkte Effekte des Praktikumstyps auf insgesamt acht Faktoren. Es werden hierfür stets systematische Gruppenunterschiede zu Gunsten von Semesterpraktika deutlich. Für die *Lehrkompetenz*-Faktoren *verständlich erklären* ($d = -0.81$), *Verknüpfungen herstellen* ($d = -0.56$), *Diagnostik: Grundeinstellung* ($d = -.46$), *Zusammenfassen und Wiederholen* ($d = -0.46$), *Diagnostik: situativ* ($d = -0.50$), *Gute Lernatmosphäre herstellen* ($d = -0.61$), und *Relevanz verdeutlichen* ($d = -0.52$) ergeben sich je signifikante Effekte.

Aber auch für den Faktor *Skriptqualität für den Versuch* aus den Konstrukten der *Lernumgebungsqualität* ergab sich ein signifikanter Effekt des Praktikumstyps mit mittlerer bis hoher Effektstärke ($d = -0.65$). In obigem Modell sind die Item-DIFs bereits berücksichtigt worden, es gab nur einen mittelstarken Item-DIF ($DIF = 0.43^*$, $SE = 0.20$) für das Item »...nennt Beispiele für mögliche Anwendungen der Experimente« des Faktors *Relevanz verdeutlichen*, also einer höheren Ausprägung in Blockpraktika.

Nur für ein Item ergab sich ein signifikanter DIF. Es handelt sich um das Item »Der/die Betreuende nennt Beispiele für mögliche Anwendungen der Experimente.« zum Faktor *Relevanz verdeutlichen*, dass systematisch in Blockpraktika um 0.43 Skalenpunkte ($SE = 0.20$) höher ausfällt als in Semesterpraktika. Für den Faktor *Relevanz verdeutlichen* liegt daher lediglich eine partielle Messinvarianz vor.

MIMIC-Analyse mit Kovariaten Eine Analyse der PraQ-B-Faktoren unter Einbezug der Kovariaten zeigt Abb. 13.10. Es ergibt sich für die Kovariate *Vorwissen (Theorie)* ein signifikanter Modellbeitrag für sechs der acht Faktoren des reinen MIMIC-Modells (s. o.).

Darüber hinaus ergab sich ein signifikanter Effekt des Geschlechts auf den Faktor *Gute Lernatmosphäre herstellen*, der zudem den Effekt des *Vorwissens (Theorie)* auf selbigen Faktor vollständig mediiert (Abb. 13.11).

13.3.4. Diskussion

PraQ-A:

Diskussion des Messmodells über Gruppen Für das Messmodell des PraQ-A ergaben sich ein akzeptabler Fit und nur 3 signifikante DIF-Items, was für eine gelungene statistische Modellierung spricht. Das erste DIF-Item »...weiß ich besser, was in welches Kapitel des Protokolls/Versuchsberichtes gehört.« des Faktors *Kommunikationskompetenz schriftlich* weist hierbei mit 0.84 Skalenpunkten einen starken DIF zu Gunsten von Blockpraktika auf, der sich aber inhaltlich rechtfertigen lässt. So können gerade Zuordnungsfragen zu Kapiteln im Versuchsprotokoll

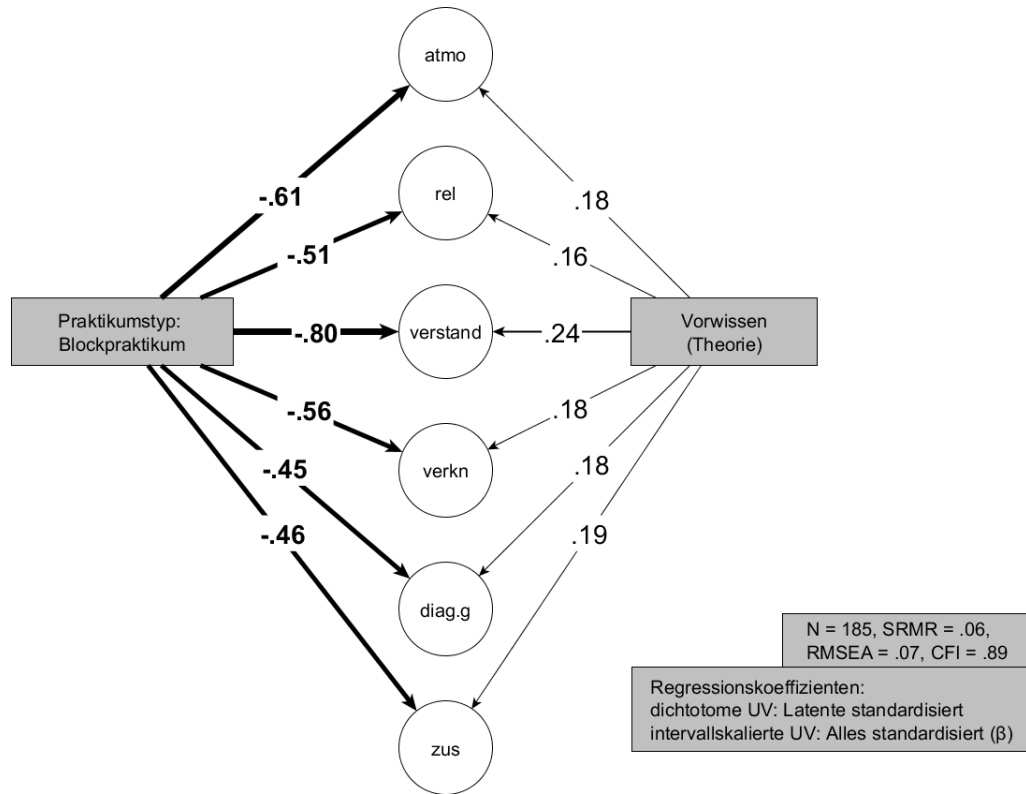


Abbildung 13.10.: MIMIC-Modell zu Gruppeneffekten Semester-/Blockpraktika PraQ-B mit Kovariate Vorwissen (Theorie). Regressionskoeffizienten > 0 indizieren eine höhere Ausprägung des betroffenen Faktors für Blockpraktika. Die von dichotomen Variablen ausgehenden Regressionspfade sind berechnet für die standardisierte latente Variable und daher als Effektstärke d zu interpretieren. Die von der Variablen Vorwissen (Theorie) ausgehenden Regressionspfade sind auf beiden Seiten standardisiert und daher wie z -Werte in Standardabweichungen zu interpretieren. Korrelationen zwischen Faktoren werden zwecks Übersichtlichkeit nicht dargestellt.

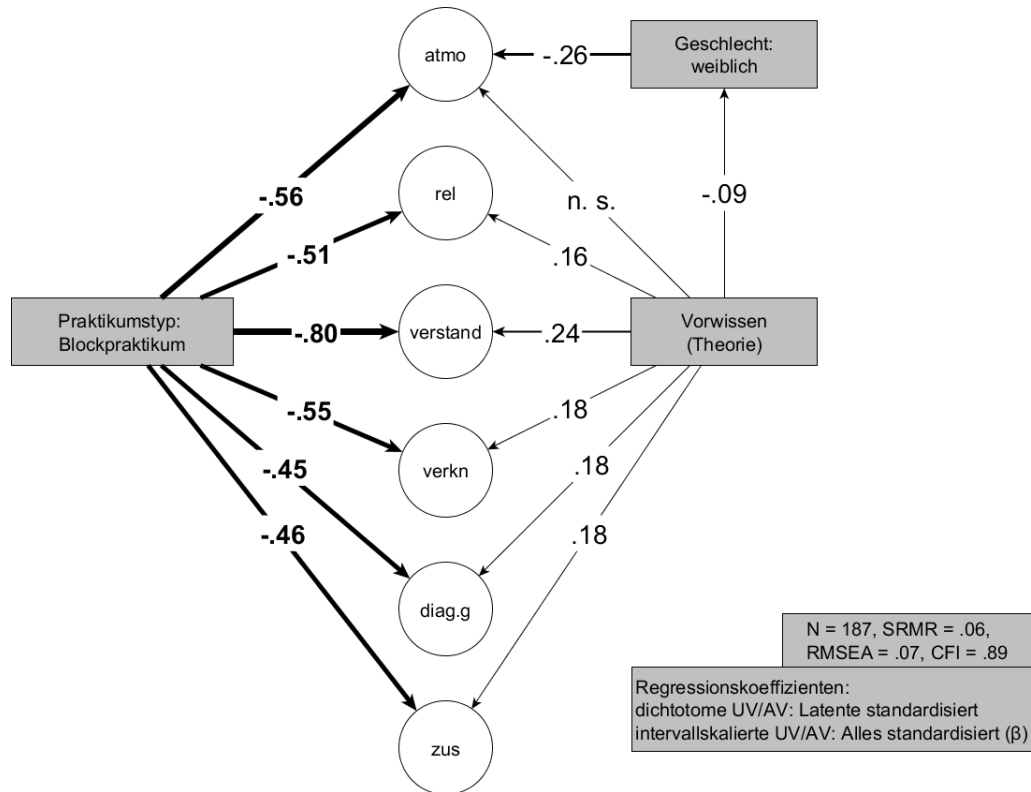


Abbildung 13.11.: MIMIC-Modell zu Gruppeneffekten Semester-/Blockpraktika PraQ-B mit Kovariaten *Vorwissen (Theorie)* und *Geschlecht*. Regressionskoeffizienten > 0 indizieren eine höhere Ausprägung des betroffenen Faktors für Blockpraktika. Die von dichotomen Variablen ausgehenden Regressionspfade sind berechnet für die standardisierte latente Variable und daher als Effektstärke d zu interpretieren. Die von der Variablen *Vorwissen (Theorie)* ausgehenden Regressionspfade sind auf beiden Seiten standardisiert und daher wie z -Werte in Standardabweichungen zu interpretieren. Korrelationen zwischen Faktoren werden zwecks Übersichtlichkeit nicht dargestellt.

während der zeitüberdauernden Anwesenheit im Blockpraktikum leichter von den Studierenden gegenüber den Betreuenden geäußert werden, ohne Fernanfragen per Email oder Ähnliches unternehmen zu müssen.

Ebenso zu Gunsten von Blockpraktika höher ausgeprägt fiel das Item » ... kann ich die einzelnen Geräte besser montieren.«, zum Faktor *Experimentelle Kompetenz: Versuchsplanung & Geräte zusammenstellen* aus, mit ca. einem halben Skalenpunkt. Erneut ist dies dadurch zu erklären, dass Betreuende in Blockpraktika mehr Zeit am Stück mit den Studierenden verbringen, wodurch die Studierenden sich eher trauen können, auch kleinlich erscheinende Fragen offen zu stellen, was wiederum einen höheren Lernzuwachs bewirken kann. Das Item » ... kann ich die Eignung von Geräten für das Experimentieren besser einschätzen.« zum Faktor *Experimentelle Kompetenz: Gerätekenntnisse deklarativ* fällt dagegen in Semesterpraktika leicht höher aus, was dadurch begründet werden könnte, dass dieses »Einschätzungswissen« mit den Geräteangaben zusammenhängt, die wiederum meist in Einzelarbeit beim Lesen von Anleitungen gewonnen werden.

Da Semesterpraktika mehr zeitlichen Freiraum gewähren, könnte sich dies systematisch positiv auf dieses Item auswirken. In jedem Fall ist festzuhalten, dass die Faktoren *Kommunikationskompetenz schriftlich*, *Experimentelle Kompetenz: Versuchsplanung & Geräte zusammenstellen* und *Experimentelle Kompetenz: Gerätekenntnisse deklarativ* nur eine partielle Messinvarianz aufweisen und daher ihre Ausprägungen mit Vorsicht in Bezug auf Gruppenunterschiede zwischen Block- und Semesterpraktika zu interpretieren sind.

Diskussion von Gruppenunterschieden Folgende Hypothesen bezüglich der Konstrukte des PraQ-A haben sich im Rahmen des Ergebnisteils bestätigt:

- f) Die Fähigkeit zum *Zeitmanagement* entwickelt sich in Semesterpraktika schlechter als in Blockpraktika.

Alle verbleibenden Hypothesen sind dagegen abzulehnen. Es hat sich erwartungskonform gezeigt, dass Blockpraktika die positive Entwicklung der eigenen Wahrnehmung des *Zeitmanagements* besser fördern als Semesterpraktika. Allerdings konnten für die Hypothesen a) bis d) keinerlei Gruppenunterschiede gefunden werden. Weder fördern Semesterpraktika die Fähigkeit zum *Rohdaten aufbereiten*, noch zum *Daten interpretieren* stärker als Blockpraktika. Ebenso ergaben sich keine Unterschiede in der Förderung des theoretischen Wissenserwerbs oder der *schriftlichen Kommunikation*.

Subsumiert überwiegen also in der Dimension *Lernzuwachs* die Argumente pro Blockpraktika, sie fördern gemäß Analyse manche Kompetenzbereiche der *Experi-*

mentellen Kompetenz besser (drei²² von neun Skalen) und helfen besser beim Erwerb überfachlicher Kompetenzen wie *Zeitmanagement* und *Kooperationskompetenz*. Aus der Perspektive des *Lernzuwachs* im 3L-Modell sind also Blockpraktika zu bevorzugen, da sie den Kompetenzerwerb der Studierenden anscheinend besser fördern können als Semesterpraktika.

PraQ-B:

Diskussion des Messmodells über Gruppen Das Messmodell für die MIMIC-Analyse des PraQ-B war erneut akzeptabel, bei nur einem einzigen signifikanten DIF. Es handelte sich hier um das Item » ... nennt Beispiele für mögliche Anwendungen der Experimente.« des Faktors *Relevanz verdeutlichen*, das mit etwa einem halben Skalenpunkt systematisch zu Gunsten von Blockpraktika ausfällt.

Diese höhere Ausprägung in Blockpraktika kann dadurch begründet werden, dass in Blockpraktika zeitnäher Verknüpfungen zwischen den Praktikumsversuchen durch die Betreuenden vorgenommen werden können, da an einem Tag mehrere Praktikumsversuche absolviert werden und den Studierenden zeitgleich präsent sind. Dies könnte ihre systematisch bessere Einschätzung der Betreuenden in diesem Item erklären.

Diskussion von Gruppenunterschieden Die Hypothese g) ist auf Basis obiger Analysen abzulehnen, es gibt deutliche Gruppenunterschiede in der Wahrnehmung der *Lehrkompetenz* des Betreuenden zwischen Semester- und Blockpraktikastudierenden. Zudem fällt auch die Skala *Skriptqualität für den Versuch* gruppenverschieden aus. Es ergaben sich im Unterschied zum PraQ-A je Gruppenunterschiede zu Gunsten von Semesterpraktika mit mittlerem bis starkem Effekt.

In Semesterpraktika werden die Fähigkeiten der Betreuenden, *verständlich zu erklären*, *Verknüpfungen herzustellen*, *eine positive Grundeinstellung zu Diagnostik zu besitzen*, *Zusammenzufassen*, *situativ zu diagnostizieren*, *eine gute Lernatmosphäre herzustellen* und die *Relevanz der Lerninhalte herauszustellen* deutlich positiver eingeschätzt als in Blockpraktika. Unter Kontrolle des *theoretischen Vorwissens* entfallen allerdings die Unterschiede im *situativ Diagnostizieren* und unter Kontrolle von Geschlechtseinflüssen entfällt ebenso der Unterschied in der wahrgenommenen Fähigkeit, *eine gute Lernatmosphäre herzustellen*.

