

Studien zum Physik- und Chemielernen

M. Hopf und M. Ropohl [Hrsg.]

388

Rike Große-Heilmann

Entwicklung fachdidaktischen Wissens zum Einsatz digitaler Medien im Fach Physik

λογος

Studien zum Physik- und Chemielernen

Herausgegeben von Martin Hopf und Mathias Ropohl

Diese Reihe im Logos Verlag Berlin lädt Forscherinnen und Forscher ein, ihre neuen wissenschaftlichen Studien zum Physik- und Chemielernen im Kontext einer Vielzahl von bereits erschienenen Arbeiten zu quantitativen und qualitativen empirischen Untersuchungen sowie evaluativ begleiteten Konzeptionsentwicklungen zu veröffentlichen. Die in den bisherigen Studien erfassten Themen und Inhalte spiegeln das breite Spektrum der Einflussfaktoren wider, die in den Lehr- und Lernprozessen in Schule und Hochschule wirksam sind.

Die Herausgeber hoffen, mit der Förderung von Publikationen, die sich mit dem Physik- und Chemielernen befassen, einen Beitrag zur weiteren Stabilisierung der physik- und chemiedidaktischen Forschung und zur Verbesserung eines an den Ergebnissen fachdidaktischer Forschung orientierten Unterrichts in den beiden Fächern zu leisten.

Martin Hopf und Mathias Ropohl

Studien zum Physik- und Chemielernen

Band 388

Rike Große-Heilmann

**Entwicklung fachdidaktischen Wissens
zum Einsatz digitaler Medien
im Fach Physik**

Logos Verlag Berlin



Studien zum Physik- und Chemielernen

Martin Hopf und Mathias Ropohl [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.



Dieses Werk ist lizenziert unter der Creative Commons Lizenz CC BY-NC-ND (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Die Bedingungen der Creative-Commons-Lizenz gelten nur für Originalmaterial. Die Wiederverwendung von Material aus anderen Quellen (gekennzeichnet mit Quellenangabe) wie z. B. Schaubilder, Abbildungen, Fotos und Textauszüge erfordert ggf. weitere Nutzungsgenehmigungen durch den jeweiligen Rechteinhaber.

Diese Veröffentlichung wurde durch den Open Access-Publikationsfond der Universität Paderborn gefördert.

Logos Verlag Berlin GmbH 2025

ISBN 978-3-8325-5943-4

ISSN 1614-8967

Logos Verlag Berlin GmbH
Georg-Knorr-Str. 4, Geb. 10
D-12681 Berlin

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

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

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

Entwicklung fachdidaktischen Wissens zum Einsatz digitaler Medien im Fach Physik

Dissertation zur Erlangung eines Doktorgrades
der Naturwissenschaftlichen Fakultät
an der Universität Paderborn

vorgelegt von

Rike Isabel Große-Heilmann

aus

Dernbach

Gutachter:

Prof. Dr. Josef Riese (Erstgutachten)

Prof. Dr. Jan-Philipp Burde (Zweitgutachten)

Disputation am 10. Februar 2025

Ein Teil oder Teile dieser Arbeit wurden zuvor veröffentlicht, und zwar in:

1. Große-Heilmann, R. & Riese, J. (2021). Erwerb physikdidaktischen Wissens zum Einsatz digitaler Medien im Physikunterricht. In S. Habig (Hrsg.), *Naturwissenschaftlicher Unterricht und Lehrerbildung im Umbruch? Gesellschaft für Didaktik der Chemie und Physik Online-Jahrestagung 2020*.
2. Große-Heilmann, R., Riese, J., Burde, J.-P., Schubatzky, T. & Weiler, D. (2021). Erwerb und Messung physikdidaktischer Kompetenzen zum Einsatz digitaler Medien. *PhyDid B - Didaktik der Physik - Beiträge zur DPG-Frühjahrstagung 2021 virtuell*, 171–178.
3. Große-Heilmann, R., Burde, J.-P., Riese, J., Schubatzky, T. & Weiler, D. (2022). Messung fachdidaktischer digitaler Kompetenzen in Physik. In S. Habig & H. van Vorst (Hrsg.), *Gesellschaft für Didaktik der Chemie und Physik Jahrestagung virtuell 2021*.
4. Schubatzky, T., Burde, J.-P., Große-Heilmann, R., Riese, J. & Weiler, D. (2022). Das Gesamtuntersuchungsdesign im Verbundprojekt DiKoLeP. In S. Habig & H. van Vorst (Hrsg.), *Gesellschaft für Didaktik der Chemie und Physik Jahrestagung virtuell 2021*.
5. Große-Heilmann, R., Riese, J., Burde, J.-P., Schubatzky, T. & Weiler, D. (2022). Fostering Pre-Service Physics Teachers' Pedagogical Content Knowledge Regarding Digital Media. *Education Sciences*, 12(7), 440. <https://doi.org/10.3390/educsci12070440>
6. Große-Heilmann, R., Burde, J.-P., Riese, J., Schubatzky, T. & Weiler, D. (2023). Erwerb und Messung fachdidaktischen Wissens zum Einsatz digitaler Medien. In H. van Vorst (Hrsg.), *Lernen, Lehren und Forschen in einer digital geprägten Welt: Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Aachen 2022*.
7. Weiler, D., Burde, J.-P., Große-Heilmann, R., Lachner, A., Riese, J. & Schubatzky, T. (2023). Förderung von digitalisierungsbezogenen Kompetenzen von angehenden Physiklehrkräften mit dem SQD-Modell im Projekt DiKoLeP. In M. Meier, G. Greefrath, M. Hammann, R. Wodzinski & K. Ziepprecht (Hrsg.), *Edition Fachdidaktiken. Lehr-Lern-Labore und*

Digitalisierung (S. 47–62). Springer VS. https://doi.org/10.1007/978-3-658-40109-2_4

8. Schubatzky, T., Burde, J.-P., Große-Heilmann, R., Haagen-Schützenhöfer, C., Riese, J. & Weiler, D. (2023). Predicting the development of digital media PCK/TPACK: The role of PCK, motivation to use digital media, interest in and previous experience with digital media. *Computers & Education*, 206, 104900. <https://doi.org/10.1016/j.compedu.2023.104900>
9. Große-Heilmann, R., Burde, J.-P., Riese, J., Schubatzky, T. & Weiler, D. (2024a). Wie sollte ein fachdidaktisches Seminar zum Einsatz digitaler Medien gestaltet sein? In H. van Vorst (Hrsg.), *Frühe naturwissenschaftliche Bildung: Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Hamburg 2023*.
10. Große-Heilmann, R., Burde, J.-P., Riese, J., Schubatzky, T. & Weiler, D. (2024b). Entwicklung fachdidaktischen Wissens zum Einsatz digitaler Medien bei Lehramtsstudierenden im Fach Physik. In B. Herzig, B. Eickelmann, F. Schwabl, J. Schulze & J. Niemann (Hrsg.), *Lehrkräftebildung in der digitalen Welt*. Waxmann. <https://doi.org/10.31244/9783830998372>
11. Große-Heilmann, R., Burde, J.-P., Riese, J., Schubatzky, T. & Weiler, D. (in Druck). Warum erwerben Studierende (kein) fachdidaktisches Wissen zu digitalen Medien im Physikunterricht? In H. van Vorst (Hrsg.), *Lernen, lehren und forschen im Schülerlabor: Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Bochum 2024*.

Danksagung

Bei der Entstehung dieser Arbeit wurde ich von vielen besonderen Menschen unterstützt, wofür ich sehr dankbar bin.

Ein besonderer Dank geht dabei an meinen Doktorvater Josef. Durch deine konstruktiven Ratschläge, deine Wertschätzung und viele hilfreiche und inspirierende Gespräche hast du mir stets Motivation und Kraft für diese Arbeit gegeben. Von dir konnte ich lernen, an welcher Stelle ein gewisser Pragmatismus sinnvoll ist, um den Fokus auf inhaltlich Wichtiges legen zu können. Danke auch für viele unterhaltsame und kulinarisch wertvolle Erlebnisse außerhalb der Arbeit. Ich werde die Zusammenarbeit mit dir vermissen.

Weiterhin danke ich meinen Kolleginnen und Kollegen in den AGs Riese für die gute Zusammenarbeit, den wertvollen inhaltlichen Austausch sowie schöne gemeinsame Zeiten auf Tagungen, in Mittags- und Kaffeepausen oder bei Weihnachtsmarktbesuchen: Ann-Kathrin, Jan und Gerrit in Aachen sowie Madeleine, Yvonne, Alex, Katja und Wilfried in Paderborn. Ich bin sehr froh, ein Teil dieser beiden AGs gewesen zu sein.

Besonders danke ich Doro, Melanie und Jannis als Schnittmenge der beiden AGs. Meiner lieben Bürokollegin Doro danke ich für die so wertvolle Unterstützung und Beratung während dieser Arbeit sowie für die unzähligen Lacher und schönen gemeinsamen Momente im und außerhalb vom Büro. Es ist so schön, dass wir uns auf diesem Weg kennengelernt haben! Liebe Melanie, glücklicherweise konnten wir unser gemeinsames Abschlussarbeiten-Schreiben fortsetzen und auch zusammen promovieren. Ich danke dir für die immer schöne Zusammenarbeit, vor allem in der gemeinsamen Lehre und freue mich schon, dass wir auch den nächsten Ausbildungsabschnitt (zumindest zeitweise) gemeinsam bewältigen können. Jannis, dir danke ich für das Beantworten aller meiner Fragen und Anliegen zu R und zur statistischen Auswertung. Ich bin sehr dankbar, den Wechsel nach Paderborn mit euch zusammen gemacht zu haben. Die gemeinsamen Zugfahrten und die schönen WG-Abende haben die manchmal mühsame Zeit (vor allem auf dem Weg von und nach Paderborn) sehr zum Positiven beeinflusst.

Auch möchte ich meinen Kollegen Jan, David und Thomas im Projekt DiKoLeP danken. Lieber Jan, danke für die Bereitschaft, das Zweitgutachten

zu übernehmen und für deine hilfreiche Unterstützung bei wichtigen Meilensteinen meiner Arbeit. Danke, lieber David, für deine große Bereitschaft und Hilfsbereitschaft bei Projektaufgaben und den immer unkomplizierten Austausch auf dem kurzen Dienstweg. Lieber Thomas, du warst mir in meiner Promotion und im Projekt eine wichtige Bezugsperson und ich danke dir für die wertvolle Beratung und Hilfe, die du mir gegeben hast. Danke euch allen (inklusive Josef) auch für die schönen und lustigen Stunden, die wir auf Tagungen und Projekttreffen zusammen hatten. Ich bin froh, dass wir die produktive und schöne Zusammenarbeit noch etwas weitergeführt haben.

Zudem möchte ich den studentischen Hilfskräften, Bacheloranden und Masteranden an der RWTH Aachen und der Universität Paderborn danken, die mich oder mein Projekt durch ihre Arbeit unterstützt haben. Ein besonderer Dank geht dabei an Sophie aus Aachen für ihre tatkräftige und wertvolle Arbeit und Hilfe in der Interviewanalyse. Ich danke außerdem allen Studierenden, die an den Befragungen zu meiner Arbeit teilgenommen haben, sowie den Kolleginnen und Kollegen der Fachdidaktik an anderen Standorten, die meine Erhebungen unterstützt haben.

Zuletzt danke ich meiner Familie und meinen Freunden, die mich während der Zeit meiner Promotion stets bestärkt haben oder mir durch schöne Erlebnisse und Erfahrungen die nötige Ablenkung geboten haben. Danke meiner lieben Mama und Doro für das Korrekturlesen der Arbeit. Abschließend danke ich besonders dir, Carsten, für deine ständige Unterstützung in dieser Zeit. Danke, dass du immer für mich da bist und mir zur Seite stehst.

Diese Arbeit und das Projekt „*Gemeinsam verschieden sein in einer digitalen Welt – Lehrerbildung an der RWTH Aachen (LeBiAC)*“ wurden im Rahmen der gemeinsamen „Qualitätsoffensive Lehrerbildung“ von Bund und Ländern aus Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01JA1813 gefördert.

Kurzfassung

Der Einsatz digitaler Medien eröffnet neue Potentiale für den Physikunterricht wie beispielsweise eine neue Qualität der Visualisierung oder die Erweiterung experimenteller Möglichkeiten. Um diese Potentiale angemessen im Unterricht nutzen zu können, benötigen angehende Physiklehrkräfte bereits im Studium entsprechende (fachdidaktische) Lerngelegenheiten, um digitalisierungsbezogene Kompetenzen – wie das fachdidaktische Wissen (FDW) zum Einsatz digitaler Medien – erwerben zu können. Zudem sind die Evaluation solcher Lerngelegenheiten hinsichtlich des Wissenserwerbs sowie die Identifikation, welche Aspekte der Lerngelegenheiten dafür förderlich sind, bedeutsam. Zur Messung dieses Wissens wird häufig auf Selbsteinschätzungen zurückgegriffen, welche in ihrer Validität jedoch insbesondere im Vergleich zu Leistungstest zunehmend kritisiert werden. Während Untersuchungen zum FDW von angehenden Physiklehrkräften unter Nutzung etablierter Leistungstests nicht den Einsatz digitaler Medien adressieren, beziehen sich Studien zum digitalisierungsbezogenen FDW nicht auf das Fach Physik.

Vor diesem Hintergrund wird in der vorliegenden Arbeit untersucht, wie sich FDW zum Einsatz digitaler Medien im Fach Physik valide messen und wie sich solches Wissen in physikdidaktischen Lehrveranstaltungen angemessen fördern lässt. Dazu wird ein physikspezifischer Leistungstest zur Erfassung des FDW zum Einsatz digitaler Medien bei Physiklehramtsstudierenden entwickelt und im Hinblick auf mehrere Validitätsaspekte untersucht. Der Leistungstest wird im Prä-Post-Design zur Evaluation eines universitären Lehrkonzepts zum Einsatz digitaler Medien im Physikunterricht genutzt, welches in Form von fachdidaktischen Seminaren an vier universitären Standorten in Deutschland und Österreich implementiert und beforscht wird. Durch anschließende retrospektive Interviews mit einem Teil der Seminarteilnehmenden wird identifiziert, welche Elemente des Lehrkonzepts mit Veränderungen in den Testantworten zusammenhängen und lernförderlich oder lernhinderlich für den Wissenserwerb scheinen. Abschließend werden daraus Schlussfolgerungen für zukünftige Lerngelegenheiten zum Einsatz digitaler Medien abgeleitet.

Aus den Untersuchungen zur Validität ergeben sich mehrere Indizien für eine angemessene Validität des entwickelten Leistungstest hinsichtlich unterschiedlicher Aspekte. Die Prä-Post-Erhebungen in den Seminaren zeigen einen signifikanten Zuwachs im FDW zum Einsatz digitaler Medien mit kleiner Effektstärke. Aus der Interviewanalyse lassen sich insbesondere praktische Erfahrungen mit digitalen Medien als lernförderlich ableiten, aber auch eher theoretische Seminarinhalte zu lernpsychologischen Hintergründen (z. B. Cognitive Load Theory) und einzelnen Medien scheinen hilfreich. Teilweise werden von den Studierenden jedoch auch (v. a. positive) Einzelerfahrungen im Seminar fälschlich verallgemeinert und daraufhin Aufgaben im Post-Test weniger kritisch beantwortet. Demzufolge scheint insbesondere das kritische Reflektieren von Medieneinsätzen und -beispielen wichtig für lernwirksame Lerngelegenheiten zum Einsatz digitaler Medien.

Inhaltsverzeichnis

| | | |
|----------|---|-----------|
| 1 | Einleitung..... | 1 |
| 2 | Fachdidaktisches Wissen zum Einsatz digitaler Medien im Fach Physik | 7 |
| 2.1 | Fachdidaktisches Wissen als Teil professioneller Handlungskompetenz | 7 |
| 2.2 | Fachdidaktisches Wissen von (angehenden) Lehrkräften in den Naturwissenschaften | 12 |
| 2.2.1 | Fachdidaktisches Wissen und pedagogical content knowledge | 12 |
| 2.2.2 | Innere Struktur des Fachdidaktischen Wissens | 16 |
| 2.3 | Fachdidaktisches Wissen zum Einsatz digitaler Medien | 18 |
| 2.3.1 | Beschreibungen zu digitalisierungsbezogenen Kompetenzen | 19 |
| 2.3.2 | Fachtypische Einsatzmöglichkeiten digitaler Medien im Physikunterricht | 22 |
| 2.3.3 | Bedingungen für einen Mehrwert des Medieneinsatzes | 24 |
| 2.4 | Zusammenfassung theoretischer Grundlagen zum Fachdidaktischen Wissen zum Einsatz digitaler Medien | 26 |
| 3 | Untersuchung von Fachdidaktischem Wissen und TPACK | 29 |
| 3.1 | Zugänge in der fachdidaktischen Forschung..... | 29 |
| 3.2 | Kompetenzmessung in der Lehrkräftebildungsforschung..... | 35 |
| 3.2.1 | Verfahren der Kompetenzmessung | 35 |
| 3.2.2 | Aufbau von Leistungstestaufgaben | 38 |
| 3.3 | Güte und Validitätsaspekte von Kompetenzmessungen | 41 |
| 3.3.1 | Hauptgütekriterien Objektivität, Reliabilität und Validität | 41 |
| 3.3.2 | Argumentbasiertes Validitätsverständnis | 44 |
| 3.4 | Messverfahren zum Fachdidaktischen Wissen | 50 |
| 3.5 | Messverfahren zum TPACK-Modell..... | 58 |
| 3.5.1 | Selbsteinschätzungsinstrumente zu TPACK | 60 |
| 3.5.2 | Leistungsbasierte Messungen zu TPACK..... | 63 |

| | | |
|----------|--|------------|
| 3.5.3 | Kombination von Selbsteinschätzung und Leistungsmessung | 68 |
| 3.6 | Zusammenfassung und Schlussfolgerungen für die Messung des Fachdidaktischen Wissens zum Einsatz digitaler Medien | 73 |
| 4 | Forschungsergebnisse zum Fachdidaktischen Wissen und TPACK | 77 |
| 4.1 | Ausgewählte Forschungsergebnisse zum Fachdidaktischen Wissen | .77 |
| 4.1.1 | Entwicklung des Fachdidaktischen Wissens im Lehramt Physik | 79 |
| 4.1.2 | Zusammenhänge von Lerngelegenheiten und Fachdidaktischen Wissen | 82 |
| 4.2 | Ausgewählte Forschungsergebnisse zum TPACK | 85 |
| 4.2.1 | Entwicklung des TPACK im Lehramt für naturwissenschaftliche Fächer | 87 |
| 4.2.2 | Zusammenhänge von Lerngelegenheiten und TPACK | 94 |
| 4.3 | Zusammenfassung des relevanten Forschungsstands zum FDW und TPACK | 97 |
| 5 | Lerngelegenheiten im Verbundprojekt DiKoLeP | 101 |
| 5.1 | Das Verbundprojekt DiKoLeP im Überblick | 101 |
| 5.2 | Fachdidaktische Lehrveranstaltungen an den beteiligten Standorten | 103 |
| 5.2.1 | Übergreifendes Lehrkonzept | 103 |
| 5.2.2 | Lehr-Lern-Seminar am Standort Aachen | 106 |
| 5.2.3 | Seminare an den Standorten Graz, Paderborn und Tübingen | 108 |
| 6 | Ziele der Arbeit | 111 |
| 6.1 | Konsequenzen aus Literatur und bisheriger Forschung | 111 |
| 6.2 | Ziele und Forschungsfragen | 115 |
| 7 | Studiendesign und methodisches Vorgehen | 121 |
| 7.1 | Vorgehen zum quantitativen Teil | 122 |
| 7.2 | Vorgehen zum qualitativen Teil | 127 |
| | Quantitativer Teil | 133 |

| | | |
|-----------|--|------------|
| 8 | Testentwicklung und Pilotierung | 135 |
| 8.1 | Modellierung des FDW zum Einsatz digitaler Medien..... | 135 |
| 8.2 | Aufgabenentwicklung anhand eines Itementwicklungsmodells... | 140 |
| 8.3 | Pilotierung des Testinstruments | 148 |
| 8.4 | Untersuchungen zur Testgüte des optimierten Testinstruments . | 156 |
| 9 | Ergebnisse zur Untersuchung der Validität des Testinstruments . | 167 |
| 9.1 | Abgrenzung von fachunabhängigem pädagogischem Wissen durch Korrelationsanalysen..... | 167 |
| 9.2 | Diskriminante Validierung durch Befragung von Lehramtsstudierenden anderer Fächer | 170 |
| 9.3 | Kognitive Validierung durch Think-Aloud-Studien..... | 173 |
| 9.4 | Inhaltlich curriculare Validierung durch schriftliche Expert:innenbefragungen | 177 |
| 9.5 | Strukturelle Validierung durch Vergleich von IRT-Modellen..... | 181 |
| 9.6 | Zusammenfassung der Ergebnisse zur Untersuchung der Validität des entwickelten FDW-Tests zum Einsatz digitaler Medien | 183 |
| 10 | Vorstellung des entwickelten Testinstruments | 187 |
| 10.1 | Übergeordnete Testmerkmale | 187 |
| 10.2 | Aufgabenbeispiele | 189 |
| 10.2.1 | Aufgabe 3: Interesse am Unterrichtsthema wecken..... | 192 |
| 10.2.2 | Aufgabe 7: Potenzial digitaler Medien zur Berücksichtigung der Heterogenität | 193 |
| 10.2.3 | Aufgabe 8: Vorteile beim Einsatz von Simulationen im Physikunterricht | 195 |
| 10.2.4 | Aufgabe 10: Zielstellung für den Einsatz digitaler Messwerterfassung..... | 196 |
| 10.2.5 | Aufgabe 13: Einbettung eines Erklärvideos in den Unterrichtseinstieg | 197 |
| 11 | Ergebnisse zur Untersuchung des fachdidaktischen Wissenserwerbs | 201 |
| 11.1 | Stichprobe der Haupterhebung zur Untersuchung des Wissenserwerbs im FDW zum Einsatz digitaler Medien | 202 |