Es verbleiben trotzdem sechs von insgesamt zehn Konstrukten der *Lehrkompetenz*, die deutlich höher in Semesterpraktika ausfallen.

²²Die Unterschiede im Faktor *EXP: Probleme lösen* sind allerdings möglicherweise ein statistisches Artefakt.

Die Betreuung gelingt also, in der Wahrnehmung der Studierenden, in Semesterpraktika besser, weshalb aus Perspektive der *Lehrkompetenz* Semesterpraktika stark zu bevorzugen sind. Aus Perspektive der *Lernumgebungswahrnehmung* könnte ebenso, mit Bezug zur besseren Wahrnehmung der *Skriptqualität für den Versuch*, geschlossen werden, dass Semesterpraktika einen effektiveren Einsatz des Skriptes begünstigen. Kontrolliert man hier erneut für das *theoretische Vorwissen*, ist dieser Unterschied allerdings nichtig und somit irrelevant.

Insgesamt zeigt sich also ein zweigeteiltes Bild zum Für und Wider von Blockpraktika versus Semesterpraktika. Erstere besitzen analysegemäß mehr Stärken in der Förderung studentischer Kompetenzen, letztere scheinen eine deutlich bessere Betreuung zu begünstigen. Eine klare Empfehlung lässt sich daher nicht ableiten, am ehesten seien hier auf Grund der stark besseren Betreuung Semesterpraktika empfohlen.

Methodenkritik Auch diese Studie betrat forschungstechnisches Neuland durch die erste systematische Erforschung der Unterschiede zwischen zwei beliebten Praktikums-Organisationsformen in Deutschland. Einschränkend muss an dieser Stelle jedoch hervorgehoben werden, dass die Gesamtstichproben für die Schätzung der vielen Parameter der MIMIC-Modelle als vergleichsweise klein zu bezeichnen sind und daher eine gewisse Ungenauigkeit der Schätzung innewohnt.

An den teils großen Standardfehlern im PraQ-B-Modell ist dies gut abzulesen. Weiter ist natürlich der Evaluationscharakter der Erhebung zu betonen. Gemessen wurden Selbsteinschätzungen der Studierenden, mit den entsprechenden möglichen Verzerrungseffekten (vgl. Abschnitt 10.1 auf Seite 107).

Zuletzt lässt sich das MIMIC-Verfahren erneut selbst kritisieren, da es Faktorladungen (Item-Konstrukt-Beziehung), Fehler(ko)varianzen (Item-Item-Beziehung) und Faktor(ko)varianzen (Faktor-Faktor-Beziehung) über Gruppen hinweg als gleich annimmt (BROWN, 2006, S. 305ff). Eine »echte« Messinvarianz über die Gruppen der Studierenden von Semester- und Blockpraktika gilt es also noch nachzuweisen.

Kapitel 14

Zusammenfassung und Ausblick

14.1. Evaluation von Praktika

Der PraQ hat sich in Erhebungen an 18 Praktika-Standorten im deutschsprachigen Raum bewähren können. Die Erhebungen liefen in 15min ab und ermöglichen einen groben, aber umfassenden Überblick über qualitätsrelevante Praktikumsmerkmale in den Dimensionen *Lernzuwachs*, *Lehrkompetenz* und *Lernumgebungsqualität*.

Die Anforderung der Ökonomie (vgl. Abschnitt 4.4 auf Seite 43) kann hiermit als erfüllt angesehen werden. Grenzen der Erfassung von Konstrukten liegen bei Tiefenanalysen von Praktika (wie bei MÜHLENBRUCH et al., 2015; REHFELDT et al., 2015). Dazu zählen z. B. die konkrete Organisation der Praktika, die spezifische Fachkompetenz der Betreuenden und die Aufbereitung und Wahl der Experimente.

Für Praktikumsevaluierende stellt sich zunächst die Frage nach der Wahl des Erhebungsinstruments. Sollen die selbsteingeschätzten Kompetenzzuwächse der Studierenden und somit ihr *Lernzuwachs* betrachtet werden, so bietet sich der PraQ-A mit Skalen zum *Fachwissen*, zur *Experimentellen Kompetenz*, *Kommunikationskompetenz*, *Bewertungskompetenz*, *Kooperationskompetenz* und *Personalkompetenz* an.

Soll Aufschluss gewonnen werden über die Qualität der Lehre, der *Lehrkompetenz* der Betreuenden und der Qualität der Praktikums-*Lernumgebung*, so ist der PraQ-B einzusetzen. Dieser enthält Skalen zu *didaktischen und pädagogischen Fertigkeiten* der Lehrenden, zur *Skriptqualität* und zur *Integration* des Praktikums in das Studium.

Mit Blick auf die Testevaluationsstudien lassen sich folgende Ergebnisse für die Messqualität des PraQ festhalten:

1. Relevante inhaltliche Aspekte von Praktikumsqualität werden abgedeckt.^a
2. Die Konstrukte der Praktikumsqualität sind voneinander abgegrenzt interpretierbar.^b
3. Es besteht ein geringer Einfluss affektiver Praktikumszufriedenheit auf die PraQ-Messung (Studie 3.2).
4. Das Niveau der Skriptbeurteilung durch Studierende ist annähernd gleichwertig zu Expert*innenbeurteilungen.

^aAusnahmen: Konstrukt *Experimentelle Kompetenz: Eigenständiges Planen von Experimenten*, Studie 1.1

^bAusnahmen: Konstrukte *Experimentelle Kompetenz: Fragestellung klären*, *Lehrkompetenz: Lerndienliches Feedback geben* (je Studie 2.1), *Lehrkompetenz: Zur Selbsttätigkeit anregen* (Studie 2.2)

14.2. Erforschung von Lernzuwachs, Lehrkompetenz und Lernumgebungsqualität

Der PraQ beinhaltet Skalen, die mannigfaltige Testevaluationsstudien erfolgreich durchlaufen haben, weshalb der Blick auf eine Verwertbarkeit der Skalen für die empirische Bildungsforschung lohnt. Folgende Skalen sieht der Autor als unabhängig vom Praktikumskontext in leicht adaptierter Form für die empirische Bildungsforschung einsetzbar an:

- Fachwissen Theorie
- Fachwissen Praxis für naturwissenschaftliche, experimentelle Kontexte
- Neun Experimentelle-Kompetenz-Skalen für naturwissenschaftliche, experimentelle Kontexte
- Kommunikationskompetenz schriftlich für naturwissenschaftliche, experimentelle Kontexte
- Kommunikationskompetenz mündlich
- Bewertungskompetenz fachlich für naturwissenschaftliche, experimentelle Kontexte
- Kooperationskompetenz
- Interesse für Lehrveranstaltungen im Studium

- verständlich erklären
- Zusammenfassen & Wiederholen für naturwissenschaftliche, experimentelle Kontexte
- Verknüpfungen herstellen
- Verständnis überprüfen für naturwissenschaftliche, experimentelle Kontexte
- Gute Lernatmosphäre herstellen
- Relevanz verdeutlichen für naturwissenschaftliche, experimentelle Kontexte
- Ermutigen und Selbstwirksamkeit stärken für naturwissenschaftliche, experimentelle Kontexte
- Effizient mit Störungen umgehen
- Diagnostik Grundeinstellung
- Diagnostik situativ für naturwissenschaftliche, experimentelle Kontexte

Die Gütekriterien dieser Skalen liegen fast alle im guten bis sehr guten Bereich (vgl. Abschnitt 12.7 auf Seite 187), die Interpretierbarkeit im Rahmen des argument-based approach lässt schlussfolgern, dass die Skalen des PraQ einen gewinnbringenden Beitrag zur Erhebungsinstrumentenlandschaft der empirischen Bildungsforschung, speziell in den Naturwissenschaftsdidaktiken, leisten können.

14.3. Erkenntnisse zu Praktika im deutschsprachigen Raum

Die PraQ-Instrumente haben darüber hinaus gezeigt, dass mit ihnen Forschungsfragen über Praktika zu untersuchen sind.

In Studie 5 wurde das erste Mal in der Geschichte der Praktikumsforschung ein großes Benchmarking der Praktikumsqualität über 18 Praktika im deutschsprachigen Raum vorgenommen. Praktika fördern demnach vor allem das *Fachwissen* der Studierenden, bei gleichzeitig vergleichsweise geringer *intrinsischer Motivation* derselben. Die Betreuung in Praktika fällt indes akzeptabel aus, in diversen pädagogischen und didaktischen Konstrukten der *Lehrkompetenz* erreichen die Betreuenden mittlere Ausprägungen. Die Qualität von *Skript* und *Integration* fällt mittelmäßig aus. In einem exemplarischen Vergleich zweier stark unterschiedlicher Praktika konnte zudem gezeigt werden, dass Qualitätsprofile von Praktika zu ermitteln und auszuwerten sind.

14. Zusammenfassung und Ausblick

Chemie und Physikpraktika weisen bei der Analyse ihrer Unterschiede folgende Eigenschaften auf: Physikpraktika fördern eher das *Dokumentieren von Messungen*, das *Aufbereiten von Rohdaten*, *experimentelles Problemlösen* und *schriftliches Kommunizieren* bei den Studierenden. Die Betreuer*innen in Physikpraktika *fassen* besser *Inhalte zusammen* und *Überprüfen* regelmäßiger das *Verständnis* ihrer Studierenden. Physikpraktika scheinen lern- und betreuungsförderlicher als Chemiepraktika in einigen Bereichen der Praktikumsqualität zu sein, welche es genauer zu analysieren gilt, um die Stärken der Physikpraktika auch in Chemiepraktika erreichen zu können.

Bezüglich der Frage nach der Organisationsform des Praktikums als Block- oder Semesterpraktikum ergab sich, dass in der Dimension *Lernzuwachs* die Argumente pro Blockpraktika überwiegen. Sie fördern gemäß Analyse das *Versuche aufbauen* und das *Zeitmanagement* bedeutsam besser als Semesterpraktika.

Betrachtet man auf der anderen Seite die Dimensionen *Lehrkompetenz* der Betreuenden und *Lernumgebung* für Block- und Semesterpraktika, so ergibt sich ein gegensätzliches Bild. In Semesterpraktika werden die Fähigkeiten der Betreuenden, *verständlich zu erklären*, die *Relevanz der Lerninhalte* zu *verdeutlichen*, *Verknüpfungen herzustellen*, eine *gute Lernatmosphäre herzustellen*, zu *Diagnostizieren* und *Zusammenzufassen* positiver eingeschätzt als in Blockpraktika. Aus Perspektive der *Lernumgebungswahrnehmung* wird zudem die *Skriptqualität für den Versuch* dort systematisch höher eingeschätzt. Bei Kontrolle lernrelevanter Einflüsse wie Vorwissen und Geschlecht entfallen lediglich die Unterschiede im Herstellen der Lernatmosphäre und im situativen Diagnostizieren. Sechs von zehn Konstrukten der *Lehrkompetenz* bleiben besser ausgeprägt in Semesterpraktika.

Semesterpraktika weisen daher eine deutlich bessere Betreuung bei etwas schlechterer Kompetenzförderung in gewissen Bereichen auf. Sie sind daher, mit Abstrichen, aus Perspektive der Praktikumsqualität eher zu empfehlen als Blockpraktika.

14.4. Ausblick

All diese Ergebnisse können nur der Anfang einer groß angelegten, systematischen empirischen Erforschung der Lehrqualität von Praktika sein. Die Validierung des PraQ geht, wie bei jedem Instrument, weiter. So wären in Anlehnung an Studie 4 auch Expert*inneneinschätzungen zur Dimension *Lernzuwachs* und zu den Konstrukten der *Lehrkompetenz* angebracht, um die Verlässlichkeit der Studierendenbeurteilungen hier besser einschätzen zu können.

Um die Erkenntnisrichtung von Studie 3 zu bedienen, könnten Vergleiche mit weiteren Kompetenztests angestrebt werden. Dazu wäre ein Leistungstest zur *Lehr-*

kompetenz in experimentellen Lernumgebungen und ein objektiver Bewertungsmaßstab für Praktikumsskripte zu entwickeln. Für letzteres liefert die Studie 4 ein erstes Kategoriensystem.

Bezüglich Messinvarianz sollten hierarchische Modelle gerechnet werden, die miteinbeziehen, dass Praktika in Experimentalgruppen eingeteilt sind, welche eine gegenseitige Beeinflussung des Evaluationsergebnisses nahelegen und somit eine adäquatere statistische Modellierung der Praktikumsqualität ermöglichen würden.

Im Laufe der Testevaluationsstudien mussten zudem Skalen aufgegeben werden, die es neuzukonstruieren gilt, da ihr inhaltlicher Wert sich in den Expert*inneninterviews ergab. Dazu gehören die Skalen *Experimentelle Kompetenz: Fragestellung klären*, *Lehrkompetenz: Lerndienliches Feedback geben* und *Lehrkompetenz: Zur Selbsttätigkeit anregen*. Bezüglich der Skalen *Skriptqualität* und *Integration* sollten zudem die Testevaluationsschritte der Studien 2 und 3 an einer größeren Stichprobe wiederholt werden.

Auf der inhaltlichen Ebene könnten weitere Konstrukte Einzug in den PraQ erhalten. So könnte ein »PraQ-C« die fachlichen Betreuer*innenkompetenzen erfassen. Dies ließe sich dann allerdings nur fachspezifisch realisieren.

Auch eine Skala *Offenheit der Lernumgebung* für den PraQ-B ist prinzipiell denkbar, wie sie bereits FRASER & McROBBIE (1995) konzipierten. Damit könnte nachgezeichnet werden, ob sich unsere Praktika tatsächlich mit der Entwicklung eines konstruktivistischen Lernbildes auch konstruktivistisch und offener entwickeln.

Die Skriptqualität gliederte sich für den PraQ-B bisher in *allgemeine* und *versuchsbezogene Skriptqualität* auf. Hinzukommen könnte noch eine Skriptqualität für die Theorieerarbeitung, mit der genauer die Funktion des Skripts zur Vorbereitung auf die fachwissenschaftliche Theorie untersucht werden könnte, etwa im Vergleich zu Vorlesungen.

Da das Praktikum eine kooperative Veranstaltung ist, könnte auch untersucht werden, inwieweit die Betreuenden dies fördern, also das Lernen in Gruppen konstruktiv unterstützen. Auch THIEL et al. (2012, S. 10) bedachten bei der Konstruktion des LeKo diesen Aspekt, verworfen aber auf Grund des wenig kooperativen Charakters von Vorlesungen diesen Aspekt.

Auf Ebene der inhaltlichen Studien zum bereits vorliegenden PraQ-Datensatz von insgesamt über 1300 Studierenden sind weitere Analysen denkbar. So wurde jüngst im Rahmen einer Masterarbeit (BECKER, 2017) untersucht, inwieweit die Physikpraktika den speziellen Bedürfnissen von Lehramtsstudierenden gerecht werden. Langfristig sollten auch Wirkketten von Interventionen in Praktika mit Skalen

des PraQ zu untersuchen sein, um zu erforschen, welche Maßnahmen die Praktikumsqualität steigern und welche nicht.

Eine Verstetigung der Evaluation von Praktika mittels PraQ in Deutschland wird auf der PLT-Tagung 2017 besprochen.