| | |
|--|------------|
| 11.2 Ergebnisse zur Untersuchung des Wissenserwerbs im FDW zum Einsatz digitaler Medien..... | 203 |
| Qualitativer Teil | 207 |
| 12 Gestaltung der retrospektiven Interviews und Vorgehen zur Auswertung..... | 209 |
| 12.1 Stichprobe der qualitativen Interviewstudie..... | 209 |
| 12.2 Interviewleitfaden, Zusatzmaterialien und Ablauf | 211 |
| 12.3 Kategoriensystem zur Interviewauswertung..... | 216 |
| 12.3.1 Vorgehen zur Entwicklung der Hauptkategorien und Subkategorien zur Untersuchung von FF3 | 218 |
| 12.3.2 Beschreibung der relevanten Hauptkategorien und Subkategorien zur Untersuchung von FF3 | 221 |
| 13 Ergebnisse der Interviewstudie..... | 227 |
| 13.1 Gründe für die Veränderungen in den Testantworten zum FDW zum Einsatz digitaler Medien..... | 227 |
| 13.2 Identifizierung lernförderlicher und lernhinderlicher Elemente des Lehrkonzepts (FF3) | 229 |
| 13.2.1 Identifikation lernförderlicher Seminarelemente | 231 |
| 13.2.2 Identifikation lernhinderlicher Seminarelemente oder Probleme | 235 |
| 13.2.3 Zusammenfassung und Vergleich mit Erkenntnissen zur subjektiv eingeschätzten Lernförderlichkeit | 238 |
| 13.3 Hypothesenentwicklung: Folgerungen für die Seminargestaltung | 240 |
| 14 Fazit | 245 |
| 14.1 Diskussion der Ergebnisse..... | 245 |
| 14.2 Grenzen der Arbeit | 255 |
| 14.3 Ausblick..... | 262 |
| 15 Literaturverzeichnis | 265 |
| Anhang..... | 291 |

Abbildungsverzeichnis

| | |
|--|-----|
| Abbildung 2-1: Modell professioneller Handlungskompetenz und innerer Struktur des Professionswissens nach Baumert und Kunter (2006, S. 482) . | 9 |
| Abbildung 2-2: Kompetenzstrukturmodell zur professionellen Handlungskompetenz von (angehenden) Lehrkräften in Physik (Riese, 2009, S. 26; in Anlehnung an Baumert & Kunter, 2006; Blömeke et al., 2008) | 10 |
| Abbildung 2-3: Darstellung des Refined Consensus Model (RCM) of PCK (Carlson et al., 2019b, S. 84) | 14 |
| Abbildung 2-4: Modell zum Fachdidaktischen Wissen in Physik nach Gramzow et al. (2013) bzw. Gramzow (2015) | 17 |
| Abbildung 2-5: TPACK-Modell (Koehler & Mishra, 2009, S. 63; Mishra & Koehler, 2006)..... | 20 |
| Abbildung 2-6: Struktur des DiKoLAN-Orientierungsrahmens für digitale Basiskompetenzen (https://dikolan.de/ , letzter Zugriff am 20.02.2024)..... | 21 |
| Abbildung 2-7: Das SAMR-Modell nach Puentedura (2012) (Übersetzung nach A. Wilke auf http://homepages.uni-paderborn.de/wilke/blog/2016/01/06/SAMR-Puentedura-deutsch/ , letzter Zugriff am 25.04.2024). | 25 |
| Abbildung 3-1: Übersicht zu offenen und geschlossenen Antwortformaten mit Beispielen sowie Vor- und Nachteilen in Anlehnung an Hartig und Jude (2007, S. 30) | 40 |
| Abbildung 3-2: Itementwicklungsmodell für das FDW in Physik im Projekt ProfiLe-P (Gramzow, 2015, S. 104)..... | 53 |
| Abbildung 5-1: Struktur des Lehrkonzepts im Verbundprojekt DiKoLeP aus gemeinsamen Kerninhalten (Theorie) und standortspezifischen praktischen Anteilen (Praxis) | 105 |
| Abbildung 7-1: Studiendesign zur Untersuchung der Forschungsfragen . | 121 |
| Abbildung 8-1: Modellierung zum FDW zum Einsatz digitaler Medien im Fach Physik mit den vier Kategorien <i>fachbezogene Grundlagen</i> , <i>digitale Messwert-erfassung (dMWE)</i> , <i>Simulationen</i> und <i>Erklärvideos (EV)</i> | 138 |

| | |
|---|-----|
| Abbildung 8-2: Itementwicklungsmodell zur Entwicklung der Testaufgaben zum FDW zum Einsatz digitaler Medien im Fach Physik (in Anlehnung an Gramzow, 2015)..... | 141 |
| Abbildung 8-3: Beispielaufgabe (Mehrfachwahlaufgabe) aus der Testversion zur Pilotierung | 151 |
| Abbildung 8-4: Beispielaufgabe (offene Aufgabe) aus der Testversion zur Pilotierung. | 152 |
| Abbildung 8-5: Wright Map (für die eindimensionale IRT-Modellierung).. | 161 |
| Abbildung 8-6: Wright Maps mit Farbkodierung nach Kategorien und kognitiven Anforderungen | 163 |
| Abbildung 9-1: Beispielaufgabe aus dem eingesetzten Test zum FDW in Physik (Jordans et al., 2022) (Facette: Fachdidaktische Konzepte; Musterlösung: ja, ja, nein, nein, ja) | 168 |
| Abbildung 9-2: Beispielaufgabe aus dem eingesetzten Test zum PW (Seifert & Schaper, 2012) (richtige Antworten: erste, vierte und fünfte Option) | 169 |
| Abbildung 9-3: Verteilung zur Kategorisierung des bei der Testbearbeitung genutzten Wissens im Rahmen von Think-Aloud-Studien (Conrads, 2022) | 175 |
| Abbildung 9-4: Beispiel aus der Befragung zur curricularen Passung der Testaufgaben..... | 178 |
| Abbildung 9-5: Häufigkeitsverteilung zur curricularen Passung der 15 Testaufgaben..... | 179 |
| Abbildung 9-6: Latente Korrelationen der vier Dimensionen (Kategorien) im vierdimensionalen IRT-Modell..... | 182 |
| Abbildung 10-1: Aufgabe 3 des finalen Testinstruments | 193 |
| Abbildung 10-2: Aufgabe 7 des finalen Testinstruments | 194 |
| Abbildung 10-3: Aufgabe 8 des finalen Testinstruments | 195 |
| Abbildung 10-4: Aufgabe 10 des finalen Testinstruments | 196 |
| Abbildung 10-5: Aufgabe 13 des finalen Testinstruments | 198 |
| Abbildung 11-1: Wright-Map zur Stichprobe in der Haupterhebung | 204 |

| | |
|---|-----|
| Abbildung 12-1: Ausschnitt aus der Tabelle zur Einschätzung der subjektiven Lernförderlichkeit der Seminarinhalte und -tätigkeiten | 213 |
| Abbildung 12-2: Ausschnitt aus dem Interviewleitfaden zur Hautphase B (individuelle Vergleiche der Testantworten in Prä- und Posttest) | 214 |
| Abbildung 12-3: Beispiel zu einer im Interview vorgelegten Testaufgabe mit Veränderungen von Prä- zu Posttest | 215 |
| Abbildung 12-4: Übersicht der entwickelten Subkategorien zur Hauptkategorie <i>Änderungen in Testantworten</i> | 222 |
| Abbildung 13-1: Übersicht zu identifizierten Gründen für Veränderungen in den Testantworten im FDW zum Einsatz digitaler Medien im Fach Physik (hervorgehoben sind die Subkategorien zu direktem oder vermutetem Seminarbezug, da diese für FF3 bedeutsam sind) | 228 |
| Abbildung 13-2: Entwickelte Hypothesen zur Gestaltung von Lerngelegenheiten zum fachdidaktisch sinnvollen Einsatz digitaler Medien im Physikunterricht | 243 |

Tabellenverzeichnis

| | |
|---|-----|
| Tabelle 7-1: Allgemeines Schema für Kategoriendefinitionen (Kuckartz, 2018, S. 40)..... | 129 |
| Tabelle 8-1: exemplarische Darstellung der tabellarischen Übersicht zur Modellierung der Facette (Digitale) Medien des FDW bzw. des FDW zum Einsatz digitaler Medien im Fach Physik | 137 |
| Tabelle 8-2: Beschreibung und Anwendung der Kategorien des Itementwicklungsmodells | 142 |
| Tabelle 8-3: Beschreibung und Anwendung der kognitiven Anforderungen des Itementwicklungsmodells | 143 |
| Tabelle 8-4: Übersicht der Stichprobe zur Pilotierung des entwickelten Testinstruments | 149 |
| Tabelle 8-5: Übersicht zur Verteilung der 17 Testaufgaben für die Pilotierung in Kategorien und kognitive Anforderungen..... | 150 |
| Tabelle 8-6: Kritische Aufgaben der Pilotierung und vorgenommene Überarbeitungen | 155 |
| Tabelle 8-7: statistische Kennwerte zu Modellpassung, EAP-Reliabilitäten und mittleren Schwierigkeiten der Aufgabenschwellen für die eindimensionale sowie die beiden mehrdimensionalen Modellierungen nach Kategorien bzw. kognitiven Anforderungen..... | 159 |
| Tabelle 9-1: Manifeste Korrelationen des FDW zum Einsatz digitaler Medien im Fach Physik mit den Leistungstests zum FDW im Fach Physik und PW (Pearson-Korrelation) | 170 |
| Tabelle 9-2: Deskriptive Statistik der befragten Lehramtsstudierenden (Physik oder andere Fächer bzw. kein Physik) zur Testleistung im FDW zum Einsatz digitaler Medien..... | 171 |
| Tabelle 9-3: Ergebnisse der Regressionsanalyse zur Testleistung (Summenscore) im FDW zum Einsatz digitaler Medien (N = 94; $R^2 = 0,193$; korr. $R^2 = 0,175$; $F(2,91) = 10,881$; $p < 0,001$)..... | 172 |

| | |
|--|-----|
| Tabelle 9-4: Beschreibung und Anwendung zu verwendeten Wissensarten bei der Testbearbeitung im Rahmen einer Masterarbeit zur kognitiven Validierung (Conrads, 2022) | 174 |
| Tabelle 9-5: Überarbeitungen der Testaufgaben nach der curricularen Validierung | 180 |
| Tabelle 9-6: Vergleich zwischen eindimensionalem IRT-Modell ohne Unterteilung (1D-Modell) und vierdimensionalem IRT-Modell mit der Unterteilung in vier Kategorien (4D-Modell)..... | 182 |
| Tabelle 10-1: Verteilung der 14 Testaufgaben des finalen Testinstruments auf Kategorien und kognitive Anforderungen..... | 190 |
| Tabelle 10-2: Anwendung der Kategorien und Zuordnung der zugehörigen Aufgaben | 190 |
| Tabelle 10-3: Anwendung der kognitiven Anforderungen und Zuordnung der zugehörigen Aufgaben | 191 |
| Tabelle 11-1: Übersicht zur Stichprobe in der Hauptstudie zur Untersuchung der Entwicklung im FDW zum Einsatz digitaler Medien in den beteiligten Seminaren (FF1)..... | 202 |
| Tabelle 11-2: Ergebnisse in Prä- und Posttest im FDW zum Einsatz digitaler Medien..... | 205 |
| Tabelle 12-1: Übersicht zur Stichprobe in den retrospektiven Interviews im Vergleich zur übergeordneten Stichprobe der quantitativen Befragung... | 210 |
| Tabelle 12-2: Grobstruktur des Interviewleitfadens zur Untersuchung von FF3..... | 212 |
| Tabelle 12-3: Übersicht zu den Hauptkategorien zur Interviewauswertung und ihrer inhaltlichen Beschreibung | 217 |
| Tabelle 13-1: Anzahl der Interviewsegmente, die Überschneidungen zwischen den Subkategorien zur Testbewertung und Subkategorien mit (vermutetem) Seminarbezug darstellen | 230 |
| Tabelle 13-2: Kategorien zur Identifikation lernförderlicher Seminarelemente und Angabe der Häufigkeiten bei Überschneidungen <i>konkreter Seminarerfahrung (Seminarbezug vermutbar)</i> mit <i>Verbesserung</i> | 232 |

Tabelle 13-3: Kategorien zur Identifikation lernhinderlicher Seminarelemente oder möglicher Probleme und Angabe der Häufigkeiten bei Überschneidungen *konkreter Seminarerfahrung (Seminarbezug vermutbar)* mit *Verschlechterung*..... 236

Abkürzungsverzeichnis

| | |
|------------|--|
| AIC | <i>Akaike Information Criterion</i> |
| BIC | <i>Bayesian Information Criterion</i> |
| BMBF | Bundesministerium für Bildung und Forschung |
| COACTIV | Verbundforschungsprojekt „Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung mathematischer Kompetenz“ |
| CLT | <i>Cognitive Load Theory</i> |
| CTML | <i>Cognitive Theory of Multimedia Learning</i> |
| DigCompEdu | Europäischer Rahmenplan „Digital Competence Framework für Educators“ |
| DiKoLAN | Orientierungsrahmen „Digitale Kompetenzen für das Lehramt Naturwissenschaften“ |
| DiKoLeP | Verbundforschungsprojekt „Digitale Kompetenzen von Lehramtsstudierenden im Fach Physik“ |
| EAP | <i>Expected-a-posteriori</i> -Schätzung |
| FALKO | Verbundforschungsprojekt „Fachspezifische Lehrerkompetenzen“ mit dem Teilprojekt FALKO-P zu Physik |
| FDW | Fachdidaktisches Wissen |
| ICAP | ICAP-Modell mit vier Stufen <i>Interactive, Constructive, Active, Passive</i> |
| IRT | <i>Item-Response</i> -Theorie |
| KeiLa | Verbundforschungsprojekt „Kompetenzentwicklung in mathematischen und naturwissenschaftlichen Lehramtsstudiengängen“ |
| KiL | Verbundforschungsprojekt „Messung professioneller Kompetenzen in mathematischen und naturwissenschaftlichen Lehramtsstudiengängen“ |
| KMK | Kultusministerkonferenz |

| | |
|-------------|--|
| LeBiAC | Projekt „Lehrerbildung in Aachen“ der Qualitätsoffensive Lehrerbildung |
| MINT-Fächer | Zusammenfassende Bezeichnung der Fächer Mathematik, Informatik, Naturwissenschaften und Technik |
| MNSQ | Mean Square |
| MNU | Deutscher Verein zur Förderung des mathematischen und naturwissenschaftlichen Unterrichts e.V. |
| PCK | <i>Pedagogical Content Knowledge</i> |
| ProfiLe-P | Verbundforschungsprojekt „Professionswissen in der Lehr- amtsausbildung Physik“ mit den anknüpfenden Projektpha- sen ProfiLe-P+ und ProfiLe-P Transfer |
| ProwiN | Verbundforschungsprojekt „Professionswissen von Lehr- kräften in den Naturwissenschaften“ |
| RCM | <i>Refined Consensus Model of PCK in Science Education</i> |
| SAMR | SAMR-Modell mit vier Ebenen <i>Substitution, Augmentation,</i> <i>Modification, Redefinition</i> |
| SMSR | <i>Standardized Root Mean Square Residual</i> |
| SPSS | Schritte zur Interviewleitfadenerstellung: Sammeln, Prüfen, Sortieren, Subsummieren |
| SQD | <i>Synthesis of Qualitative Evidence</i> |
| SWK | Ständige Wissenschaftliche Kommission der Kultusminis- terkonferenz |
| RMSD | <i>Root Mean Square Error of Approximation</i> |
| TPACK | <i>Technological Pedagogical Content Knowledge</i> |
| WLE | <i>Weighted Likelihood Estimation</i> |

1 Einleitung

„Digitale Medien ermöglichen unzählige technologische Aufbereitungs- und Interaktionsmöglichkeiten für Inhalte. Diese Möglichkeiten dürfen bei der Entwicklung digitaler Medien nicht beliebig ausgeschöpft werden, sondern müssen in ihrer fachdidaktischen Funktion auf den jeweiligen fachlichen Kompetenzaspekt abgestimmt sein.“ (Ständige Wissenschaftliche Kommission der Kultusministerkonferenz [SWK], 2022, S. 41–42)

Die Digitalisierung stellt eine große Herausforderung für Gesellschaft und Bildung dar. Dies bezieht sich auf alle Phasen und Arten von Bildung und betrifft damit sowohl die Bildung an (allgemeinbildenden) Schulen als auch die Lehrkräftebildung (Kultusministerkonferenz [KMK], 2016). Die Bedeutung digitaler Medien in Schule und Unterricht wird dadurch zunehmend größer.

Dabei ist die Nutzung digitaler Medien im Unterricht nicht nur bildungspolitisch erwünscht, sondern bietet bei sinnvollem Einsatz auch Potentiale für den Unterricht und das fachliche Lernen (Hillmayr et al., 2020). So ermöglichen digitale Medien die Nutzung verschiedener Darstellungsformen, multipler Perspektiven und dynamisch-interaktiver Visualisierungen und bieten damit ein reichhaltiges Informationsangebot sowie die Möglichkeit unterschiedlicher Zugänge auf einen Lerngegenstand (SWK, 2022). Gerade im naturwissenschaftlichen Unterricht kann durch die Nutzung digitaler Medien eine neue Qualität der Visualisierung geschaffen werden, um Lernende beim Erwerb konzeptionellen Verständnisses zu unterstützen (Girwidz, 2020c). Der Einsatz digitaler Medien im Unterricht ist dabei jedoch nicht als Automatismus oder Selbstläufer für fachliches Lernen zu verstehen. Vielmehr sollte für einen gelungenen Medieneinsatz dieser stets passend zum fachlichen Lernziel oder Lerngegenstand legitimiert sein (Girwidz, 2020b; Schwanewedel et al., 2018; SWK, 2022).

Um digitale Medien entsprechend sinnstiftend in den eigenen Fachunterricht implementieren und die möglichen Potentiale angemessen nutzen zu können, sollten angehende Lehrkräfte bereits in der ersten Phase der Lehrkräftebildung im Hinblick auf den Erwerb (fachdidaktischer) digitalisierungsbe-

zogener Kompetenzen gefördert werden (KMK, 2016; SWK, 2022). Es ist demnach notwendig, dass entsprechende lernwirksame Lerngelegenheiten bereits in der ersten Phase der Lehrkräftebildung integriert werden, um angehende Lehrkräfte auf den sinnstiftenden Einsatz digitaler Medien im Fachunterricht vorzubereiten (KMK, 2016). Die Implementation solcher Lerngelegenheiten zur Förderung digitalisierungsbezogener Kompetenzen ist dabei nicht nur Aufgabe der Bildungswissenschaften oder einer übergeordneten Mediendidaktik, sondern ebenso der einzelnen Fächer und Fachdidaktiken (KMK, 2016; 2019; Schmid et al., 2020; SWK, 2022). Dass dieser Notwendigkeit nachgegangen wird, äußert sich beispielsweise bereits in einigen innovativen Lehrveranstaltungskonzepten zur Förderung digitalisierungsbezogener Kompetenzen (z. B. Bundesministerium für Bildung und Forschung [BMBF], 2018). Um jedoch auch die Frage der Lernwirksamkeit zu berücksichtigen, ist eine Evaluation solcher Lerngelegenheiten unerlässlich.

In diesem Zusammenhang wird zur Untersuchung des digitalisierungsbezogenen Wissenserwerbs in den Lerngelegenheiten häufig das TPACK-Modell nach Mishra und Koehler (2006) herangezogen, welches verschiedene (technologiebezogene) Wissensbereiche des Professionswissens von Lehrkräften unterscheidet und als zentralen Wissensbereich technologiebezogenes fachdidaktisches Wissen bzw. TPACK berücksichtigt. Zur Erfassung dieser Wissensbereiche wird in der Regel auf Messinstrumente zur Selbsteinschätzung zurückgegriffen (Wang et al., 2018; Willermark, 2018). Diese Selbsteinschätzungsinstrumente sind jedoch hinsichtlich ihrer Validität und mangelnden Fachspezifität in Kritik (z. B. Akyuz, 2018; Kopcha et al., 2014; Kotzebue, 2022a; Lachner et al., 2019; Willermark, 2018) und messen scheinbar eher Selbstwirksamkeitserwartungen anstelle von Professionswissen (Backfisch et al., 2020b; Kotzebue, 2022a; Lachner et al., 2019; Scherer et al., 2017). Neben der Erfassung des technologie- oder digitalisierungsbezogenen Professionswissens ist für die Evaluation entsprechender Lerngelegenheiten zum Einsatz digitaler Medien weiterhin bedeutsam, welche Aspekte lernwirksame Lerngelegenheiten ausmachen und förderlich für die Wissensentwicklung sind.