14.4.1. Gewichtung der Konstrukte

Je nach Praktikum und Curriculum können natürlich unterschiedliche Schwerpunkte der Ziele von Praktika (vgl. Abschnitt 5 auf Seite 55) gelegt werden. Das zu spezifizierende Modell soll dabei vor allem aufzeigenden Charakter haben, welche maximalen Qualitätsanforderungen an ein Praktikum gestellt werden können, die einzelne Gewichtung der Dimensionen und Facetten ist indes eine Aufgabe der Evaluierenden für eine später geplante, reguläre Implementation als Evaluationsinstrument.

14.4.2. Handlungsempfehlungen für Praktika

Die meisten früh im Studium verorteten naturwissenschaftlichen Experimentalpraktika sind gemäß Forschungsstand dem »expository-style« zuzuordnen. Diese expository-style-Praktika (vgl. Abschnitt 3.4.4 auf Seite 35) sind nicht nur weit verbreitet, sondern aus lerntheoretischer Sicht besonders interessant. Dies kommt daher, da sie eine stark geschlossene, inputorientiert ausgerichtete Lernumgebung enthalten, wie sie von Bologna-Reform-Befürworter*innen (vgl. Abschnitt 2.1 auf Seite 19) und auch konstruktivistisch orientierten Fachdidaktiker*innen abgelehnt würde und damit sich die Frage offenbart, warum trotzdem noch so viele dieser Praktika in dieser Form in Betrieb sind.

Dieser Abschnitt soll daher Handlungsempfehlungen für expository-style-Praktika bieten. Es wird von mehreren Autor*innen bisheriger Praktikumsstudien eine offenere und für die Studierenden selbstbestimmtere Gestaltung der Grundpraktika vorgeschlagen und gefordert (HUCKE, 1999; HALLER, 1999, S. 70, DIEMER, 1998, S. 7), wo nicht alle Arbeitsschritte vorher in der Versuchsanleitung des Skripts vorgegeben werden, also die Methode geöffnet wird. Die Entwicklung würde daher in Richtung eines »problem-based« Praktikums gehen (DOMIN, 1999). In eine ähnliche Richtung gehen Vorschläge, die für mehr selbstständiges Aufbauen der Versuche plädieren (SERE, 2002).

Liegt in der methodischen Öffnung der Praktika vor allem ein pädagogisches Argument für mehr Kreativität und Eigentätigkeit der Studierenden, so ist fachdidaktisch auch über Veränderungen des im Praktikum vorgesehenen Erkenntnisgewinnungsprozesses nachgedacht worden. So fordern mehrere Autoren die Etablierung einer Hypothesen-Bildungs- und Planungs-Kultur (SCHULZ, 2011, S. 39, 41, TIBERGHIE

et al., 1998, S. 496), also eine Betonung der ersten Phasen des Erkenntnisgewinnungsprozesses, die sonst vom Skript deduktiv gegeben sind.

Auch wird der Einsatz moderner Geräte gefordert, sei es bezogen auf die Praktikumsgeräte selbst (DIEMER, 1998, S. 6, ENGELS & HÜLSBUSCH, 2008, S. 59) oder auf den zeitgemäßen Einsatz von Multimedia und PC im Praktikum (DIEMER, 1998; ZASTROW, 2001; HUCKE, 1999; SANDER, 2000; NAGEL, 2009; FRICKE et al., 2011, MÜHLENBRUCH et al., 2015), um die Vorbereitungs- und Durchführungsphase zu modernisieren und mittels Multimedia ein besseres Lernen zu ermöglichen.

Auch Forderungen zu Betreuer*innenschulungen (SANDER, 2000, S. 254, BOND-ROBINSON & RODRIQUES, 2006, S. 313–314) im Sinne von pädagogischem und fachdidaktischem Basiswissen existieren. Just diese sind auf Basis der hier vorgestellten Forschungsergebnisse durchaus sinnvoll anzusehen. In den untersuchten 18 Praktika im deutschsprachigen Raum fiel die Betreuung im Mittel durchschnittlich bis leicht unterdurchschnittlich aus in den didaktischen und pädagogischen Konstrukten der Qualitätsdimension *Lehrkompetenz*.

Auf Basis der Ergebnisse stark zu fordern ist eine Etablierung höherer *intrinsischer Motivation* der Studierenden. Diese kann auch über eine lernförderliche und wertschätzende Betreuung gefördert werden. Vermutlich wahrscheinlicher ist aber, dass die eher mittelmäßige Integration der Praktika in sonstige Lehrveranstaltungen und die hohe Betonung innerfachlichen theoretischen Wissens Anteil an der geringen *intrinsischen Motivation* der Studierenden in Praktika haben. Von struktureller Seite sollte daher mehr auf vernetztes, anwendungsbezogenes Wissen abgezielt werden, indem Praktika sich inhaltlich mit Vorlesungen vernetzen und potentielle Anwendungsfelder der Themen und Methoden stärker herausstellen.

Die Vernetzung mit Vorlesungen kann beispielsweise so erfolgen, dass sich wöchentlich auf Inhalte geeinigt wird, die dann theoretisch in der Vorlesung und praktisch in den Praktika behandelt werden. Dabei sollte bei der Behandlung der Themen stets mit im Blick sein und auch betont werden, an welchen Stellen des zukünftigen Studiums die Inhalte und Methoden Relevanz besitzen werden – vor allem für fachfremde Studierende. Auch eine Betonung der Relevanz für den später angestrebten Beruf der Studierenden ist möglich.

Die Umgestaltung des Praktikums in den »problem-based«-Typ kann über die Erhöhung der Selbstständigkeit und -verantwortlichkeit die Interesse steigern. Als Organisationsform für Praktika sind hierbei sowohl Block- als auch Semesterpraktika denkbar, mit leichten Qualitätsvorteilen auf Seiten der Semesterpraktika. Ist Raum für Hospitationen oder eine Beauftragung von Forschungsarbeiten in Praktika, so

empfiehlt sich, ein von der PraQ-Qualität her gutes Physikpraktikum genauer zu analysieren und daraus Interventionsansätze abzuleiten.

Zuletzt soll noch ein Blick auf die Übertragbarkeit des Einsatzes als Evaluationsinstrument auf andere etablierte Lehrformate ähnlichen Charakters wie Praktika herausgestellt werden (Ziel 2, vgl. Kapitel 8 auf Seite 99).

14.4.3. Demonstrationspraktika (Lehramt)

Demonstrationspraktika sind in vielen Lehramtsstudiengängen der naturwissenschaftlichen Fächer in Deutschland zu finden und in den Fachdidaktik-Modulen verortet. Auf die Evaluation von Demonstrationspraktika, wie sie häufig in der Lehramtsausbildung vorkommen, lassen sich nicht alle Konstrukte des PraQ übertragen und auch neue Konstrukte wären für eine angepasste Evaluation mutmaßlich von Nöten. So wird hier z. B. die selbstständige Recherche, Planung und das manuelle Aufbauen von Experimenten deutlich stärker betont, als in Grundpraktika (BOCZIANOWSKI et al., 2012). Entsprechende Skalen hierfür könnten nachkonstruiert und testevaluiert werden.

Des Weiteren wird in Demonstrationspraktika die *mündliche Kommunikationskompetenz* mutmaßlich weit stärker gefordert, als in Grundpraktika. So wird bei der Demonstration von Experimenten mit dem Publikum der Aufbau und die Funktionsweise und vieles mehr besprochen, was einerseits eine strukturierte und lerndienliche Kommunikation seitens des/der Demonstrierenden erfordert und andererseits eine stärker adressatengerechte Kommunikation, vor allem dann, wenn die anderen Lehramtsstudierenden Schüler*innen simulieren sollen. Es wird hier zudem stärker auf die korrekte Fachsprache zu achten sein, da eine echte Demonstration vor Publikum vollführt wird und nicht das Experimentieren in geschützten Kleingruppen im Grundpraktikum.

Seitens der Lehrenden in Demonstrationspraktika erscheinen indes das 3L-Modell und somit auch die Skalen des PraQ durchaus übertragbar. So untersuchte jüngst VOLKMER (2017) im Rahmen einer Masterarbeit, welche Konstrukte der Experimentellen Kompetenz nach 3L-Modell in Demonstrationspraktika besonders gefördert werden und nutzte dazu die Skalen des PraQ-A. Ergebnis der Studie ($N = 20$) waren Tendenzen hin zu einer stärkeren Betonung der *Planung von Experimenten* und einer im Vergleich zu Grundpraktika unterdurchschnittlichen Ausprägung der Kompetenzen zur *Auswertung von Experimenten* in Demonstrationspraktika (ebd., S. 40).

Lediglich bei der *Lernumgebungs*-Dimension ist die *Skriptqualität* nur selten geeignet, da schlichtweg in Demonstrationspraktika wenige Skripte eingesetzt werden.

Eine *Integration der Vorlesungsinhalte* wäre indes wünschenswert, allerdings vornehmlich mit Didaktik-Lehrveranstaltungen, da das Demonstrationspraktikum eben jenen Bereich tangiert.

14.4.4. Fortgeschrittenen-Praktika

Bezüglich der Evaluation von sogenannten F-Praktika gilt ähnliches wie für Demonstrationspraktika, es lässt sich manches adaptieren und manches müsste neu hinzukommen. F-Praktika unterscheiden sich von Grundpraktika in der Art, dass weniger Studierende an deutlich komplexeren und (zeit-)aufwändigeren Versuchsaufbauten arbeiten (ZWICKL et al., 2013).

Auch hier wird ähnlich wie im Demonstrationspraktikum, mehr das selbstständige Arbeiten betont, was wiederum andere Kompetenzen fördern kann. So ist beispielsweise eine Kompetenz in forschendem Arbeiten und Recherche zu erwerben, das Instrument nach THIEL & BÖTTCHER (2014) bietet hierfür ergänzende Skalen für F-Praktika über deren Einsatz erneut Expert*innen beraten sollten. Ebenso sind Skalen zur *Versuchsplanung* und zum *Aufbau komplexerer Apparaturen* denkbar. Insgesamt wären je nach Ausgestaltung der F-Praktika also alle PraQ-Skalen für die Evaluation denkbar, plus weitere dem fortgeschrittenen Rahmen angepasste Skalen.

14.4.5. Projektlabore (offenere Praktikumsformen)

Projektlabore stehen teils in direkter Konkurrenz zu naturwissenschaftlichen Grundpraktika (TECHNISCHE-UNIVERSITÄT-BERLIN, 2017). Sie zeichnet in Ähnlichkeit zur Planungs- und Aufbauphase eines Demonstrationspraktikums aus, eine größere Offenheit bezüglich der physischen Realisierung von Versuchsaufbauten zu bieten. Die Präsentation der erarbeiteten Projektergebnisse kann wiederum mit der Demonstration eines Experiments assoziiert werden. Es gelten also für eine Übertragung des PraQ-Instruments auf Projektlabore ähnliche Empfehlungen wie für die oben beschriebenen Demonstrationspraktika.

Literaturverzeichnis

- AAPT (1998). Goals of the Introductory Physics Laboratory. In *American Journal of Physics Teachers* Bd. 66 (6): S. 483.
- ABEL, A. (2014). Beurteilung von Lehrbüchern mithilfe einer Kriterienliste. In *PDN Physik in der Schule* Bd. 3.
- ABRAHAM, M. R.; CRAOLICE, M. S.; GRAVES, A. P. et al. (1997). The nature and state of general chemistry laboratory courses offered by colleges and universities in the United States. In *J. Chem. Educ* Bd. 74 (5): S. 591.
- AERA; APA & NCME (2014). *The Standards for Educational and Psychological Testing*. Washington, D.C.
- AINLEY, J. (1990). School laboratory work: Some issues of policy. In *The student laboratory and the science curriculum*, Routledge, London.
- AKKREDITIERUNGSRAT (2007). *Mission Statement (Beschluss des Akkreditierungsrates vom 18.06.2007)*.
- ALBRECHT, A. (2011). *Längsschnittstudie zur Identifikation von Risikofaktoren für einen erfolgreichen Studieneinstieg in das Fach Physik*. Freie Universität Berlin, Berlin.
- ARENZ, C. & KNOLL, A. (2011). Organisch-Chemisches Praktikum: Skript zum Organisch-Chemischen Praktikum für Studierende der Biologie. 14.07.2017, URL https://www2.chemie.hu-berlin.de/arenz/OC_Praktikumsskript.pdf.
- BECKER, D. (2017). *Untersuchung zur Betreuungs- und Strukturqualität naturwissenschaftlicher Experimentalpraktika*. Masterarbeit an der Freien Universität Berlin in der Didaktik der Physik.
- BERENDT, B. (1994). Gut geplant ist halb gewonnen... In *Handbuch Hochschullehre* Bd. 1: S. 1.
- BERK, R. A. (2005). Survey of 12 Strategies to Measure Teaching Effectiveness. In *International Journal of Teaching and Learning in Higher Education* Bd. 17 (1): S. 48.

- BERLHG (2011). *Gesetz über die Hochschulen im Land Berlin in der Fassung vom 26. Juli 2011*. Berliner Vorschrifteninformationssystem.
- BLK (2002). Modularisierung in Hochschulen. Handreichung zur Modularisierung und Einführung von Bachelor- und Masterstudiengängen. Bd. 101, Bonn.
- BLÜTHMANN, I.; STENZEL, A. & AMELANG, T. (2017). *Unizensus: Key-User-Runde. Präsentation am 03. Mai (intern)*. Freie Universität Berlin, Berlin.
- BLÜTHMANN, I. (2012). *Studierbarkeit, Studienzufriedenheit und Studienabbruch: Analysen von Bedingungsfaktoren in den Bachelorstudiengängen*. Freie Universität Berlin, Berlin.
- BOCZIANOWSKI, F.; MÜLLER, M. & WESTPHAL, N. (2012). Das Demonstrationspraktikum der Humboldt-Universität zu Berlin – eine offene Lehrveranstaltung. In *PhyDid B – Didaktik der Physik: Beiträge zur DPG-Frühjahrstagung*, Berlin.
- BOND, T. G. & FOX, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Lawrence Erlbaum Associates, New Jersey.
- BOND-ROBINSON, J. & RODRIQUES, R. A. B. (2006). Catalyzing graduate teaching assistants' laboratory teaching through design research. In *J. Chem. Educ* Bd. 83 (2): S. 313.
- BORAWSKI, H. & HEINKE, H. (2005). Entwicklung eines Physikpraktikums für Studierende der Biologie. In V. Nordmeier & A. Oberländer (Hrsg.) *Didaktik der Physik Frühjahrstagung Berlin 2005*, Berlin.
- BORTZ, J. (2010). *Statistik für Human- und Sozialwissenschaftler*. Springer, Berlin.
- BORTZ, J. & DÖRING, N. (2006). *Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler*. Springer, Heidelberg.
- BOSBACH, E.; HOCHSCHULREKTORENKONFERENZ & SERVICE-STELLE-BOLOGNA (Hrsg.) (2007). *Neue Texte und Hilfestellungen*

- zur Umsetzung der Ziele des Bologna-Prozesses an deutschen Hochschulen, *Bologna-Reader II*, Bd. 5. Hochschulrektorenkonferenz, Bonn.
- BRAUN, E. & GUSY, B. (2006). Perspektiven der Lehrevaluation. In G. Krampen & H. Zayer (Hrsg.) *Didaktik und Evaluation in der Psychologie*, S. 152, Hogrefe, Göttingen, Bern, Wien, Toronto, Seattle, Oxford, Prag.
- BRAUN, E.; GUSY, B.; LEIDNER, B. et al. (2008). Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). In *Diagnostica* Bd. 1 (54): S. 30.
- BROWN, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. Guilford Press, New York, London.
- BRÜGGEMANN, V. (2015). *Validierung einer Skala zur Erfassung der Skriptqualität in Physik-Experimentierpraktika*,. Abschlussarbeit zum Erwerb des akademischen Grades Master of Education.
- BRYAN, L. A. (2012). *Research on Science Teacher Beliefs*, S. 477. Springer Netherlands, Dordrecht.
- BUCHHAAS-BIRKHOLZ, D. (2009). Die »empirische Wende« in der Bildungspolitik und in der Bildungsforschung. Zum Paradigmenwechsel des BMBF im Bereich der Forschungsförderung. In *Erziehungswissenschaft* Bd. 20 (39): S. 27.
- BÜHNER, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. Pearson Studium, München.
- BUW (2013). *Prüfungsordnung für den Bachelor-Studiengang Physik (zuletzt geändert durch Gesetz vom 18.12.2012)*. Bergische Universität Wuppertal.
- BYRNE, B. N. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Lawrence Erlbaum Associates.
- CHANG, H.-P.; CHEN, C.-C.; GUO, G.-J. et al. (2011). The development of a competence scale for learning science: Inquiry and commu-

- nication. In *International Journal of Science and Mathematics Education* Bd. 9 (5): S. 1213.
- CHEUNG, G. W. & RENSVD, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. In *Structural Equation Modeling: A Multidisciplinary Journal* Bd. 9 (2): S. 233.
- COLLINS, A.; BROWN, J. S. & NEWMAN, S. E. (1987). Cognitive Apprenticeship: Teaching the Craft of Reading, Writing, and Mathematics. In *Thinking: The Journal of Philosophy for Children* Bd. 8 (1): S. 2.
- CRAMER, C. (2010). Kompetenzerwartungen Lehramtsstudierender. Grenzen und Perspektiven selbsteingeschätzter Kompetenzen in der Lehrerbildungsforschung. In *Bildungsstandards und Kompetenzmodelle. Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht*, S. 85–97, Klinkhardt, Bad Heilbrunn.
- CSONKA, N. (2014). *Evaluation von Lehrveranstaltungen an der Humboldt-Universität zu Berlin: Praxisleitfaden für Evaluationsbeauftragte an Fakultäten und Instituten*, Bd. 8. Stabsstelle Qualitätsmanagement der Humboldt-Universität zu Berlin.
- DECI, E. L. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. In *Zeitschrift für Pädagogik* Bd. 2 (39): S. 224.
- DEGEVAL (2008). *Standards für Evaluation der DeGEval – Gesellschaft für Evaluation ('DeGEval-Standards')*.
- DIEMER, U.; BASER, B. & JODL, H.-J. (1998). *Computer im Praktikum: Moderne physikalische Versuche*. Springer, Berlin.
- DITTON, H. & MERZ, D. (1995). Qualität von Schule und Unterricht. In *Zeitschrift für Pädagogik* Bd. 41 (4): S. 507.
- DOMIN, D. S. (1999). A review of laboratory instruction styles. In *J. Chem. Educ* Bd. 76 (4): S. 543.
- DONABENDIAN (1966). Evaluating the Quality of Medical Care. In *The Milbank Quarterly* Bd. 44: S. 166.

- EGGERT, S. (2008). *Bewertungskompetenz für den Biologieunterricht- Vom Modell zur empirischen Überprüfung*. Universität Göttingen, Göttingen.
- EID, M. (2013). *Statistik und Forschungsmethoden*. Beltz, Weinheim.
- EINSIEDLER, W. (2002). Das Konzept »Unterrichtsqualität«. In *Unterrichtswissenschaft* Bd. 53 (2): S. 68.
- ELLIOTT, M. J.; STEWART, K. K. & LAGOWSKI, J. J. (2008). The Role of the Laboratory in Chemistry Instruction. In *Journal of Chemical Education* Bd. 85 (1): S. 145.
- ENGELS, S. & HÜLSBUSCH, C. M. (2008). Vermitteln und Lernen von chemischen Inhalten. Hochschuldidaktische Veränderungen im Rahmen der Veranstaltung »Chemisches Praktikum für Studierende der Biologie und Medizin«. In *Das Hochschulwesen* Bd. 56.
- ENQA (2009). *Standards and Guidelines for Quality Assurance in the European Higher Education Area*. European Association for Quality Assurance in Higher Education, Helsinki.
- ERPENBECK, J. (2003). Einführung. In *Handbuch Kompetenzmessung: Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis*, Bd. 2, Schaeffer & Poeschel, Stuttgart.
- FAUL, F.; ERDFELDER, E.; BUCHNER, A. et al. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. In *Behavior research methods* Bd. 41 (4): S. 1149.
- FIELD, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage, Los Angeles, London, New Delhi, Singapore, Washington DC.
- FISCHER, H. E. (1994). *Physiklernen: eine Herausforderung für Unterrichtsforschung, Didaktik und Naturwissenschaft*, Bd. 3. Lang, Frankfurt.
- FISCHLER, H. (1979). Das Schulbuch im Physikunterricht der Sekundarstufe I. Ungelöste Probleme mit dem Physikbuch. In *Lehrmittel Aktuell: Informationen für die Unterrichtspraxis* S. 28–41.

- FRASER, B. J. & McROBBIE, C. J. (1995). Science Laboratory Classroom Environments at Schools and Universities: A Cross-National Study. In *Educational Research and Evaluation* Bd. 1 (4): S. 289.
- FRESE, B. (2001). *Eine berufsorientierte Ausbildung für Lehramtsstudenten im Fach Anorganische Chemie: Entwicklung und Evaluation eines neuen Praktikumkonzepts auf der Grundlage einer empirischen Erhebung bei Referendaren*. Justus-Liebig-Universität Gießen, Gießen.
- FREY, A. (2005). *Kompetenzstrukturen von Studierenden in der ersten und zweiten Phase der Lehrerbildung*. Verlag Empirische Pädagogik e.V., Landau.
- FREYER, K. (2013). *Zum Einfluss von Studieneingangsvoraussetzungen auf den Studienerfolg Erstsemesterstudierender im Fach Chemie*. Logos, Berlin.
- FRICKE, A. & SCHECKER, H. (2011). Computergestützte Vor- und Nachbereitung des physikalischen Praktikums. In D. Höttecke (Hrsg.) *Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie: Gesellschaft für Didaktik der Chemie und Physik: Jahrestagung in Potsdam 2010*, S. 528–530, LIT, Münster.
- FRICKE, A.; SCHECKER, H. & RÜCKMANN, I. (2011). Hypermedia in der Vorbereitung auf das Physikalische Praktikum. In *PhyDid B – Didaktik der Physik: Beiträge zur DPG-Frühjahrstagung*, Berlin.
- FRÜHWEIN, O. & HEINKE, H. (2007). Umsetzung eines Anfängerpraktikums für Lehramtsstudierende mit Fach Physik. Berufsfeldorientierte Gestaltung einer Fachveranstaltung. In V. Nordmeier & A. Oberländer (Hrsg.) *Didaktik der Physik Frühjahrstagung Regensburg 2007*, Regensburg.
- FUB (2012a). *Entwicklung, Ausgestaltung und Umsetzung von Modulen*. Freie Universität Berlin.
- FUB (2012b). *Studienordnung/Prüfungsordnung für den Bachelorstudiengang Physik des Fachbereichs Physik der Freien Universität Berlin*. Freie Universität Berlin.

- GRÖSCHNER, A.; SCHMITT, C. & SEIDEL, T. (2013). Veränderung subjektiver Kompetenzeinschätzungen von Lehramtsstudierenden im Praxissemester. In *Zeitschrift für pädagogische Psychologie* Bd. 27 (1): S. 77.
- GUTZLER, T.; REHFELDT, D. & NORDMEIER, V. (2014). Technology SUPPORTed Labs (TSL) – multimedial ergänztes Lernen im Praktikum. In *Grundfragen Multimedialen Lehrens und Lernens. Der Qualitätspakt E-Learning im Hochschulpakt 2020. (Tagungsband)* S. 224–236.
- HAAKE, F. & MÜLLER, R. (2011). Physikpraktikum für Studierende der Biologie. In D. Höttecke (Hrsg.) *Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie: Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Potsdam 2010*, S. 540–542, LIT, Münster.
- HAIR, J. F.; ANDERSON, R. E.; TATHAM, R. L. et al. (1998). *Multivariate data analysis*. Prentice Hall, NJ.
- HALLER, K. (1999). *Über den Zusammenhang von Handlungen und Zielen. Eine empirische Untersuchung zu Lernprozessen im physikalischen Praktikum*. Logos, Berlin.
- HARTIG, J. & JUDE, N. (2007). Empirische Erfassung von Kompetenzen und psychometrische Kompetenzmodelle. In *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik*, S. 17–36, BMBF, Bonn, Berlin.
- HARVEY, L. & GREEN, D. (2000). *Qualität und Qualitätssicherung im Bildungsbereich; Schule, Sozialpädagogik, Hochschule*, Kap. Qualität definieren. Fünf unterschiedliche Ansätze, S. 17–39. Beltz, Weinheim.
- HATTIE, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge, New York.
- HEIDRICH, J.; NEUMANN, K. & PETERSEN, S. (2013). *Entwicklung und Validierung eines Experimentiertests*, Gesellschaft für Didaktik der Chemie und Physik, Bd. 33, S. 677–679. LIT, Berlin.

- HEINE, C.; BRIEDIS, K.; DIDI, H. J. et al. (2006). Auswahl- und Eignungsfeststellungsverfahren beim Hochschulzugang in Deutschland und ausgewählten Ländern. Eine Bestandsaufnahme. Report, HIS Hochschul-Informationssystem GmbH.
- HENSE, J. (2012). Innovative Hochschullehre zur Reduktion der Studienabbruchsquote in MINT-Fächern. In D. Oerterer; H. Frey & L. Mandl (Hrsg.) *Universitäre Bildung - Fachidiot oder Persönlichkeit*, S. 70–91, Rainer Hampp, München.
- HENSON, R. K. & ROBERTS, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. In *Educational and Psychological measurement* Bd. 66 (3): S. 393.
- HOFSTEIN, A. (2004). the laboratory in chemistry education: thirty years of experience with developments, implementation, and research. In *chemistry education: research and practice* Bd. 5 (3): S. 247.
- HOFSTEIN, A. & LUNETTA, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. In *Science education* Bd. 88 (1): S. 28.
- HOLM, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. In *Scandinavian Journal of Statistics* Bd. 6 (2): S. 65.
- HOPBACH, A. (2005a). *Hochschule entwickeln, Qualität managen: Studierende als (Mittel)punkt: die Rolle der Studierenden im Prozess der Qualitätssicherung und -entwicklung. 6. Arbeitstagung zur Evaluierung an Hochschulen am 10. und 11. März 2005 an der Fachhochschule für Verwaltung und Rechtspflege Berlin*, Kap. Hochschulen im Wandel: Akteure, Interessen, Strategien, S. 13–22. 10, HRK, Bonn.
- HOPBACH, A. (2005b). *Qualifikationsrahmen für deutsche Hochschulabschlüsse*. KMK, Berlin. Im Zusammenwirken von Hochschulrektorenkonferenz, Kultusministerkonferenz und Bundesministerium für Bildung und Forschung erarbeitet und von der Kultusministerkonferenz am 21.04.2005 beschlossen.
- HRK (2011). *Deutscher Qualifikationsrahmen für lebenslanges Lernen*. Arbeitskreis Deutscher Qualifikationsrahmen.

- HUCKE, L. (1999). *Handlungsregulation und Wissenserwerb in traditionellen und computergestützten Experimenten des physikalischen Praktikums*. Logos, Berlin.
- VAN DER JAGT, S.; VAN RENS, L.; SCHALK, H. H. et al. (2012). Development of a Student Self-Evaluation Instrument in Inquiries. In *Conference Proceedings NARST*.
- JUSSIM, L. (2012). *Social Perception and Social Reality: Why Accuracy Dominates Bias and Self-Fulfilling Prophecy*. Oxford University Press, Oxford, New York.
- KAISER, H. F. (1960). The application of electronic computers to factor analysis. In *Educational and Psychological Measurement* (20): S. 141.
- KANE, M. T. (2001). Current Concerns in Validity Theory. In *Journal of Educational Measurement* Bd. 38 (4): S. 319.
- KAPVO (2012). *Gesetz- und Verordnungsblatt für Berlin*, Bd. 68, Kap. Neunzehnte Verordnung zur Änderung der Kapazitätsverordnung, S. 68–84.
- KATTMANN, U.; DUIT, R.; GROPENGIESSER, H. et al. (1997). Das Modell der Didaktischen Rekonstruktion – Ein Rahmen für naturwissenschaftsdidaktische Forschung und Entwicklung. In *Zeitschrift für Didaktik der Naturwissenschaften* Bd. 3 (3): S. 3.
- KEMBER, D.; LEUNG, D. Y. P. & KWAN, K. P. (2002). Does the Use of Student Feedback Questionnaires Improve the Overall Quality of Teaching? In *Assessment & Evaluation in Higher Education* Bd. 27 (5): S. 411.
- KFP (2010). *Zur Konzeption von Bachelor- und Master-Studiengängen in der Physik. Handreichung der Konferenz der Fachbereiche Physik (KFP)*. Berlin.
- KIND, P. M. & KIND, V. (2007). Creativity in science education: Perspectives and challenges for developing school science. In *Studies in Science Education* Bd. 43 (1): S. 1.

- KISSMANN, F.; MÜLLER, R.; SCHUMACHER, D. et al. (2010). Brauchen Studierende der Biologie ein Physikpraktikum? In *PhyDid B – Didaktik der Physik: Beiträge zur DPG-Frühjahrstagung*, Berlin.
- KLIEME, E.; ARTELT, C. & STANAT, P. (2002). Fächerübergreifende Kompetenzen. In *Leistungsmessungen in der Schule*, S. 203–218, Beltz, Weinheim.
- KLIEME, E.; AVENARIUS, H.; BLUM, W. et al. (2003). *Zur Entwicklung nationaler Bildungsstandards. Grundpositionen einer Expertise*. BMBF, Frankfurt a. M.
- KLINE, R. B. (2011). *Principles and Practice of Structural Equation Modeling, (Methodology In The Social Sciences)*. Guilford Publications, New York, London.
- KMK (2005). *Bildungsstandards im Fach Physik (Chemie / Biologie) für den Mittleren Schulabschluss*. Luchterhand, München.
- KMK (2010). *Ländergemeinsame Strukturvorgaben für die Akkreditierung von Bachelor und Masterstudiengängen. Beschluss der Kultusministerkonferenz vom 10.10.2003 i.d.F. vom 04.02.2010*.
- KOCH, E. (2004). *Gute Hochschullehre: theoriebezogene Herleitung und empirische Erfassung relevanter Lehraspekte*. Kovac, Hamburg.
- KRATSCHMAR, J. (2017). *Erhebung zur Experimentellen Kompetenz in physikalischen Praktika und Validierung eines Evaluationsinstruments zur Praktikumsqualität (PraQ)*. Masterarbeit an der Freien Universität Berlin in der Didaktik der Physik.
- KREITEN, M. (2012). *Chancen und Potenziale web-basierter Aufgaben im physikalischen Praktikum*. Logos, Berlin.
- KRÜGER, D.; NORDMEIER, V.; PANT, H. A. et al. (2017). ValiDiS – Kompetenzmodellierung und -erfassung: Validierungsstudie zum wissenschaftlichen Denken im naturwissenschaftlichen Studium. Förderkennzeichen 01PK15004B.
- KROMREY, H. (2000). Qualität und Evaluation im System Hochschule. In R. Stockmann (Hrsg.) *Evaluationsforschung, Sozialwissenschaftliche Evaluationsforschung*, S. 233–258, Leske + Budrich, Opladen.