Die vorliegende Arbeit beschäftigt sich vor diesem Hintergrund mit den zentralen Fragestellungen (1) wie sich TPACK bzw. fachdidaktisches Wissen zum

Einsatz digitaler Medien im Fach Physik valide messen lässt und (2) wie sich dieses Wissen in physikdidaktischen Lehrveranstaltungen angemessen fördern lässt. Die Arbeit ist Teil des Verbundprojekts DiKoLeP („Digitale Kompetenzen von Lehramtsstudierenden im Fach Physik“) der Universitäten Graz, Innsbruck, Paderborn und Tübingen, welches sich aufgrund oben beschriebener Ausgangslage der Entwicklung und Evaluation eines universitären Lehrkonzepts zur Förderung digitaler Kompetenzen im Fach Physik widmet (Schubatzky et al., 2022). Die Evaluation des Lehrkonzepts erfolgt dabei unter anderem im Hinblick auf den Erwerb fachdidaktischen Wissens zum Einsatz digitaler Medien. Dabei wird ein physikdidaktischer Leistungstest als proximales Messverfahren verwendet, um Validitätsprobleme von Selbsteinschätzungen in diesem Bereich zu überwinden.

Das erste Ziel dieser Arbeit stellt daher die Modellierung des fachdidaktischen Wissens (FDW) zum Einsatz digitaler Medien im Fach Physik als Ausgangslage für eine anknüpfende Testerstellung dar. Das zweite Ziel bildet daraufhin die Entwicklung eines Leistungstests zur Erfassung dieses Wissens bei Lehramtsstudierenden im Fach Physik sowie die Untersuchung der Validität im Hinblick auf eine angemessene Nutzbarkeit des Tests. Mithilfe des entwickelten Testinstruments wird das dritte Ziel dieser Arbeit verfolgt, die Entwicklung des FDW zum Einsatz digitaler Medien bei Lehramtsstudierenden in den physikdidaktischen Seminaren im Projekt DiKoLeP zu untersuchen. Das vierte Ziel der Arbeit stellt die Erkundung von Gründen für Veränderungen im gemessenen FDW dar sowie die Identifikation von lernförderlichen oder lernhinderlichen Seminarelementen. Dazu folgt anknüpfend an die quantitativen Erhebungen eine retrospektive Interviewstudie mit einem Teil der Seminarteilnehmenden. In dieser Interviewstudie wird erkundet, wie die Teilnehmenden ihre Veränderungen im gemessenen FDW zum Einsatz digitaler Medien über das Seminar begründen und welche Seminarelemente sich als lernförderlich erweisen. Die Erkenntnisse der Interviewstudie werden genutzt, um im Rahmen des fünften Ziels Hypothesen zur Gestaltung lernwirksamer Lerngelegenheiten zum Einsatz digitaler Medien im Fach Physik abzuleiten.

Struktur der Arbeit

An dieser Stelle wird die Struktur der vorliegenden Arbeit vorgestellt. Die nachfolgenden Kapitel 2 bis 4 beschreiben die theoretischen Grundlagen und den relevanten Forschungsstand für diese Arbeit. In Kapitel 2 wird ausgehend vom übergeordneten FDW in Physik (2.1 und 2.2) das zentrale Konstrukt dieser Arbeit – das FDW zum Einsatz digitaler Medien im Fach Physik – unter Berücksichtigung angrenzender und ähnlicher Modellierungen in diesem Bereich (wie z. B. TPACK) dargestellt (2.3). Kapitel 3 thematisiert die Untersuchung von FDW und TPACK. Ausgehend von einer kurzen allgemeinen Darstellung quantitativer und qualitativer Zugänge in der fachdidaktischen Forschung (3.1), wird der Fokus auf die quantitative Kompetenzmessung gelegt. So erfolgt in 3.2 die Erläuterung von Verfahren der Kompetenzmessung in der Lehrkräftebildungsforschung sowie in 3.3 von Anforderungen hinsichtlich Qualitäts- und Validitätsaspekten bei Kompetenzmessungen. Anschließend werden verschiedene Messmodelle und -verfahren zur Messung von FDW (3.4) und TPACK (3.5) vorgestellt. Kapitel 4 zeigt den Forschungsstand in Bezug auf die Entwicklung von FDW im Allgemeinen (4.1) und von FDW zum Einsatz digitaler Medien bzw. TPACK (4.2) in der Lehrkräftebildung (im Fach Physik bzw. den Naturwissenschaften) auf.

Kapitel 5 beschreibt die fachdidaktischen Lerngelegenheiten an den beteiligten Standorten im Verbundprojekt DiKoLeP als Kontext der vorliegenden Arbeit. Anknüpfend an einen Überblick zum Verbundprojekt (5.1) werden das übergreifende Lehrkonzept (5.2.1) und die standortspezifischen Ausprägungen dargestellt (5.2.2 und 5.2.3). Die Formulierung der Ziele und Forschungsfragen dieser Arbeit erfolgt in Kapitel 6 auf Basis des beschriebenen Forschungsstandes und daraus abgeleiteten Desideraten. Anschließend wird in Kapitel 7 das methodische Vorgehen und Studiendesign der Arbeit skizziert. Darauf folgt eine Unterteilung in einen quantitativen und einen qualitativen Teil zur Vorstellung der Operationalisierungen und vorgenommenen Untersuchungen sowie der Ergebnisse dieser Arbeit.

Der quantitative Teil der Arbeit wird in den Kapiteln 8 bis 11 beschrieben. Zunächst wird die Testentwicklung (Kapitel 8) vorgestellt. Dazu zählen die Modellierung des FDW zum Einsatz digitaler Medien (8.1) sowie die systematische Aufgabenentwicklung anhand eines Itementwicklungsmodells (8.2).

Die Pilotierung des vorläufigen Testinstruments sowie anschließende Optimierungen werden in 8.3 erläutert. In 8.4 werden Ergebnisse zur Untersuchung der Testgüte der optimierten Testinstruments beschrieben. Kapitel 9 beinhaltet die einzelnen Teilstudien zur Untersuchung der Validität des entwickelten Testinstruments (9.1 bis 9.5) sowie eine Zusammenfassung dieser (9.6). In Kapitel 10 wird schließlich das entwickelte Testinstrument in seiner finalen Form inklusive einigen Aufgabenbeispielen (10.2) vorgestellt. Kapitel 11 zeigt die Ergebnisse zur Entwicklung des FDW zum Einsatz digitaler Medien über die untersuchten Seminare mithilfe des entwickelten Leistungstests.

Im qualitativen Teil der Arbeit werden in Kapitel 12 zunächst die Gestaltung und Auswertung der qualitativen retrospektiven Interviewstudie beschrieben, indem der Interviewleitfaden und zusätzliche Interviewmaterialien (12.2) sowie das entwickelte Kategoriensystem zur Analyse der Interviews dargestellt werden (12.3). Kapitel 13 beinhaltet die Ergebnisse der retrospektiven Interviewstudie in Bezug auf die Erkundung von Gründen für die Veränderungen im Testverhalten (13.1) und die Identifikation von lernförderlichen und lernhinderlichen Seminarelementen (13.2). Die entwickelten Hypothesen zur Gestaltung von Lehrveranstaltungen zum Einsatz digitaler Medien im Fach Physik werden in 13.3 vorgestellt.

Das abschließende Kapitel 14 beinhaltet das Fazit dieser Arbeit. Es werden die Ergebnisse der Arbeit zusammengefasst und diskutiert (14.1) sowie die Grenzen der vorliegenden Arbeit beleuchtet (14.2). Zum Abschluss wird ein Ausblick hinsichtlich möglicher Anknüpfungspunkte in Forschung und Lehrkräftebildung gegeben (14.3).

2 Fachdidaktisches Wissen zum Einsatz digitaler Medien im Fach Physik

Dieses Kapitel beschreibt den theoretischen Hintergrund zum fachdidaktischen Wissen (FDW) zum Einsatz digitaler Medien im Fach Physik beschrieben – jener Teil des Professionswissens von angehenden Physiklehrkräften, der in dieser Arbeit im Vordergrund steht. Zunächst wird das FDW im Modell der professionellen Handlungskompetenz von angehenden Physiklehrkräften verortet (2.1) und anschließend inhaltlich beschrieben (2.2). Schließlich wird der Fokus auf den Einsatz digitaler Medien im Unterricht gelegt und unter Berücksichtigung verschiedener Strukturierungen und Modelle in diesem Bereich das fokussierte Konstrukt dieser Arbeit – FDW zum Einsatz digitaler Medien im Fach Physik – definiert (2.3). In 2.4 folgt eine kurze Zusammenfassung der wesentlichen Inhalte des Kapitels.

2.1 Fachdidaktisches Wissen als Teil professioneller Handlungskompetenz

Die Lehrkräftebildungsforschung beschäftigt sich unter anderem mit der Überprüfung des Ausbildungserfolgs der universitären Lehrkräftebildung. Dabei stellt der Erwerb professioneller Kompetenz ein Ziel dieser ersten Phase der Lehrkräftebildung dar (z. B. Baumert & Kunter, 2006; Terhart, 2002). Um dies empirisch erfassen zu können, ist eine möglichst klare Definition des untersuchten Kompetenzbegriffs sowie eine passende Kompetenzmodellierung vonnöten (Riese, 2009). Als Basis wird dazu in der Lehrkräftebildungsforschung Weinerts (2001a) Konzept der *Handlungskompetenz* aufgegriffen.

Der Begriff Kompetenz erscheint aus dem Alltag bekannt, lässt sich jedoch in diesem Verständnis kaum von synonym verwendeten Begriffen wie „Können“ und „Fähigkeit“ abgrenzen und damit präzise definieren (Weinert, 2001a). Dennoch lässt sich in Bereichen wie Philosophie, Psychologie und Soziologie Kompetenz vergleichsweise einheitlich als ein grob spezialisiertes System von Fähigkeiten, Können und Kenntnissen verstehen, welche zum Erreichen eines speziellen Ziels notwendig oder hinreichend sind (Weinert,

2001a). Allgemein kann Kompetenz nach Weinerts Verständnis damit als kognitive Fähigkeit und Fertigkeit verstanden werden:

„Dabei versteht man unter Kompetenzen die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können.“ (Weinert, 2001b, S. 27–28)

Die Verknüpfung von kognitiven Fähigkeiten mit motivationalen und sozialen Voraussetzungen, welche zum erfolgreichen Lernen und Handeln vonnöten oder verfügbar sind, bezeichnet Weinert dabei als *Handlungskompetenz*, welche von allgemein kognitiven Kompetenzen (z. B. fachlichen und überfachlichen Kompetenzen) zu unterscheiden ist (Weinert, 2001a, 2001b). In der dargestellten Definition wird dabei Erlernbarkeit als eine Eigenschaft von Kompetenz deutlich, die für die Lehrkräftebildungsforschung eine wichtige Rolle spielt. Dass Kompetenz erlernbar ist, macht die Messung dieser z. B. zur Untersuchung der Wirksamkeit der Lehrkräftebildung überhaupt sinnvoll (Riese, 2009).