- KRÜGER, D.; PARCHMANN, I. & SCHECKER, H. (2013). *Methoden in der naturwissenschaftsdidaktischen Forschung*. Springer Spektrum, Berlin, Heidelberg.
- KULGEMEYER, C. (2010). *Physikalische Kommunikationskompetenz: Modellierung und Diagnostik*. Logos, Berlin.
- LAGOWSKI, J. J. (2002). the role of the laboratory in chemical education. In *International Conference on Chemical Education*, Beijing.
- LAMMERTZ, I. & HEINKE, H. (2014). Schreiben und Peer-Feedback im Physikpraktikum. In *PhyDid B – Didaktik der Physik: Beiträge zur DPG-Frühjahrstagung*, Berlin.
- MAISEYENKA, V. (2014). *Modellbasiertes Experimentieren im Unterricht: Praxistauglichkeit und Lernwirkungen*. Logos, Berlin.
- MANDERLINK, G. & HARACKIEWICZ, J. M. (1984). Proximal versus distal goal setting and intrinsic motivation. In *Journal of Personality and Social Psychology* Bd. 47 (4): S. 918.
- MARSH & BAILEY (1993). Multidimensional Students' Evaluations of Teaching Effectiveness: A Profile Analysis. In *The Journal of Higher Education* Bd. 64 (1): S. 1.
- MARSH, H. W. (1982). SEEQ: A Reliable, Valid, and Useful Instrument for Collecting Students' Evaluations of University Teaching. In *British Journal of Educational Psychology* Bd. 52 (1): S. 77.
- MARSH, H. W. (1987). Students' evaluations of University teaching: Research findings, methodological issues, and directions for future research. In *International Journal of Educational Research* Bd. 11 (3): S. 253.
- MARSH, H. W. & ROCHE, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. In *American Psychologist* Bd. 52 (11).
- MAYER, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen. In *Theorien in der biologiedidaktischen Forschung*, S. 177–186, Springer, Berlin, Heidelberg.

- MAYRING, P. (2001). Combination and Integration of Qualitative and Quantitative Analysis. In *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* Bd. 2 (1).
- MERKT, M. (2016). Zwischen individueller Kompetenzentwicklung und strategischem Qualitätsmanagement. In R. Egger & M. Merkt (Hrsg.) *Teaching Skills Assessments*, Lernweltforschung, S. 7–26, Springer Fachmedien, Wiesbaden.
- MERRITT, M. V.; SCHNEIDER, M. J. & DARLINGTON, J. A. (1993). Experimental design in the general chemistry laboratory. In *J. Chem. Educ* Bd. 70 (8): S. 660.
- MERZYN, G. (1994). *Physikschulbücher, Physiklehrer und Physikunterricht*. IPN, Kiel.
- MESSICK, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. In *American psychologist* Bd. 50 (9): S. 741.
- METZGER, C. & SCHULMEISTER, R. (2011). Die tatsächliche Workload im Bachelorstudium. Eine empirische Untersuchung durch Zeitbudget-Analysen. In *Der Bologna-Prozess aus Sicht der Hochschulforschung*, S. 68, CHE, Gütersloh.
- MEYER-ARNDT, M. (2000). *Interaktionen im Physikpraktikum zwischen Studierenden und Betreuern: Feldstudie zu Bedeutungsentwicklungsprozessen im physikalischen Praktikum*. Logos, Berlin.
- MÜHLENBRUCH, T.; REHFELDT, D. & NORDMEIER, V. (2014). TSL: Interventionsgestaltung im Nebenfachpraktikum. In *PhyDid B – Didaktik der Physik: Beiträge zur DPG-Frühjahrstagung*, Berlin.
- NAGEL, C. C. (2009). *eLearning im Physikalischen Anfängerpraktikum*. Logos, Berlin.
- NASSER & FRESKO (2002). Faculty Views of Student Evaluation of College Teaching. In *Assessment & Evaluation in Higher Education* Bd. 27 (1): S. 187.
- NEUMANN, K. (2004). *Didaktische Rekonstruktion eines physikalischen Praktikums für Physiker*. Logos, Berlin.

- NEUMANN, R. (2000). Communicating Student Evaluation of Teaching Results: Rating Interpretation Guides (RIGs). In *Assessment & Evaluation in Higher Education* Bd. 25 (2): S. 121.
- PLOMER, M. (2011). *Physik physiologisch passend Praktiziert: Eine Studie zur Lernwirksamkeit von traditionellen und adressatenspezifischen Physikpraktika für die Physiologie*. Logos, Berlin.
- POHLENZ, P. (2008). Lehrevaluation und Qualitätsmanagement: neue Anforderungen für die Hochschulsteuerung. In *Sozialwissenschaften und Berufspraxis* Bd. 31 (1): S. 66.
- PORST, R. (2000). Question Wording – Zur Formulierung von Fragebogen-Fragen. Report 2, Zentrum für Umfragen, Methoden und Analysen, Mannheim.
- PRENZEL, M. (1997). Sechs Möglichkeiten, Lernende zu demotivieren. In H. Gruber & A. Renkl (Hrsg.) *Wege zum Können. Determinanten des Kompetenzerwerbs*, S. 32–44, Hans Huber, Bern.
- PRÜFER, P. & REXROTH, M. (2005). Kognitive Interviews. Techn. Ber., Zentrum für Umfragen, Methoden und Analysen, Mannheim.
- PSILLOS, D. & NIEDDERER, H. (2002). Issues and Questions Regarding the Effectiveness of Labwork. In D. Psillos & H. Niedderer (Hrsg.) *Teaching and Learning in the Science Laboratory*, Science & Technology Education Library, S. 21–30, Springer, Netherlands.
- RAMSDEN, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. In *Studies in higher education* Bd. 16 (2): S. 129.
- REHFELDT, D.; BRÜGGEMANN, V. & NORDMEIER, V. (2015a). Der PraQ: Ein Instrument zur Einschätzung von Praktikumsqualität – Skriptqualität. In *PhyDid B – Didaktik der Physik: Beiträge zur DPG-Frühjahrstagung*, Berlin.
- REHFELDT, D.; MÜHLENBRUCH, T. & NORDMEIER, V. (2014). TSL: Quantitative Bedarfs- und Problemanalyse im Physikalischen Praktikum für NaturwissenschaftlerInnen. In *PhyDid B – Didaktik der Physik: Beiträge zur DPG-Frühjahrstagung*, Berlin.

- REHFELDT, D.; MÜHLENBRUCH, T. & NORDMEIER, V. (2015b). TSL: Ergebnisse einer quantitativen Studie zur Problemanalyse im Physikalischen Praktikum für Naturwissenschaftler. In J. Stiller & C. Laschke (Hrsg.) *Berlin-Brandenburger Beiträge zur Bildungsforschung 2015: Herausforderungen, Befunde und Perspektiven interdisziplinärer Bildungsforschung*, S. 203–228, Peter Lang, Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien.
- REINHOLD, P. (2004). Naturwissenschaftsdidaktische Forschung in der Lehrerbildung. In *Zeitschrift für Didaktik der Naturwissenschaften* Bd. 10.
- RICHARDSON, J. T. E. (2005). Instruments for obtaining student feedback: a review of the literature. In *Assessment & Evaluation in Higher Education* Bd. 30 (4): S. 387.
- RIESE, J. & REINHOLD, P. (2009). Fachbezogene Kompetenzmessung und Kompetenzentwicklung bei Lehramtsstudierenden der Physik im Vergleich verschiedener Studiengänge. In *Lehrerbildung auf dem Prüfstand* Bd. 2 (1): S. 104.
- RINDERMAN, H. (1996). *Untersuchungen zur Brauchbarkeit studentischer Lehrevaluationen*. Empirische Pädagogik, Landau.
- ROLLNICK, M.; LUBBEN, F.; LOTZ, S. et al. (2002). What do Underprepared Students Learn about Measurement from Introductory Laboratory Work? In *Research in Science Education* Bd. 32 (1): S. 1.
- ROST, J. (1996). *Lehrbuch Testtheorie, Testkonstruktion*. Huber, Bern.
- ROUQUETTE, A. & FALISSARD, B. (2011). Sample size requirements for the internal validation of psychiatric scales. In *International Journal of Methods in Psychiatric Research* Bd. 20 (4): S. 235.
- RUICKOLDT, G. (1996). Ergebnisse einer Umfrage zum Physikalischen Praktikum. In *Physikalische Blätter* Bd. 10 (52): S. 1022.
- RUNGE, B. (2017). Physikalische Praktika an Universitäten und Hochschulen. 14.07.2017, URL <http://www.physikalische-praktika.de/service/praktika.html>.

- SACHER, M. D.; PROBST, H. M. & REINHOLD, P. J. (2015). Neue Wege zur Konzeption eines kompetenzfördernden, studierenden-zentrierten Laborpraktikums. In *Tagungsband zum 2. HDMINT Symposium*, S. 106–113, Technische Hochschule Nürnberg, Nürnberg.
- SANDER, F. (2000). *Verbindung von Theorie und Experiment im physikalischen Praktikum – Eine empirische Untersuchung zum handlungsbezogenen Vorverständnis und dem Einsatz grafikorientierter Modellbildung im Praktikum*. Logos, Berlin.
- SARIS, W. E.; SATORRA, A. & VAN DER VELD, W. M. (2009). Testing Structural Equation Models or Detection of Misspecifications? In *Structural Equation Modeling: A Multidisciplinary Journal* Bd. 16 (4): S. 561.
- SCHIMANK, U. (2010). Humboldt in Bologna – falscher Mann am falschen Ort? In *Perspektive Studienqualität. Themen und Forschungsergebnisse der HIS-Fachtagung*, S. 44–61, Bertelsmann, Bielefeld.
- SCHMELKIN, L. P.; SPENCER, K. J. & GELLMAN, E. S. (1997). Faculty Perspectives on Course and Teacher Evaluations. In *Research in Higher Education* Bd. 38 (5): S. 575.
- SCHMIDT, U. (2009). Evaluation an deutschen Hochschulen – Entwicklung, Stand und Perspektiven. In T. Widmer; W. Beywl & C. Fabian (Hrsg.) *Evaluation*, S. 163–169, VS Verlag für Sozialwissenschaften, Wiesbaden.
- SCHNEIDER, C.; BODENSOHN, R. & MÜLLER, A. (2012). Das Zusammenspiel pädagogischer Handlungskompetenz und fachdidaktischer Kompetenz in Selbsteinschätzungen angehender Lehrer mathematisch-naturwissenschaftlicher Fächer in der ersten Phase der Ausbildung. In *Lehrerbildung auf dem Prüfstand* Bd. 5 (1): S. 47.
- SCHNEIDER, W. & HASSELHORN, M. (2008). *Handbuch der pädagogischen Psychologie*. Hogrefe, Göttingen, Bern, Wien, Paris, Oxford, Prag, Toronto, Cambridge, Amsterdam, Kopenhagen.
- SCHREIBER, N. (2012). *Diagnostik experimenteller Kompetenz: Validierung technologiegestützter Testverfahren im Rahmen eines Kompetenzstrukturmodells*. Logos, Berlin.

- SCHREIBER, N.; THEYSEN, H. & DICKMANN, M. (2016). Wie genau beurteilen Schülerinnen und Schüler ihre eigenen experimentellen Fähigkeiten? Ein Ansatz zur praktikablen Diagnostik experimenteller Fähigkeiten im Unterrichtsalltag. In *PhyDid A – Physik und Didaktik in Schule und Hochschule* Bd. 1 (15): S. 49.
- SCHROEDTER, S. & KÖRNER, H. (2011). Entwicklung eines Fragebogens zur Selbstwirksamkeitserwartung beim Experimentieren (SWE_EX). In *Vortrag im hochschulinternen Forschungskolloquium der Pädagogischen Hochschule Schwäbisch Gmünd*, Schwäbisch Gmünd.
- SCHULZ, A. (2011). *Experimentierspezifische Qualitätsmerkmale im Chemieunterricht*. Studien zum Physik- und Chemielernen, Logos, Berlin.
- SCHWARZ-HAHN, S. & REHBURG, M. (2003). *Bachelor und Master in Deutschland. Empirische Befunde zur Studienstrukturreform*. Wissenschaftliches Zentrum für Berufs- und Hochschulforschung Universität Kassel.
- SELDIN, P. (1999). *Changing Practices in Evaluating Teaching: A Practical Guide to Improved Faculty Performance and Promotion/Tenure Decisions*, Kap. Current Practices – Good and Bad – Nationally, S. 1–24. Anker, Bolton, MA.
- SERE, M.-G. (2002). Towards renewed research questions from the outcomes of the European project Labwork in Science Education. In *Science Education* Bd. 86 (5): S. 624.
- STAUFENBIEL, T. (2000). Fragebogen zur Evaluation von universitären Lehrveranstaltungen durch Studierende und Lehrende. In *Diagnostica* Bd. 46 (4): S. 169.
- STEFANI, L. A. J. (1994). Peer, self and tutor assessment: Relative reliabilities. In *Studies in Higher Education* Bd. 19 (1): S. 69.
- STRAUBE, P. (2016). *Modellierung und Erfassung von Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung bei (Lehramts-) Studierenden im Fach Physik*, Bd. 20. Logos, Berlin.

- STRAUBE, P. & NORDMEIER, V. (2013). Ko-WADiS - Kompetenzmodell der Erkenntnisgewinnung. In *PhyDid B – Didaktik der Physik: Beiträge zur DPG-Frühjahrstagung*, Berlin.
- TECHNISCHE-UNIVERSITÄT-BERLIN (2017). Projektlabor: Das Projektlabor. 14.07.2017, URL https://www.projektlabor.tu-berlin.de/menue/informationen/das_projektlabor/.
- THEYSSEN, H. (2000). *Didaktische Rekonstruktion eines Physikpraktikums für Medizinstudierende*. Logos, Berlin.
- THIEL, F.; BLÜTHMANN, I. & RICHTER, M. (2010). Ergebnisse der Befragung der Studierenden in den Bachelorstudiengängen an der Freien Universität Berlin - Sommersemester 2010. Report, FU Berlin.
- THIEL, F.; BLÜTHMANN, I. & WATERMANN, R. (2012). Konstruktion eines Fragebogens zur Erfassung der Lehrkompetenz (LeKo). In B. Berendt & H. P. Voss (Hrsg.) *Neues Handbuch Hochschullehre. Lehren und Lernen effizient gestalten.*, Raabe, Berlin.
- THIEL, F. & BÖTTCHER, F. (2014). Modellierung fächerübergreifender Forschungskompetenzen. In B. Berendt & H. P. Voss (Hrsg.) *Neues Handbuch Hochschullehre. Lehren und Lernen effizient gestalten.*, S. 109–124, Raabe, Berlin.
- TIBERGHIE, A.; VEILLARD, L.; MERECHAL, J.-F. L. et al. (1998). An Analysis of Labwork Tasks Used in Science Teaching at Upper Secondary School and University Levels in Several European Countries. In *Science Education* Bd. 85 (5): S. 483.
- UNIVERSITÄT-KÖLN (2010). *Prüfungsordnung für den Bachelorstudiengang Physik der Mathematisch-Naturwissenschaftlichen Fakultät der Universität zu Köln*. Universität zu Köln.
- UNIVERSITÄT-TÜBINGEN (2009). *Prüfungs- und Studienordnung der Universität Tübingen für den Studiengang Physik mit akademischer Abschlussprüfung (Bachelor/ Master-Studiengang)*. Universität Tübingen.
- UNIVERSITÄT-TÜBINGEN (2013). *Prüfungs- und Studienordnung der Universität Tübingen für den Studiengang Physik mit akademischer*

- Abschlussprüfung (Bachelor/ Master-Studiengang)*. Universität Tübingen.
- UNIVERSITÄT-WIEN (2011). *Curriculum für das Bachelorstudium Physik*. Universität Wien.
- URBAN, D. & MAYERL, J. (2014). *Strukturgleichungsmodellierung*. Springer Fachmedien, Wiesbaden.
- VOLKMER, R. M. F. (2017). *Erwerb experimenteller Kompetenz im Lehramtsspezifischen Experimentierpraktikum*. Masterarbeit an der Freien Universität Berlin in der Didaktik der Physik.
- WEAVER, B. & MAXWELL, H. (2014). Exploratory factor analysis and reliability analysis with missing data: A simple method for SPSS users. In *The Quantitative Methods for Psychology* Bd. 10 (2): S. 143.
- WEINERT, F. E. (2001). *Leistungsmessungen in Schulen*. Beltz, Weinheim.
- WEINERT, F. E. (2002). Vergleichende Leistungsmessungen in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.) *Leistungsmessungen in Schulen*, S. 17–32, Beltz, Weinheim, Basel.
- WELLINGTON, J. (1998). *Practical work in school science - Which way now?*, Kap. Practical work in Science: Time for a reappraisal. Routledge, London.
- WELZEL, M.; HALLER, K.; BANDIERA, M. et al. (1998). Ziele, die Lehrende mit dem Experimentieren in der naturwissenschaftlichen Ausbildung verbinden - Ergebnisse einer europäischen Umfrage. In *Zeitschrift für Didaktik der Naturwissenschaften* Bd. 4 (1): S. 29.
- WESTPHAL, W. (1937). Die physikalischen Übungen an der Technischen Hochschule Berlin. In *Zeitschrift für den physikalischen und chemischen Unterricht* Bd. 4: S. 147.
- WILD, K.-P. & SCHIEFELE, U. (1994). Lernstrategien im Studium: Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens. In *Zeitschrift für differentielle und diagnostische Psychologie* Bd. 15 (4): S. 185.

ZASTROW, M. U. (2001). *Interaktive Experimentieranleitungen. Entwicklung und Evaluation eines Konzeptes zur Vorbereitung auf das Experimentieren mit Messgeräten im Physikalischen Praktikum*. Logos, Berlin.

ZWICKL, B. M.; FINKELSTEIN, N. & LEWANDOWSKI, H. J. (2013). The process of transforming an advanced lab course: Goals, curriculum, and assessments. In *American Journal of Physics* Bd. 81 (1): S. 63.

Anhang A

Konstrukte und Items des PraQ – Finale Version

Legende:

= Item mit Ausweichoption »passt nicht«;

* = umgepoltes Item,

fwtheo: Fachwissen Theorie

fwpraxis: Fachwissen Praxis

experw: Experimentelle Kompetenz: Erwartungen formulieren

expverger: Experimentelle Kompetenz: Versuchsplanung & Geräte zusammenstellen

expgerdekla: Experimentelle Kompetenz: Gerätekenntnisse deklarativ

expverauf: Experimentelle Kompetenz: Versuchsanordnung aufbauen

expmess: Experimentelle Kompetenz: Messungen durchführen

expmessdok: Experimentelle Kompetenz: Messungen dokumentieren

expprob: Experimentelle Kompetenz: mit Problemen und Fehlern umgehen

exproh: Experimentelle Kompetenz: Rohdaten graphisch tabellarisch darstellen

expdat: Experimentelle Kompetenz: Daten interpretieren

kommuschrift: Kommunikationskompetenz schriftlich

kommumuendl: Kommunikationskompetenz mündlich

bewfach: Bewertungskompetenz fachlich

koop: Kooperationskompetenz

intr: Interesse

persozeit: Personalkompetenz: Zeitmanagement

ver: verständlich erklären

A. Konstrukte und Items des PraQ – Finale Version

zus: Zusammenfassen & Wiederholen

verkn: Verknüpfungen herstellen

verstand: Verständnis überprüfen

atmo: Gute Lernatmosphäre herstellen

rel: Relevanz verdeutlichen

swe: Ermutigen und Selbstwirksamkeit stärken

stoer: Effizient mit Störungen umgehen

diag.g: Diagnostik: Grundeinstellung

diag.s: Diagnostik: Situativ

skriptver: Skriptqualität für Versuchsablauf

skript: Skriptqualität

int: Integration der Vorlesungsinhalte in das Praktikum

A.1. PraQ-A: Lernzuwachs

Aufgrund dieses Praktikums...

fwtheo.theo ... kann ich wichtige Theorien zu dem behandelten Thema wiedergeben.

fwtheo.ueberblick ... kann ich einen Überblick über die meisten behandelten Themen geben.

fwtheo.typisch ... sehe ich mich nun in der Lage, eine typische Fragestellung des behandelten Gegenstandsbereiches zu bearbeiten.

fwpraxis.quali ... kann ich die Qualität experimenteller Arbeitsweisen besser beurteilen.

fwpraxis.theoexp ... kann ich Bezüge zwischen Theorie und Experiment herstellen.

fwpraxis.fehler ... kenne ich verschiedene Fehlerquellen, die beim Experimentieren auftreten können.

experw.erw ... kann ich konkrete Erwartungen an den Ausgang eines Experiments formulieren.

experw.ausgang ... kann ich nun besser einschätzen, wie ein Experiment "ausgehen" wird (Endergebnis).

experw.konkret: ... sind meine Erwartungen an die experimentellen Ergebnisse eines Versuchs konkreter geworden.

experw.theo: ... kann ich besser einschätzen, welche theoretischen Aspekte ich mir vor dem Experimentieren anschauen muss.

#expverger.mess ... weiß ich in der Planung besser, welche Messgrößen ich während des Experimentierens verändern muss.

#expverger.geraet ... kann ich zum Experiment gehörende Gerätschaften besser auswählen.

#expverger.montieren ... kann ich die einzelnen Geräte besser montieren.

#expverger.eigen ... kann ich Experimente besser nach meinem eigenen Versuchsplan durchführen.

#expverger.problem ... kann ich beim Experimentieren besser auf unerwartete Probleme reagieren.

expgerdekla.eignung ... kann ich die Eignung von Geräten für das Experimentieren besser einschätzen.

expgerdekla.vor ... erkenne ich besser, welches Gerät ich vor mir habe.

expgerdekla.welche ... weiß ich besser, welche Geräte ich konkret für den Versuch/das Experiment benötigen werde.

expverauf.auf ... kann ich Versuchsanordnungen besser aufbauen.

expverauf.test ... kann ich das Funktionieren von Versuchsanordnungen/-aufbauten besser testen.

expverauf.anbringen ... kann ich Bestandteile des Versuchsaufbaus besser anbringen.

expverauf.fehler ... erkenne ich Aufbaufehler besser.

expmess.reihenfolge ... bin ich besser in der Lage, eine Datenaufnahme Schritt für Schritt in logischer Reihenfolge zu gewährleisten.

A. Konstrukte und Items des PraQ – Finale Version

expmess.umgang ... kann ich besser mit Messgeräten umgehen.

expmess.vollstaendig ... kann ich eine vollständige Datenaufnahme besser gewährleisten.

expmess.verfahren ... beherrsche ich mehr Messverfahren als vorher.

expmessdok.proto ... kann ich das Messprotokoll/Laborbuch nachvollziehbarer erstellen.

expmessdok.doku ... bin ich besser in der Lage, eine vollständige Dokumentation der Datenaufnahme zu gewährleisten.

expmessdok.welche ... habe ich einen besseren Überblick, welche Daten im Messprotokoll/Laborbuch dokumentiert werden müssen.

expprob.unsicherheit ... kann ich besser mit Messunsicherheiten umgehen.

expprob.fehler ... beachte ich experimentelle Fehlerquellen adäquater.

expprob.bericht ... kann ich Messunsicherheiten besser berichten (z. B. im Protokoll/Versuchsbericht).

expprob.erkennen ... erkenne ich besser mögliche Fehlerquellen des Experiments.

exproh.tabell ... kann ich Rohdaten besser tabellarisch darstellen.

exproh.graph ... kann ich Rohdaten besser graphisch darstellen.

exproh.geeignet ... fällt es mir leichter, geeignete Darstellungsformen (Grafiken, Tabellen, Schemata) zu erstellen.

expdat.interpretieren ... kann ich besser Messwerte interpretieren.

expdat.frage ... fällt es mir leichter, die Ergebnisse des Versuchs auf die angedachte Fragestellung zu beziehen.

expdat.erwartung ... bin ich besser in der Lage, die Ergebnisse des Versuchs mit meinen anfänglichen Erwartungen zu vergleichen.

expdat.schluss ... kann ich besser Schlussfolgerungen aus meinen Ergebnissen ziehen.

expdat.quali ... kann ich die Qualität meiner Ergebnisse besser einschätzen.

kommumuendl.meinung ... fällt es mir leichter, meine eigenen Eindrücke/Meinungen zu äußern.

kommumuendl.nachfrage ... fällt es mir leichter, nachzufragen, wenn ich etwas nicht verstanden habe.

kommumuendl.diskurs ... kann ich Fragen fachsprachlich besser diskutieren.

kommumuendl.fach ... fällt mir der Gebrauch der korrekten Fachsprache leichter (mündlich).

kommumuendl.verstaendlich

kommuschriftl.praezise ... formuliere ich in meinen Protokollen/Versuchsberichten präziser.

kommuschriftl.verstaendlich ... schreibe ich meine Protokolle/Versuchsberichte verständlicher.

kommuschriftl.gliedern ... kann ich meine Protokolle/Versuchsberichte klarer gliedern.

kommuschriftl.repro ... kann ich Sachverhalte im Protokoll/Versuchsbericht besser schriftlich reproduzieren.

kommuschriftl.kapit ... weiß ich besser, was in welches Kapitel des Protokolls/Versuchsberichtes gehört.

kommuschriftl.fach ... fällt mir der Gebrauch der korrekten Fachsprache leichter (schriftlich).

FILTERFRAGE Ich habe in diesem Praktikum mindestens zwei Wochen in Teams gearbeitet (Partner- oder Gruppenarbeit). [Ja / Nein]

FILTERFRAGE (Bitte beantworten Sie die folgenden fünf Fragen nur, wenn Sie in einer Arbeitsgruppe zusammengearbeitet haben.) [Ja / Nein]

koop.verteilung Durch die Zusammenarbeit mit anderen Studierenden im Praktikum gelingt es mir nun leichter, mich an der Aufgabenverteilung innerhalb einer Gruppe zu beteiligen.

koop.zurueck Durch die Zusammenarbeit mit anderen Studierenden im Praktikum fällt es mir nun leichter, meine eigenen Vorschläge auch mal zurückzunehmen.

koop.konstruktiv Durch die Zusammenarbeit mit anderen Studierenden im Praktikum kann ich mich nun besser für eine konstruktive Arbeitsatmosphäre innerhalb von Teams einsetzen.

koop.absprache Durch die Zusammenarbeit mit anderen Studierenden im Praktikum halte ich mich nun besser an die Absprachen innerhalb einer Gruppe.

intr.aufgabe ...habe ich Spaß daran entwickelt, die an mich gestellten Aufgaben zu lösen.

intr.ausserhalb ...beschäftige ich mich aus Spaß außerhalb der Hochschule mit dem Gegenstandsbereich.

intr.interessant Jetzt finde ich die Themen des Praktikums interessanter als zu Beginn des Praktikums.

intr.begeisterung In diesem Praktikum habe ich Sachen gelernt, die mich begeistern.

intr.studium Das Praktikum hat mich darin bestärkt, mein Studium fortzusetzen.

persozeit.halten ...halte ich mich besser an meine Zeitplanungen für das Praktikum.

persozeit.vorbereitung ...kann ich besser Zeiten festlegen, zu denen ich Praktikumstermine vorbereite.

persozeit.nachbereitung ...kann ich besser Zeiten festlegen, zu denen ich Praktikumstermine nachbereite.

persozeit.orga ...kann ich meine experimentelle Arbeit besser zeitlich organisieren.

bewfach.ziele ...kann ich Ziele und Bedeutung des Versuchs besser schildern.

bewfach.fehler ...kann ich den möglichen Einfluss von Fehlerquellen besser diskutieren.

bewfach.kritisch ...kann ich Ergebnisse von Experimenten besser kritisch diskutieren.

bewfach.gueltig ...kann ich die Gültigkeit empirischer Ergebnisse besser beurteilen.

bewfach.referenz ...kann ich Versuchsergebnisse besser mit Referenzwerten vergleichen.

A.2. PraQ-B: Lehrkompetenz des/der Betreuenden & Lernumgebung

Präambel: Bitte beziehen Sie die folgenden Items NUR auf Ihre/n zuletzt erlebten BetreuerIn (Versuch abgeschlossen).

Der/die Betreuende...

ver.nachvoll ... erklärt Begriffe und Sachverhalte klar und nachvollziehbar (z. B. in Gesprächen; im Vorgespräch; Nachbesprechungen etc.).

ver.komplex ... ist in der Lage, auch einen komplexen Sachverhalt verständlich zu erklären (z. B. in Gesprächen; im Vorgespräch; Nachbesprechungen etc.).

ver.bsp ... gibt anschauliche Beispiele, die zum Verständnis des Sachverhaltes beitragen (z. B. in Gesprächen; im Vorgespräch; Nachbesprechungen etc.).

ver.impuls ... gibt Impulse, die zu einem tieferen Verständnis des Sachverhalts anregen (z. B. in Gesprächen; im Vorgespräch; Nachbesprechungen etc.).

ver.hinweis ... gibt bei Verständnisschwierigkeiten hilfreiche Hinweise.

zus.theohervor ... hebt wichtige Aspekte der Theorie besonders hervor (z. B. in Gesprächen; im Vorgespräch; Nachbesprechungen etc.).

zus.theozusammen ... fasst die wichtigsten Aspekte der Theorie zusammen (z. B. in Gesprächen; im Vorgespräch; Nachbesprechungen etc.).

zus.theovertieft ... wiederholt und vertieft besonders schwierige Aspekte der Theorie ausreichend (z. B. in Gesprächen; im Vorgespräch; Nachbesprechungen etc.).

zus.exphervor ... hebt wichtige Aspekte des Experiments besonders hervor (z. B. in Gesprächen; im Vorgespräch; beim Experimentieren etc.).

zus.expzusammen ... fasst die wichtigsten Aspekte der experimentellen Grundlagen zusammen (z. B. in Gesprächen; im Vorgespräch; beim Experimentieren etc.).

zus.expvertieft ... vertieft besonders schwierige Schritte des Experiments ausreichend (z. B. in Gesprächen; im Vorgespräch; beim Experimentieren etc.).

verkn.vorwissen ... stellt immer wieder Bezüge zum Vorwissen der Studierenden her (z. B. aus der Vorlesung, vorangegangenen Versuchen, Schulwissen etc.).

verkn.behandelt ...stellt immer wieder Bezüge zu den bereits behandelten Sachverhalten her (z.B. aus der Vorlesung, vorangegangenen Versuchen, etc.).

verkn.zusammenhaenge ...verdeutlicht die Zusammenhänge zwischen unterschiedlichen Aspekten der Theorie.

verstand.expfrag ...stellt Fragen, die den Studierenden die Gelegenheit geben, zu überprüfen, ob sie das Experiment verstanden haben.

verstand.expverstanden ...vergewissert sich, dass die Studierenden zentrale Aspekte des Experiments verstanden haben, bevor sie weitermachen.

verstand.theoverstanden ...stellt Fragen, die den Studierenden die Gelegenheit geben, zu überprüfen, ob sie den theoretischen Hintergrund verstanden haben.

atmo.atmo ...sorgt für eine angenehme Lernatmosphäre.

atmo.wert ...zeigt gegenüber allen Studierenden ein respektvolles und wertschätzendes Verhalten.

rel.studium ...macht den Stellenwert der behandelten Sachverhalte für das weitere Studium klar.

rel.bsp ...nennt Beispiele für mögliche Anwendungen der Experimente.

rel.anwendung ...verdeutlicht die Relevanz der Experimente für unterschiedliche Anwendungszusammenhänge.

swe.vertrauen ...vermittelt grundlegendes Vertrauen in die Fähigkeit der Studierenden.

swe.expschwierig ...ermutigt die Studierenden bei der Bewältigung schwieriger Experimente.

swe.misserfolg ...stärkt Studierende bei der Bewältigung von Misserfolgen im Experimentierprozess.

stoer.ruhig ...ist in der Lage, eine ruhige und ungestörte Lernsituation herzustellen.

stoer.effizient ...geht angemessen und effizient mit Unruhe und Störungen um.

stoer.durchsetzung ...beweist bei Unruhe und Störungen das nötige Durchsetzungsvermögen.