Weinerts (2001a, 2001b) Begriff der Handlungskompetenz stellt seitdem eine Grundlage für die Entwicklung von Kompetenzmodellen zur (fachspezifischen) Untersuchung der Lehrkräftebildung dar, welche dabei neben kognitiven Aspekten auch motivationale und volitionale Aspekte berücksichtigen. Populär und Ausgangspunkt einiger nationaler Studien ist das Modell der professionellen Handlungskompetenz von Baumert und Kunter (2006), welches im Rahmen eines Forschungsprogramms zur Untersuchung der Kompetenz von Mathematiklehrkräften (COACTIV: Kunter et al., 2011) entwickelt wurde. In diesem Modell wird für den kognitiven Bereich das Professionswissen als wesentlicher Teil der Handlungskompetenz einer Lehrkraft betrachtet. Für die weitere Spezifikation des Professionswissens werden unterschiedliche Wissensbereiche differenziert. Ausgehend von Shulmans (1986) Topologie zum professionellen Wissen von Lehrkräften (2.2) hat sich dabei die Unterscheidung in die drei Wissensbereiche pädagogisches Wissen, Fachwissen und FDW praktisch durchgesetzt (Baumert & Kunter, 2006).

Die einzelnen Wissensbereiche weisen nochmals eine Unterscheidung in verschiedene Facetten auf, welche für Fachwissen und FDW domänenspezifisch sind (Abbildung 2-1).

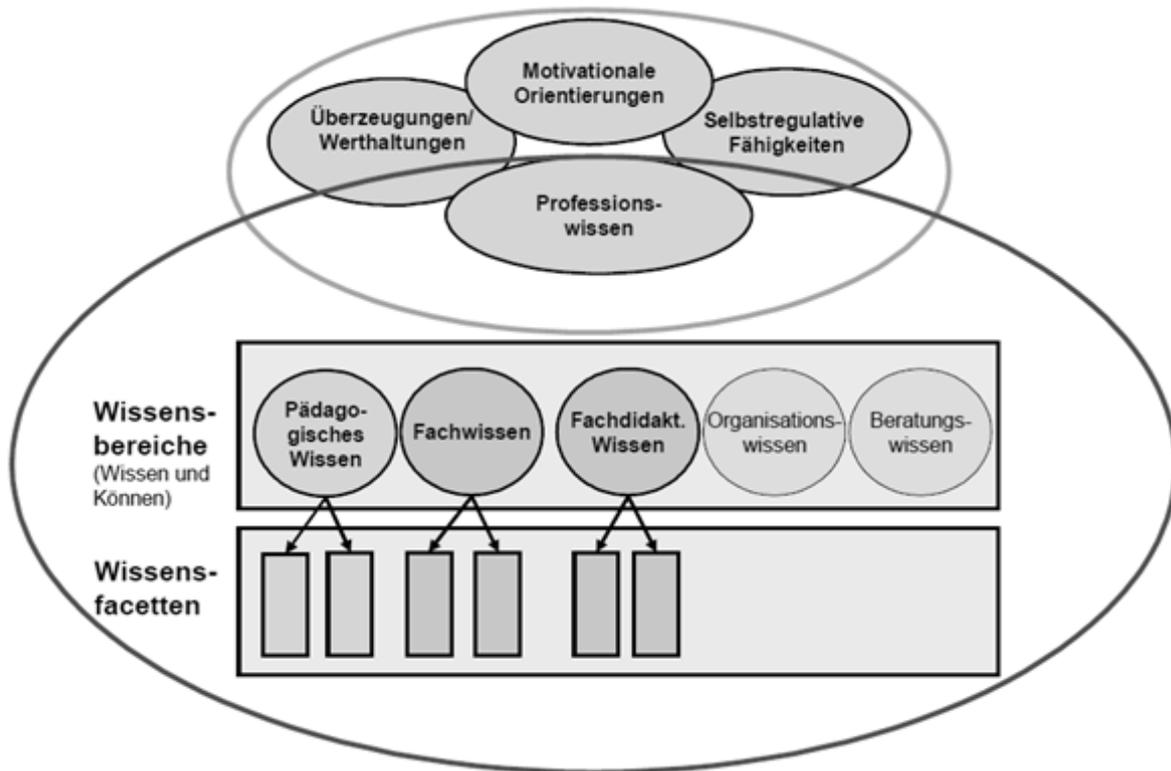


Abbildung 2-1: Modell professioneller Handlungskompetenz und innerer Struktur des Professionswissens nach Baumert und Kunter (2006, S. 482)

Daran orientiert entwickelte Riese (2009) für das Fach Physik ein entsprechendes Modell der professionellen Handlungskompetenz von (angehenden) Physiklehrkräften. Auch hier wird eine Aufteilung der professionellen Handlungskompetenz in einen kognitiven Kompetenzbereich mit dem Professionswissen von Lehrkräften sowie einen affektiven Bereich, der motivationale, volitionale und soziale Bereitschaften und Fähigkeiten enthält, vorgenommen. Ebenso wird für das Professionswissen die Dreiteilung in fachliches Wissen, FDW und pädagogisches Wissen aufgegriffen (Abbildung 2-2).



Abbildung 2-2: Kompetenzstrukturmodell zur professionellen Handlungskompetenz von (angehenden) Lehrkräften in Physik (Riese, 2009, S. 26; in Anlehnung an Baumert & Kunter, 2006; Blömeke et al., 2008)

Neben diesen dargestellten Kompetenzmodellen gibt es weitere, zum Teil unterschiedliche Auffassungen von Kompetenz und entsprechende Kompetenzmodellierungen. Um vorherrschende Differenzen im Kompetenzverständnis zu überwinden und Gemeinsamkeiten vermeintlich unterschiedlicher Kompetenzverständnisse hervorzuheben, schlagen Blömeke et al. (2015) ein Modell vor, in dem Kompetenz als Kontinuum von kognitiven und affektiven Voraussetzungen bis zum beobachtbaren Handeln verstanden werden kann. Dabei bildet das Professionswissen die kognitiven Voraussetzungen, welche neben den affektiven Voraussetzungen Dispositionen für Fähigkeiten und Fertigkeiten darstellen, die wiederum für erfolgreiches Handeln gebraucht werden (Blömeke et al., 2015).

Unabhängig davon, welche Kompetenzmodellierung konkret betrachtet wird, zeigt sich, dass das Professionswissen als eine wichtige kognitive Komponente der Kompetenz angenommen wird, die für erfolgreiches Lehrkräftehandeln von Bedeutung ist (Kaiser et al., 2020). Dies gilt ebenso für das FDW, welches neben Fachwissen und pädagogischen Wissen Teil des Professionswissens einer Lehrkraft bildet (z. B. Kaiser et al., 2020; Abbildung 2-1; Abbildung 2-2). In verschiedenen Studien konnte dieser angenommene Zusammenhang zwischen FDW und handlungsnahen Fähigkeiten (z. B. Blömeke et

al., 2022; Kulgemeyer & Riese, 2018; Riese et al., 2022b) oder erfolgreichem Lehrkräftehandeln (z. B. Baumert et al., 2010; Keller et al., 2017) empirisch gestützt werden (4.1).

Während das in dieser Arbeit fokussierte FDW in den nachfolgenden Kapiteln genauer betrachtet wird, soll auf die anderen beiden Professionswissensbereiche hier nur kurz eingegangen werden. Fachwissen stellt das Wissen einer Lehrkraft über spezifische Inhalte des zu unterrichtenden Fachs im Sinne eines profunden fachlichen Hintergrundverständnisses dar (z. B. Baumert & Kunter, 2006). Für das Fachwissen in Physik unterscheidet beispielsweise Riese (2009) die Bereiche oder Niveaustufen Schulwissen, vertieftes Wissen und universitäres Wissen. Pädagogisches Wissen wird z. T. mehr oder weniger unterrichtsnah beschrieben und erfasst (Kaiser et al., 2020). So fokussieren Blömeke und König (2010) das Unterrichten als Kernaufgabe von Lehrkräften und operationalisieren den fachübergreifenden Wissensbereich des Professionswissens unterrichtsnah über die Bereiche Strukturierung von Unterricht, Klassenführung, Motivierung, Umgang mit Heterogenität und Leistungsbeurteilung. Weniger unterrichtsnah, sondern unter Berücksichtigung weiterer beruflicher Anforderungen außerhalb des Unterrichts, differenzieren Kunter et al. (2017) in ihrer Modellierung des fachunabhängigen Wissensbereichs die sechs Dimensionen Unterrichtsgestaltung, Schulorganisation, Bildungstheorie, Lernen und Entwicklung, Diagnostik und Evaluation, Lehrerberuf als Profession. Die drei Professionswissensbereiche Fachwissen, pädagogisches Wissen und FDW finden sich auch in der universitären Phase der deutschen Lehrkräftebildung als wesentliche Studienbereiche wieder (KMK, 2019; 2022). Weiterhin ist die Struktur des Professionswissens mit seinen drei Wissensbereichen für das Fach Physik bereits empirisch untersucht (z. B. Riese et al., 2017).

Das in dieser Arbeit fokussierte FDW in Physik stellt demnach einen relevanten Teil des Professionswissens und damit der professionellen Handlungskompetenz von (angehenden) Physiklehrkräften dar. Die Untersuchung des FDW bei Lehramtsstudierenden leistet dabei einen Beitrag zur Überprüfung der Wirksamkeit der universitären Lehrkräftebildung im Hinblick auf den Erwerb professioneller Handlungskompetenz. Im folgenden Abschnitt wird da-

her das FDW sowie seine innere Struktur für das Fach Physik bzw. die Naturwissenschaften inhaltlich beschrieben (2.2). Anschließend wird der fokussierte Wissensbereich des FDW zum Einsatz digitaler Medien im Fach Physik in dieser Arbeit definiert und erläutert (2.3).

2.2 Fachdidaktisches Wissen von (angehenden) Lehrkräften in den Naturwissenschaften

Wie im vorausgehenden Abschnitt dargestellt, wird das FDW im Modell der professionellen Handlungskompetenz neben Fachwissen und pädagogischem Wissen als ein Wissensbereich des Professionswissens von Lehrkräften aufgefasst. Diese Topologie geht auf Shulman (1986) zurück, der der Frage nachging, inwiefern allgemein pädagogisches Wissen und Fachwissen miteinander in Relation stehen und wodurch solches Wissen erlernt werden könne. Damit führte Shulman das *pedagogical content knowledge* (PCK) als jenes fachspezifische Wissen ein, das eine Lehrkraft zum Unterrichten und Vermitteln fachlicher Inhalte benötigt. Es beinhaltet Wissen über hilfreiche Repräsentationen, Darstellungen, Beispiele zur verständlichen Vermittlung der jeweiligen Fachinhalte und über mögliche Lernschwierigkeiten und Präkonzepte, die Lernende in den Fachunterricht mitbringen (Shulman, 1986). PCK unterscheidet Fachlehrkräfte damit von nicht unterrichtenden Expert:innen dieses Fachs und von Lehrkräften, die das Fach nicht unterrichten (Marks, 1990).