- diag.g.zeit** ...nimmt sich die Zeit, um einzelnen Studierenden Dinge zu erklären, die sie nicht verstanden haben.
- diag.g.passiert** ...bemüht sich, dass alle im Praktikum verstehen, was gerade passiert.
- diag.s.sofort** ...merkt sofort, wenn ein/e Studierende/r etwas nicht versteht.
- diag.s.theo** ...merkt, wenn es Probleme beim Verständnis der Theorie gibt.
- diag.s.exp** ...merkt, wenn es Probleme beim Experimentieren gibt.
- skriptver.vorstellung** Das Skript (ggf. Zusatzmaterial) unterstützt mich dabei, eine gute Vorstellung vom Versuchsaufbau zu erhalten.
- skriptver.geraet** Das Skript (ggf. Zusatzmaterial) unterstützt mich dabei, mich mit vorher unbekannten/wenig genutzten Geräten vertraut zu machen.
- skriptver.durchfuehrung** Das Skript (ggf. Zusatzmaterial) unterstützt mich dabei, die Versuche ordnungsgemäß durchzuführen.
- skript.klar** Das Skript erklärt Begriffe und Sachverhalte klar und nachvollziehbar.
- skript.zusammen** Das Skript fasst die wichtigsten Sachverhalte zusammen.
- skript.theo** Das Skript hilft mir beim Erarbeiten der versuchsrelevanten Theorie.
- skript.genutzt** Ich habe die Theorie aus dem Skript auch bei der Versuchsdurchführung genutzt.
- skript.proto** Das Skript (ggf. Zusatzmaterial) unterstützt mich dabei, ordnungsgemäße Protokolle/Versuchsberichte anzufertigen.
- skript.protokorrekt** Das Skript (ggf. Zusatzmaterial) hilft mir, meine Protokolle/-Versuchsberichte fachlich korrekt zu schreiben.
- skript.struktur** Das Skript ist klar und gut strukturiert.
- skript.lernziel** Die Lernziele des jeweiligen Versuchs werden im Skript verdeutlicht.
- skript.graph** Das Skript enthält angemessene Modelle, Graphiken oder Schemata, die das Verständnis komplexer Sachverhalte erleichtern.
- skript.interesse** Aufgrund des Skripts interessiere ich mich mehr für die behandelten Themen.

A. Konstrukte und Items des PraQ – Finale Version

skript.niveau Das Anforderungsniveau des Skripts ist den Voraussetzungen der Studierenden angepasst.

skript.selbststaendig Das Skript ermöglicht selbstständiges Lernen (Vorbereitung, Nachbereitung etc.).

int.vl Die Themen der Vorlesung passen zu den Themen im Praktikum.

int.sachverhalt Die Sachverhalte der Vorlesung sind nützlich für die experimentelle Arbeit im Praktikum.

int.verstehen Die experimentelle Arbeit im Praktikum hilft dabei, die Themen der Vorlesung besser zu verstehen.

Anhang B

Skalenscores PraQ-A Gesamtstichprobe

Tabelle B.1.: Mittlere Skalenscores der PraQ-A-Gesamtstichprobe ($N = 813$).

Konstrukt	M	SD	SE
bewfach	4.14	1.00	0.04
experw	4.15	0.94	0.03
expgerdekla	4.12	1.11	0.04
expmess	4.44	0.95	0.04
expmessdok	4.45	1.11	0.04
expprob	4.52	1.00	0.04
exproh	4.25	1.22	0.04
expdat	4.23	0.94	0.03
expverauf	3.93	1.08	0.04
fwpraxis	4.67	0.93	0.03
fwtheo	4.17	0.99	0.04
intr	3.42	1.29	0.05
kommumuendl	3.64	1.18	0.04
kommuschriftl	3.98	1.11	0.04
koop	3.88	1.26	0.05
persozeit	3.52	1.37	0.05

fwtheo: Fachwissen Theorie; fwpraxis: Fachwissen Praxis; experw: Experimentelle Kompetenz: Erwartungen formulieren; expverger: Experimentelle Kompetenz Versuchsplanung & Geräte zusammenstellen; expgerdekla: Experimentelle Kompetenz: Gerätekenntnisse deklarativ; expverauf: Experimentelle Kompetenz: Versuchsanordnung aufbauen; expmess: Experimentelle Kompetenz: Messungen durchführen; expmessdok: Experimentelle Kompetenz: Messungen dokumentieren; expprob: Experimentelle Kompetenz: mit Problemen und Fehlern umgehen; exproh: Experimentelle Kompetenz: Rohdaten graphisch tabellarisch darstellen; expdat: Experimentelle Kompetenz: Daten interpretieren; kommuschriftl: Kommunikationskompetenz schriftlich; kommumuendl: Kommunikationskompetenz mündlich; bewfach: Bewertungskompetenz fachlich; koop: Kooperationskompetenz; intr: Interesse; persozeit: Personalkompetenz: Zeitmanagement

Anhang C

Skalenscores PraQ-B Gesamtstichprobe

Tabelle C.1.: Mittlere Skalenscores der PraQ-B-Gesamtstichprobe ($N = 501$ bzw. $N = 301$ für *skript* und *int*).

Konstrukt	M	SD	SE
ver	4.69	0.95	0.04
zus	4.58	0.94	0.04
verkn	4.17	0.11	0.05
verstand	4.61	0.10	0.05
atmo	4.84	0.11	0.05
rel	4.01	0.11	0.05
swe	4.16	0.11	0.05
stoer	4.86	0.99	0.04
diag.g	4.88	0.11	0.05
diag.s	4.27	0.11	0.05
skript	4.15	1.13	0.06
int	4.13	1.26	0.07

ver: verständlich erklären; zus: Zusammenfassen & Wiederholen; verkn: Verknüpfungen herstellen; verstand: Verständnis überprüfen; atmo: Gute Lernatmosphäre herstellen; rel: Relevanz verdeutlichen; swe: Ermutigen und Selbstwirksamkeit stärken; stoer: Effizient mit Störungen umgehen; diag.g: Diagnostik: Grundeinstellung; diag.s: Diagnostik: Situativ; skriptver: Skriptqualität für Versuchsablauf; skript: Skriptqualität; int: Integration der Vorlesungsinhalte in das Praktikum

Anhang D

Leitfaden für die kognitiven Expert*inneninterviews der Studie 1.1

Leitfaden für ExpertInneninterviews zu PraKo (Krüger, Parchmann, & Schecker, 2014, p. 122), ggf. Bogner, 2005 06.08.2014

Forschungsfragen

Welche Dimensionen/Items des PraKo stoßen bei ExpertInnen auf Widerstand, der nicht auf mangelnde Passung zu ihrem Praktikum zurückzuführen ist? => ethnographische Dimension Welche Dimensionen/Items des PraKo stoßen bei ExpertInnen auf Widerstand, der auf mangelnde Passung zu Ihrem Praktikum zurückzuführen ist? Dimension »Testkonstruktion«

Grundlagen

1. Ermittelnd, nicht VERmittelnd
2. Alltagssprache
3. Forschungsfrage ungleich Interviewfrage => operationalisieren Interventionen (= Redebeiträge des Interviewers) werden erst gesammelt, dann verdichtet und geordnet,
4. Ziel: wenige gute Fragen mit zunächst wenig Tiefgang
5. Ablauf
 - a) Gesprächsklima herstellen
 - b) Allgemeiner Einstieg [allg] „Wir sitzen hier, weil mir Ihre Vorstellungen zu meinem Fragebogen wichtig sind. Mit unserer Unterhaltung möchte ich herausfinden, wie Sie darüber denken. Das Ziel meiner Untersuchung ist die Verbesserung von Praktika. Ihre Angaben werde ich vertraulich behandeln und anonymisieren. Ich möchte ja gerade Ihre Ansichten kennenlernen. Selbstverständlich bin ich gern bereit, nachher auch Ihre Fragen zu beantworten. Haben Sie dazu noch Fragen?“
 - c) Erzählgenerierende Aufforderung [gut] „Beschreiben Sie bitte einmal kurz, was Ihnen zum Thema „gutes Praktikum“ einfällt.

- d) Bei Fachwörtern nachfragen: [fach], z.B. „Was genau meinen Sie mit experimentelle Fähigkeit?“ Spontane, vertiefende Nachfragen
- e) Filterfragen zu Beginn: [filt1-4] Wird bei Ihnen im Praktikum ein Skript zur Vorbereitung eingesetzt? Wird bei Ihnen im Praktikum ein Vorgespräch/Antestat/Kolloquium o.Ä. vom Betreuenden durchgeführt? Ist es notenrelevant? Wird bei Ihnen im Praktikum immer in Einzelarbeit gearbeitet? Wird bei Ihnen im Praktikum ein Versuchsprotokoll geschrieben? Wie ist es (ganz grob) aufgebaut?
- f) [beva] Welche dieser Bestandteile der Fach/Methoden/Kommunikations-/Kooperations/Personal/Bewertkompetenz sind für ihr Praktikum relevant, welche nicht? Warum? Genauer: Was von den ausgeschlossenen Dingen würden Sie sich in einem idealen Praktikum wünschen?
- g) [leko]Jetzt geht es um die Rolle der Betreuer. Hier sind einige Aspekte festgehalten, die die Betreuer möglicherweise leisten sollen. Welche sind für ihr Praktikum relevant, welche nicht? Warum? Genauer: Was von den ausgeschlossenen Dingen würden Sie sich in einem idealen Praktikum dennoch wünschen?
- h) [lv] Die folgenden Aspekte betreffen die Lehrveranstaltung an sich und die Materialien. Welche sind für ihr Praktikum relevant, welche nicht? Warum? Genauer: Was von den ausgeschlossenen Dingen würden Sie sich in einem idealen Praktikum wünschen? Bei zu starker eigener Zurückhaltung:
- i) [zurück] „Ich weiß, Sie haben das schon einmal erläutert, aber Sie helfen mir, das besser zu verstehen, wenn Sie das jetzt nochmal beschreiben.“
- j) Bei Pausen: Aushalten, bei langen Pausen: [paus] „Bitte sagen Sie, woran Sie jetzt denken.“
- k) Schlussintervention: [vergeß] „Haben wir etwas vergessen, was Sie gerne noch ansprechen möchten?“

Anhang E

Manuale für Erhebungen mittels PraQ außerhalb Berlins

Manual für PraktikumsleiterInnen

Erhebung mit dem Fragebogen zu Praktikumskompetenzen (PraKo)

Daniel Rehfeldt

Generelle Informationen zur Studie

Es handelt sich um eine Pilotierung, d.h. der Fragebogen wird zum ersten mal eingesetzt. Die Pilotierung schaut dabei nur auf das »Funktionieren« des Fragebogens. Eine inhaltliche Auswertung ist überhaupt nur bedingt möglich und dürfte maximal an die PraktikumsleiterInnen weitergetragen werden, ich halte mich vollständig an den Datenschutz und die Forschungsrichtlinien.

Durch den Fragebogen sollen langfristig Erkenntnisse zu Praktika unterschiedlicher Fachrichtungen gewonnen werden, z.B. ob Chemie-Praktika mit Physik-Praktika vergleichbar sind. Auch können daraus Erkenntnisse zur weiteren Entwicklung von Praktika erwachsen.

Ablauf

Bitte geben Sie Ihren Praktikums-BetreuerInnen Folgendes mit:

1. Je BetreuerIn ein »Manual für BetreuerInnen« übergeben und zum durchlesen auffordern
2. Je BetreuerIn EINEN Gruppencode (Kärtchen mit »A2« usw. drauf) übergeben
3. Je BetreuerIn genügend Fragebögen übergeben
4. Je BetreuerIn genügend Süßigkeiten übergeben

Am Ende erhalten Sie von den BetreuerInnen alle Fragebögen. Diese bitte nicht einsehen (Datenschutz), ich werte die Daten aus.

Vielen Dank für Ihre Mithilfe zur Verbesserung von Praktika!

Manual für BetreuerInnen

Erhebung mit dem Fragebogen zu Praktikumskompetenzen (PraKo)

Daniel Rehfeldt

Generelle Informationen zur Studie

Es handelt sich um eine Pilotierung, d.h. der Fragebogen wird zum ersten mal eingesetzt. Die Pilotierung schaut dabei nur auf das »Funktionieren« des Fragebogens. Eine inhaltliche Auswertung ist überhaupt nur bedingt möglich und dürfte maximal an die PraktikumsleiterInnen weitergetragen werden, ich halte mich vollständig an den Datenschutz und die Forschungsrichtlinien.

Durch den Fragebogen sollen langfristig Erkenntnisse zu Praktika unterschiedlicher Fachrichtungen gewonnen werden, z.B. ob Chemie-Praktika mit Physik-Praktika vergleichbar sind. Auch können daraus Erkenntnisse zur weiteren Entwicklung von Praktika erwachsen.

Ablauf

Zeitlicher Umfang der Erhebung

- Einführung: ca. 3min, **Text siehe unten**
- Ausfüllen-Lassen des Fragebogens: ca. 15min
- Einsammeln: ca. 5min

Benötigte Materialien (hat die/der Praktikumsleitende)

- Karton mit Fragebögen in ausreichender Zahl
- Karton mit Süßigkeiten (Süßigkeiten werden **gemeinsam** mit dem Bogen ausgeteilt)
- 1×Kärtchen mit Gruppencode (z.B. »A08«) ⇒ hat der Praktikumsleiter zusammen mit den Fragebögen übergeben

Konkretes Vorgehen

Vor der Erhebung

Bitte dieses Manual durchlesen. Bitte den Gruppencode (z.B. »A8«) bereit halten.