2.2.1 Fachdidaktisches Wissen und *pedagogical content knowledge*

FDW wird häufig als deutsche Übersetzung des PCK angegeben, jedoch sind die beiden Begriffe aufgrund unterschiedlicher zugrundeliegender Theorien nicht gleichbedeutend. Das FDW hat seinen Ursprung in der deutschen Didaktik- und Bildungstradition, das PCK hingegen in der anglo-amerikanischen Bildungstradition (Gramzow, 2015). Die Grundideen beider Konzepte sind jedoch vergleichbar, weswegen die Begriffe meist synonym verwendet werden (Gramzow, 2015). Diese Arbeit ist am Konzept des FDW als Teil der professionellen Handlungskompetenz (2.1) orientiert, weswegen im Laufe der Arbeit in der Regel auch der Begriff des *fachdidaktischen Wissens* (FDW) verwendet wird. Nichtsdestotrotz wird auch auf Literaturbezüge zum PCK

zurückgegriffen, wobei zu beachten ist, dass die Vergleichbarkeit von jeweiligen Forschungsbeiträgen aufgrund der unterschiedlichen Genese und Modellierungen beider Konzepte nicht immer uneingeschränkt gegeben ist.

In diesem Zusammenhang sei auf Unterschiede in der Beschreibung oder Modellierung von PCK hingewiesen, welche sich nach Gess-Newsome (1999) zwischen zwei Extremen bewegen: vom integrativen Modell zum transformativen Modell. Im integrativen Modell besteht das Wissen einer Lehrkraft aus separaten Konstrukten oder Wissensarten Fachwissen, pädagogisches Wissen und Wissen über den Kontext, welche beim Unterrichten (zu PCK) integriert werden. Im transformativen Modell stellt PCK die Synthese aller notwendigen Konstrukte oder Wissensbereiche dar, um effektiv zu unterrichten. PCK ist dabei eine eigene Wissensform, die von Fachwissen, pädagogischen Wissen und Wissen über den Kontext beeinflusst ist und für praktisches Unterrichten von Bedeutung ist (Gess-Newsome, 1999). Im deutschsprachigen Raum entspricht das FDW im Modell der professionellen Handlungskompetenz nach Baumert und Kunter (2006) bzw. Riese (2009) eher dem Verständnis im transformativen Modell, wenngleich diese Einordnung aufgrund der unterschiedlichen Konzeptionen des für den Unterricht relevanten Wissens in beiden Sprachbereichen nicht eindeutig ist (Gramzow, 2015).

Um unterschiedlichen Konzeptualisierungen des PCK zu begegnen, gab es in der jüngeren Vergangenheit Bestrebungen, einen internationalen Konsens in der Beschreibung von PCK zu finden. Mit diesem Ziel entwickelte eine Gruppe von Wissenschaftler:innen in gemeinsamen Austauschtreffen (dem sogenannten *PCK Summit*) zunächst das *Consensus Model* (Gess-Newsome, 2015) und später als Weiterentwicklung das *Refined Consensus Model* (Hume et al., 2019). Letzteres wird aufgrund seiner Aktualität und Domänenspezifität für Naturwissenschaften hier kurz erläutert. Im *Refined Consensus Model of PCK in Science Education* (RCM) wird zwischen drei voneinander abhängigen Bereichen des PCK unterschieden: *collective PCK*, *personal PCK* und *enacted PCK* (Carlson et al., 2019b; Abbildung 2-3).

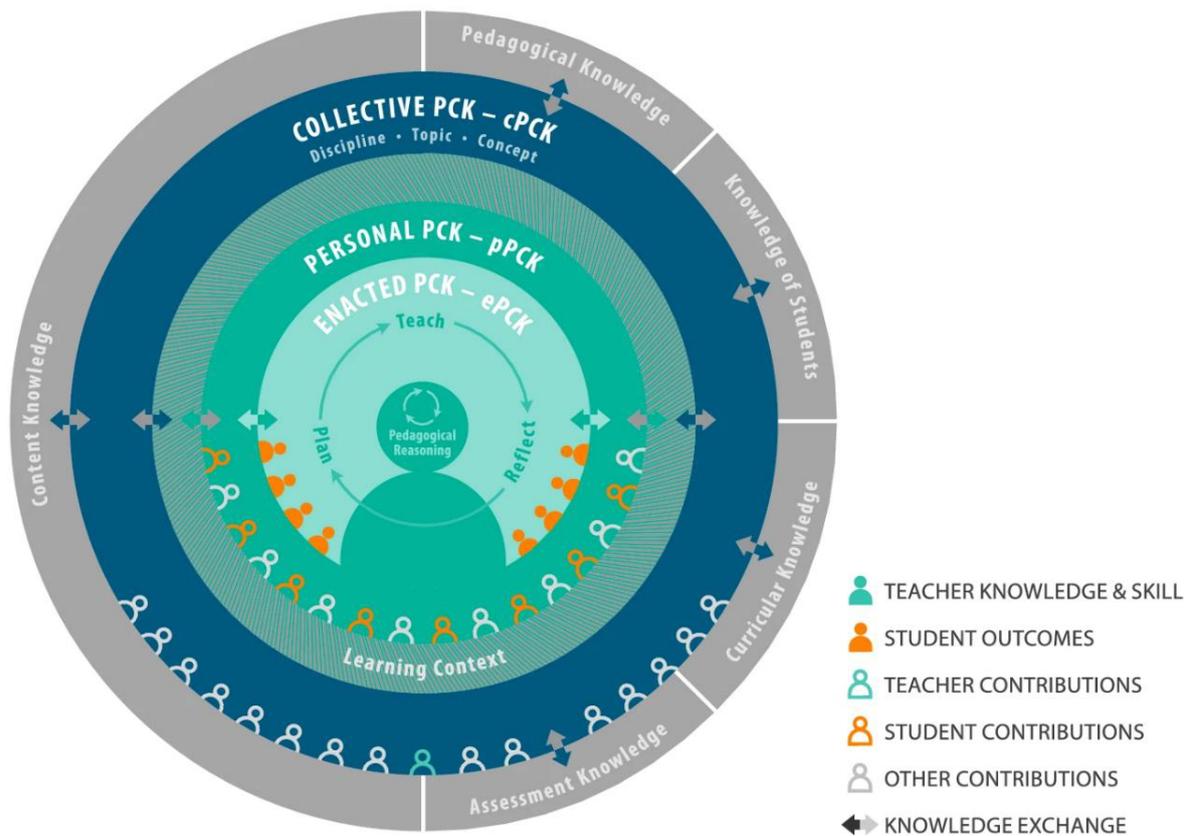


Abbildung 2-3: Darstellung des Refined Consensus Model (RCM) of PCK (Carlson et al., 2019b, S. 84)

Collective PCK stellt dabei unter Lehrkräften und Wissenschaftler:innen gemeinsam geteiltes PCK bzw. Wissen im Zusammenhang mit naturwissenschaftlichen Unterricht dar. Dieser Wissensbereich ist oft literatur- oder evidenzbasiert, meint aber auch lokaleres Wissen, das beispielsweise von Fachkolleg:innen einer Schule geteilt wird. Personal PCK entspricht hingegen dem individuellen PCK bzw. Wissenstand einer Lehrkraft und beinhaltet damit auch persönliche Erfahrungen zum Lehren im jeweiligen Fach. Enacted PCK ist das individuelle Wissen einer Lehrkraft, das sie beim unterrichtsbezogenen Handeln nutzt – wobei dazu neben dem eigentlichen Unterrichten auch die Planung von Unterricht und die Reflexion über Unterricht und Lernenergebnisse zählen (Carlson et al., 2019b). Da enacted PCK das Handeln einer Lehrkraft in den Blick nimmt, lässt sich dieser Wissensbereich nicht in angemessener Weise über schriftliche Wissens- oder Leistungstests messen. Solche Testinstrumente bieten sich jedoch für die Messung von personal PCK oder collective PCK an, wie es in der Forschung bereits häufig gemacht

wird (z. B. Carlson et al., 2019a; Kulgemeyer et al., 2020; Sorge et al., 2019). Es kann hingegen auch argumentiert werden, dass nur collective PCK mit Wissens- oder Leistungstests gemessen werden kann, da bei solchen Tests eine Bewertung anhand einer Musterlösung erfolgt, wodurch immer eine Form von geteiltem Expert:innenwissen untersucht wird (Carlson et al., 2019a).

Im RCM werden auch die Wissensbereiche Fachwissen (*content knowledge*, CK) und pädagogisches Wissen (*pedagogical knowledge*, PK) als Teile einer breiteren Basis des professionellen Wissens einer Lehrkraft aufgegriffen (Carlson et al., 2019b). Die Bereiche werden dabei als essenzielle Grundlagen bezeichnet, ohne die eine angemessene Ausprägung von PCK nicht möglich ist. In diesem Verständnis ist das PCK den Wissensbereichen CK und PK übergeordnet und umfasst mit enacted PCK bereits explizit die unterrichtsbezogene Praxis (Carlson et al., 2019b). In der deutschsprachigen Tradition stellt das FDW hingegen einen eigenständigen Wissensbereich neben Fachwissen und pädagogischen Wissen dar, wobei das Zusammenspiel dieser drei Bereiche für erfolgreiches Lehrkräftehandeln notwendig ist (Baumert & Kunter, 2006; Riese, 2009). Demnach wird deutlich, dass die Konzepte zum FDW und PCK nicht unbedingt in allen Aspekten vergleichbar sind, wenngleich oberflächlich betrachtet eine große Ähnlichkeit besteht.

Diese Arbeit bezieht sich aufgrund der Verortung in der deutschsprachigen Lehrkräftebildungsforschung und der Nutzung fachdidaktischer Vorarbeiten zur Messung des FDW auf das Konzept zum FDW im Sinne der professionellen Handlungskompetenz (2.1). Es wird demnach jenes universitär erwerb-bare FDW betrachtet, das im Zusammenspiel mit den anderen Bereichen des Professionswissens sowie affektiven und motivationalen Bereitschaften zur professionellen Handlungskompetenz von Lehrkräften beiträgt. Auch wenn das zuvor beschriebene RCM hier nicht als primäres Bezugsmodell verwendet wird, sei dennoch auf den Vorteil der internationalen Vergleichbarkeit, die durch dieses Modell ermöglicht wird, hingewiesen. Weiterhin sind auch im RCM verschiedene Auffassungen zu PCK und FDW berücksichtigt, welche teilweise auch in dem in dieser Arbeit verwendeten Bezugsmodell zum FDW (2.2.2) einbezogen wurden. Bei einer Verortung des in dieser Arbeit adres-

sierten FDW im RCM entspräche das Wissen, welches im Rahmen der Lehrveranstaltungen in der vorliegenden Studie untersucht wird, dem personal PCK der Lehramtsstudierenden, da kein „Lehrbuchwissen“ im Sinne des collective PCK, sondern der individuell vorhandene Wissensstand der Studierenden erfragt wird. Dieses personal PCK dient wiederum zur Vorbereitung auf erfolgreiches Unterrichtshandeln (enacted PCK).