Während der Erhebung

Bitte diejenige Zeit nutzen, die Ihnen der Praktikumsleitende gesagt hat. Falls es keine solche Absprache gab, wäre eine Erhebung zu Beginn des Praktikumstermins wünschenswert, notfalls am Ende. Bitte dazu den folgenden Text vorlesen (auch gern sagen, dass das jetzt vorgelesen ist):

Einführungstext *Heute wird bei uns eine Befragung durchgeführt, in der es um das Praktikum geht. Langfristiges Ziel ist die Erforschung und Verbesserung von Praktika. Das Ganze läuft über eine Doktorarbeit von Daniel Rehfeldt aus Berlin. Es sollte nicht länger als 15min dauern und bietet euch eine Möglichkeit zum Reflektieren über das Praktikum. Es gibt Süßigkeiten als Anerkennung. Wir freuen uns über jeden ausgefüllten Fragebogen, denn jeder Bogen zählt, um ein interpretierbares Ergebnis zu erhalten. Der Rest steht auf dem Fragebogen. Ergebnisse können dann bei Daniel Rehfeldt per Mail erfragt werden.*

Danach bitte die Fragebögen **inkl. je einer Süßigkeit** austeilen und den Gruppencode an eine Tafel etc. schreiben bzw. bei Nachfrage der Studierenden den Gruppencode nennen.

Nach der Erhebung

- Fragebögen bitte umgedreht einsammeln, nicht einsehen (Datenschutz) und dem Praktikumsleiter geben.

Vielen Dank für Ihre Mithilfe zur Verbesserung von Praktika!

Anhang F

Praktikumsskript der Humboldt-Universität zu Berlin: Auszug für Studie 4

Quelle: <https://gpr.physik.hu-berlin.de/Skripten/Elektrodynamik%20und%20Optik/PDF-Dateien/Elektrodynamik%20und%20Optik.pdf>, zuletzt abgerufen am 30.06.2017

Seiten 71–76: »O8 Fraunhofersche Beugung«

Anhang G

Mindestkriterien CFA

Tabelle G.1.: Kriterien der Analyse innerhalb der CFA.

Kriterium	Quelle
Korrelationen der Faktoren in $[-.90; +.90]$	CHANG et al. (2011, S. 1227)
Standardisierte Faktorladungen in $ [.50; .95]$	BROWN (2006)
Faktorvarianz sign. > 0	BROWN (2006)
Standardisierte Residuen größtenteils vom Betrag < 2.58	BROWN (2006)
SRMR < 0.10	BYRNE (2001)
RMSEA < 0.10	BROWN (2006)
CFI $> .90$	KLINE (2011)
TLI $> .90$	BROWN (2006)
modification indices: Standardized expected parameter changes for correlated residuals $< .20$	SARIS et al. (2009, S. 573)

Anhang H

Beispielitem Ko-WADiS-Test zu Auswertung und Interpretation (STRAUBE, 2016)

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 $s = 1 \text{ cm}$ vor einem Geiger-Müller-Zählrohr. Zwischen Zählrohr und Präparat werden nacheinander mehrere, jeweils $b = 2 \text{ mm}$ starke Bleiplatten eingebracht und jeweils für den Zeitraum von $t = 60 \text{ s}$ 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

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.

Danksagung

Ich danke zunächst meinem Doktorvater Herrn Prof. Nordmeier für die freundlichen und produktiven Gespräche zu allen Anliegen meiner Dissertation. Danke, dass Sie mich stets in eigenen Gedanken bestärkt und mich in vielerlei Hinsicht gefördert haben. Ein Dank sei auch an Herrn Prof. Watermann ausgesprochen. Danke dafür, dass Sie mich in empirischen Fragen umfassend beraten haben und mir auch Vorstellungen meines Projekts in Ihrer AG ermöglicht haben.

Für meine empirische Grundausbildung in der Promotionszeit danke ich Dr. André Albrecht für einen psychometrisch angereicherten Promotionsstart, sowie erneut Herrn Prof. Nordmeier für die vielfältigen Fortbildungen, die ich besuchen durfte. Für die kritischen und produktiven Rückmeldungen in den Doktorierendenrunden sei zudem ein Dank an Joachim Haupt, Nikola Schild, Volker Brüggemann, René Dohrmann, Julia Milster, Dorothee Ermel, Philipp Straube, Helen Krofta, Tobias Mühlenbruch, Christina Müller und Stephanie Strelow, sowie Herrn Fischler ausgesprochen.

Für die Teilnahme und tatkräftige Unterstützung bei meinen Studien danke ich den Praktikumsleitenden Herrn Beck, Herrn Bernhardt, Herrn Brockhinke, Herrn Detemple, Herrn Salinga, Herrn Durst, Herrn Feist, Herrn Flesch, Herrn Gonzalez, Herrn Hehl, Herrn Weber, Herrn Lützenkirchen-Hecht, Herrn Koethe, Herrn Müller, Herrn Nagel, Herrn Radtke, Herrn Roth, Frau Schattat, Herrn Schmidt, Herrn Schöps und Frau Stumpff. Allen teilnehmenden Studierenden sei zudem mein aufrichtiger Dank ausgesprochen.

Für die tatkräftige Unterstützung, nicht nur in den letzten Monaten der Doktorarbeit, danke ich von Herzen meiner Frau Dr. Miriam Rehfeldt, die mich nicht nur inspirierte, sondern auch seelisch stets mein Anker war.

Meinem Bruder Roman Rehfeldt sei gedankt für das Lesen und Rückmeldungen zu meiner Arbeit, meinen Eltern Annette und Thomas Binder, sowie Ulla Weber sei gedankt für wichtige Phasen der Erholung.

Des Weiteren sei David Seibert, Christiane Klempin, Tobias Mehrrens, Sandro Wrzalek, Jason Frisch, Cynthia Heiner, Eva Terzer und all den Anderen gedankt, die mir das wissenschaftliche Leben an der Freien Universität so zur Freude machten und machen.

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 Physiknachhilfeunterricht 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 Haerberlen: 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 Metakzepten 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: Physiklehrervorstellungen 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 Maike 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 quasixperimentelle 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 instrukionaler 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 Physiklehrkräften
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 Maiseyenka: 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 Sekundarstufenlehrkräfte*
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 Maike 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
- 200 Helmut Fischler, Elke Sumfleth (Hrsg.): Professionelle Kompetenz von Lehrkräften der Chemie und Physik
ISBN 978-3-8325-4523-9 34.00 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: Modellanwendung 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
- 207 Lars Oettinghaus: Lehrerüberzeugungen und physikbezogenes Professionswissen. *Vergleich von Absolventinnen und Absolventen verschiedener Ausbildungswege im Physikreferendariat*
ISBN 978-3-8325-4319-8 38.50 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
- 211 Markus Bohlmann: Science Education. Empirie, Kulturen und Mechanismen der Didaktik der Naturwissenschaften
ISBN 978-3-8325-4377-8 44.00 EUR
- 212 Martin Draude: Die Kompetenz von Physiklehrkräften, Schwierigkeiten von Schülerinnen und Schülern beim eigenständigen Experimentieren zu diagnostizieren
ISBN 978-3-8325-4382-2 37.50 EUR
- 213 Henning Rode: Prototypen evidenzbasierten Physikunterrichts. *Zwei empirische Studien zum Einsatz von Feedback und Blackboxes in der Sekundarstufe*
ISBN 978-3-8325-4389-1 42.00 EUR
- 214 Jan-Henrik Kechel: Schülerschwierigkeiten beim eigenständigen Experimentieren. *Eine qualitative Studie am Beispiel einer Experimentieraufgabe zum Hooke'schen Gesetz*
ISBN 978-3-8325-4392-1 55.00 EUR
- 215 Katharina Fricke: Classroom Management and its Impact on Lesson Outcomes in Physics. *A multi-perspective comparison of teaching practices in primary and secondary schools*
ISBN 978-3-8325-4394-5 40.00 EUR
- 216 Hannes Sander: Orientierungen von Jugendlichen beim Urteilen und Entscheiden in Kontexten nachhaltiger Entwicklung. *Eine rekonstruktive Perspektive auf Bewertungskompetenz in der Didaktik der Naturwissenschaft*
ISBN 978-3-8325-4434-8 46.00 EUR

- 217 Inka Haak: Maßnahmen zur Unterstützung kognitiver und metakognitiver Prozesse in der Studieneingangsphase. *Eine Design-Based-Research-Studie zum universitären Lernzentrum Physiktreff*
ISBN 978-3-8325-4437-9 46.50 EUR
- 218 Martina Brandenburger: Was beeinflusst den Erfolg beim Problemlösen in der Physik? *Eine Untersuchung mit Studierenden*
ISBN 978-3-8325-4409-6 42.50 EUR
- 219 Corinna Helms: Entwicklung und Evaluation eines Trainings zur Verbesserung der Erklärqualität von Schülerinnen und Schülern im Gruppenpuzzle
ISBN 978-3-8325-4454-6 42.50 EUR
- 220 Viktoria Rath: Diagnostische Kompetenz von angehenden Physiklehrkräften. *Modellierung, Testinstrumentenentwicklung und Erhebung der Performanz bei der Diagnose von Schülervorstellungen in der Mechanik*
ISBN 978-3-8325-4456-0 42.50 EUR
- 221 Janne Krüger: Schülerperspektiven auf die zeitliche Entwicklung der Naturwissenschaften
ISBN 978-3-8325-4457-7 45.50 EUR
- 222 Stefan Mutke: Das Professionswissen von Chemiereferendarinnen und -referendaren in Nordrhein-Westfalen. *Eine Längsschnittstudie*
ISBN 978-3-8325-4458-4 37.50 EUR
- 223 Sebastian Habig: Systematisch variierte Kontextaufgaben und ihr Einfluss auf kognitive und affektive Schülerfaktoren
ISBN 978-3-8325-4467-6 40.50 EUR
- 224 Sven Liepertz: Zusammenhang zwischen dem Professionswissen von Physiklehrkräften, dem sachstrukturellen Angebot des Unterrichts und der Schülerleistung
ISBN 978-3-8325-4480-5 34.00 EUR
- 225 Elina Platova: Optimierung eines Laborpraktikums durch kognitive Aktivierung
ISBN 978-3-8325-4481-2 39.00 EUR
- 226 Tim Reschke: Lesegeschichten im Chemieunterricht der Sekundarstufe I zur Unterstützung von situationalem Interesse und Lernerfolg
ISBN 978-3-8325-4487-4 41.00 EUR
- 227 Lena Mareike Walper: Entwicklung der physikbezogenen Interessen und selbstbezogenen Kognitionen von Schülerinnen und Schülern in der Übergangsphase von der Primar- in die Sekundarstufe. *Eine Längsschnittanalyse vom vierten bis zum siebten Schuljahr*
ISBN 978-3-8325-4495-9 43.00 EUR
- 228 Stefan Anthofer: Förderung des fachspezifischen Professionswissens von Chemielehramtsstudierenden
ISBN 978-3-8325-4498-0 39.50 EUR
- 229 Marcel Bullinger: Handlungsorientiertes Physiklernen mit instruierten Selbsterklärungen in der Primarstufe. *Eine experimentelle Laborstudie*
ISBN 978-3-8325-4504-8 44.00 EUR

- 230 Thomas Amenda: Bedeutung fachlicher Elementarisierungen für das Verständnis der Kinematik
ISBN 978-3-8325-4531-4 43.50 EUR
- 231 Sabrina Milke: Beeinflusst *Priming* das Physiklernen?
Eine empirische Studie zum Dritten Newtonschen Axiom
ISBN 978-3-8325-4549-4 42.00 EUR
- 232 Corinna Erfmann: Ein anschaulicher Weg zum Verständnis der elektromagnetischen Induktion. *Evaluation eines Unterrichtsvorschlags und Validierung eines Leistungsdiagnoseinstruments*
ISBN 978-3-8325-4550-5 49.50 EUR
- 233 Hanne Rautenstrauch: Erhebung des (Fach-)Sprachstandes bei Lehramtsstudierenden im Kontext des Faches Chemie
ISBN 978-3-8325-4556-7 40.50 EUR
- 234 Tobias Klug: Wirkung kontextorientierter physikalischer Praktikumsversuche auf Lernprozesse von Studierenden der Medizin
ISBN 978-3-8325-4558-1 37.00 EUR
- 235 Mareike Bohrmann: Zur Förderung des Verständnisses der Variablenkontrolle im naturwissenschaftlichen Sachunterricht
ISBN 978-3-8325-4559-8 52.00 EUR
- 236 Anja Schödl: FALKE-Physik – Fachspezifische Lehrerkompetenzen im Fach Physik. *Entwicklung und Validierung eines Testinstruments zur Erfassung des fachspezifischen Professionswissens von Physiklehrkräften*
ISBN 978-3-8325-4553-6 40.50 EUR
- 237 Hilda Scheuermann: Entwicklung und Evaluation von Unterstützungsmaßnahmen zur Förderung der Variablenkontrollstrategie beim Planen von Experimenten
ISBN 978-3-8325-4568-0 39.00 EUR
- 238 Christian G. Strippel: Naturwissenschaftliche Erkenntnisgewinnung an chemischen Inhalten vermitteln. *Konzeption und empirische Untersuchung einer Ausstellung mit Experimentierstation*
ISBN 978-3-8325-4577-2 41.50 EUR
- 239 Sarah Rau: Durchführung von Sachunterricht im Vorbereitungsdienst. *Eine längsschnittliche, videobasierte Unterrichtsanalyse*
ISBN 978-3-8325-4579-6 46.00 EUR
- 240 Thomas Plotz: Lernprozesse zu nicht-sichtbarer Strahlung. *Empirische Untersuchungen in der Sekundarstufe 2*
ISBN 978-3-8325-4624-3 39.50 EUR

- 241 Wolfgang Aschauer: Elektrische und magnetische Felder. *Eine empirische Studie zu Lernprozessen in der Sekundarstufe II*
ISBN 978-3-8325-4625-0 50.00 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 Experimentalpraktika an deutschen Hochschulen wurden bisher Befunde erarbeitet, die zeigen, dass es erhebliche Unterschiede zwischen den intendierten Zielen und deren Realisierung gibt. Aus den Erkenntnissen der Praktikumsforschung und den Anforderungen guter Lehrevaluation erwuchs das theoretische, fächerübergreifende Modell der Praktikumsqualität. Der zugehörige Praktikumsqualitäts-Fragebogen (PraQ) wurde ratingbasiert für Studierende in insgesamt 18 Praktika im deutschsprachigen Raum erprobt (N = 1314). Die Testevaluationsstudien zu den Inhalten, zur Struktur und zur Konstruktvalidität zeigten Evidenzen für ein gelungenes Messinstrument. Die Analyse der Forschungsfragen brachte hervor, dass ein Qualitätsbenchmarking für Praktika mittels PraQ gelingt. Physikpraktika wurden in einigen Bereichen der Experimentellen Kompetenz und der Lehrkompetenz der Betreuenden besser bewertet als Chemiepraktika. Semesterpraktika wiesen eine deutlich bessere Betreuungsqualität auf als Blockpraktika, allerdings bewirkten sie teils geringere Kompetenzzuwächse, z. B. im Aufbauen der Versuchsanordnung und im Zeitmanagement.

Als Konsequenz aus diesen Ergebnissen können die jeweiligen Stärken der Praktikumsformen in der Zukunft genauer beobachtet und zielgerichtete Interventionen für schwächere Praktika daraus abgeleitet werden.

Logos Verlag Berlin

ISBN 978-3-8325-4590-1