Nachdem nun Differenzen zwischen den verwandten Konzepten FDW und PCK aufgezeigt wurden, soll im Folgenden auf die innere Struktur des FDW eingegangen und schließlich das in dieser Arbeit verwendete Bezugsmodell des FDW im Fach Physik vorgestellt werden. Anschließend erfolgt in 2.3 die Beschreibung der in dieser Arbeit fokussierten Facette des FDW zum Einsatz digitaler Medien.

2.2.2 Innere Struktur des Fachdidaktischen Wissens

Zur Beschreibung der inneren Struktur des FDW sind in der naturwissenschaftsdidaktischen Forschung bereits einige Modellierungen entwickelt worden, z. T. auch als Ausgangslange für die Operationalisierung des FDW in Form von Testinstrumenten (3.4). In solchen Modellen wird die innere Struktur des FDW meist in verschiedenen Dimensionen beschrieben (z. B. dreidimensional in Kröger et al., 2013; Tepner et al., 2012; zweidimensional in Gramzow et al., 2013). Gemeinsamkeit dieser Modelle ist eine Dimension zur Differenzierung von verschiedenen fachdidaktischen Inhalten, Facetten oder Komponenten, welche das Wissen in verschiedenen fachdidaktischen Aspekten beschreibt, aus denen sich das FDW einer Lehrkraft zusammensetzt. In der Literatur wird dies unterschiedlich bezeichnet; im Folgenden wird für eine derartige Differenzierung der Begriff *Facetten* des FDW genutzt. Während einzelne Modelle z. T. unterschiedliche Facetten beinhalten, treten die zwei Facetten Instruktionsstrategien und Schülerverständnis bzw. Schülervorstellungen fast immer auf – sowohl in Modellierungen zum FDW nach deutschsprachiger Tradition (z. B. Baumert & Kunter, 2006; Dollny, 2011; Gramzow, 2015; Kirschner, 2013) als auch zum PCK (z. B. Chan & Hume, 2019; Magnusson et al., 1999; Park & Oliver, 2008). Diese beiden Facetten stellen wiederum auch die zentralen Inhalte in der Beschreibung des PCK nach Shulman (1986) dar. In den naturwissenschaftlichen Fächern wird in

manchen Modellierungen Wissen zu Experimenten als eigene Facette (z. B. Gramzow et al., 2013) oder innerhalb der Facette Instruktionsstrategien (z. B. Kirschner et al., 2016) einbezogen.

Im Folgenden wird das Modell zum FDW in Physik nach Gramzow et al. (2013) vorgestellt, welches als Synthese bestehender Modellierungen zum FDW und PCK (z. B. Magnusson et al., 1999; Olszewski, 2010) sowie unter Nutzung von Literatur zur fachdidaktischen Ausbildung (z. B. Korneck et al., 2010) und Lehrbüchern zur Physikdidaktik (z. B. Kircher et al., 2009; Mikelskis, 2006) entwickelt wurde. Das Modell diente der differenzierten Beschreibung der inneren Struktur des FDW in Physik und fokussierte auf Wissen, das für Lehramtsstudierende der Physik an der Universität erwerbbar ist (Gramzow, 2015). Die Entwicklung des Modells erfolgte im Rahmen des Verbundprojekts ProfiLe-P („Professionswissen in der Lehramtsausbildung Physik“: Kulgemeyer et al., 2012; Riese et al., 2015). Das entwickelte Modell ist in Abbildung 2-4 dargestellt. Die innere Struktur des FDW in Physik wird in zwei Dimensionen beschrieben. Die erste Dimension unterscheidet fachliche Inhaltsbereiche, welche im Modell noch nicht näher spezifiziert werden. Die zweite Dimension unterscheidet das FDW in acht fachdidaktischen Facetten. Dabei treten erneut die beiden zentralen Facetten Instruktionsstrategien sowie Schülervorstellungen auf; ebenso wird eine Facette zu Experimenten im Modell berücksichtigt (Abbildung 2-4).

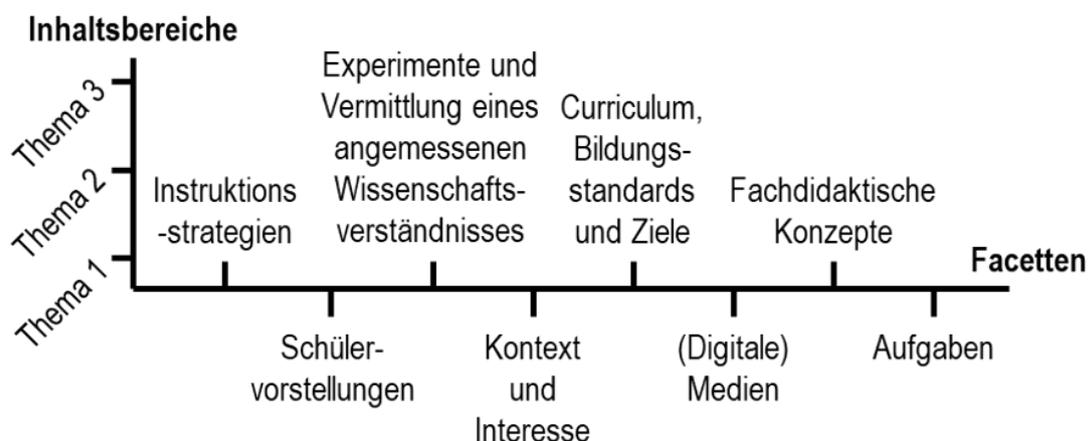


Abbildung 2-4: Modell zum Fachdidaktischen Wissen in Physik nach Gramzow et al. (2013) bzw. Gramzow (2015)

Weiterhin enthält das Modell die Facette (*Digitale Medien*) als Teil des FDW in Physik. Diese Facette wird aufgrund der zunehmenden Relevanz digitaler Medien für den Fachunterricht (Kapitel 1) auch für das FDW bedeutender. Im Projekt ProfiLe-P wurde jedoch zur Messung des FDW bei Physiklehramtsstudierenden eine Reduktion des Modells für die Operationalisierung vorgenommen, welche vier andere Facetten des Modells fokussiert (Riese et al., 2017; 3.4). Demnach liegt bislang keine differenzierte Beschreibung oder Operationalisierung der Facette (Digitale) Medien vor. Diese Facette steht als FDW zum Einsatz digitaler Medien jedoch im Fokus der vorliegenden Arbeit. Im nachfolgenden Kapitel wird daher ausgehend von der Definition im FDW-Modell nach Gramzow et al. (2013) eine differenzierte Beschreibung dieser Facette bzw. des FDW zum Einsatz digitaler Medien vorgenommen.

2.3 Fachdidaktisches Wissen zum Einsatz digitaler Medien

Die Facette (Digitale) Medien des FDW in Physik umfasst das Wissen über Möglichkeiten und Anforderungen beim Einsatz digitaler Medien sowie über die inhaltsspezifisch angemessene Nutzung digitaler Medien im Physikunterricht. Die Facette beinhaltet weiterhin Wissen zum Computereinsatz zur Messwerterfassung und -auswertung sowie zum Einsatz von Simulationen (Korneck et al., 2010, zitiert nach Gramzow et al., 2013). Diese Definition aus dem zugrundeliegenden FDW-Modell bietet sich als Ausgangslage an, die für diese Arbeit jedoch aufgrund von Entwicklungen durch die Digitalisierung aktualisiert und zur vertieften Beschreibung des Wissensbereichs weiter ausgeschärft wird. Im Folgenden werden daher weitere Konzeptualisierungen im Bereich digitalisierungsbezogenen Wissens- oder Kompetenzbereichen betrachtet, um die Beschreibung des Fachdidaktischen Wissens zum Einsatz digitaler Medien zu schärfen.

2.3.1 Beschreibungen zu digitalisierungsbezogenen Kompetenzen

Andere Konzeptualisierungen zu digitalisierungs- oder medienbezogenem Wissen oder digitalen Kompetenzen¹ (angehender) Lehrkräfte in diesem Zusammenhang sind beispielweise normative Rahmenmodelle (z. B. *DigCompEdu*: Redecker, 2017) oder nationale Anforderungen des Kultusministeriums (KMK, 2016; 2019). Die dort genannten Vorschläge zur Beschreibung digitalisierungsbezogener Kompetenzen sind jedoch sehr allgemein formuliert. Dadurch sind sie für die vorliegende Arbeit mit dem Ziel einer differenzierten Beschreibung des interessierenden Bereichs des FDW speziell für das Fach Physik kaum nutzbar.

Weiterhin ist das TPACK-Modell (Koehler et al., 2013; Mishra & Koehler, 2006) zur Beschreibung des technologiebezogenen Professionswissens von Lehrkräften international verbreitet. In diesem Modell wird Shulmans (1986) Topologie des Professionswissens einer Lehrkraft aus den Wissenskomponenten CK und PK und deren Kombination PCK bzw. FDW um eine weitere Wissenskomponente des *technological knowledge* bzw. technologiebezogenen Wissens (TK) erweitert (Abbildung 2-5). Damit entstehen durch Schnittmengen der bisherigen Wissenskomponenten CK, PK und PCK mit dem neuen TK weitere Wissensbereiche: *technological content knowledge* bzw. technologiebezogenes Fachwissen (TCK), *technological pedagogical knowledge* bzw. technologiebezogenes pädagogisches Wissen (TPK) und das zentrale *technological pedagogical content knowledge* bzw. technologiebezogene FDW (TPACK). Dieser zentrale Wissensbereich TPACK beinhaltet jenes Wissen, das eine Lehrkraft dazu befähigt, digitale Technologien oder Medien im Unterricht einzusetzen, um Fachinhalte konstruktiv zu vermitteln und das Lernen der Schülerinnen und Schüler zu unterstützen (Mishra & Koehler, 2006). TPACK ist demnach vergleichbar mit jenem Wissen, das die Facette (Digitale) Medien des FDW beschreibt.

¹ Es ist zu beachten, dass in diesem Kontext häufig von „digitalen Kompetenzen“ (und nicht von Wissen) von Lehrkräften gesprochen wird. In dieser Arbeit liegt der Fokus auf dem FDW als eine der kognitiven Wissenskomponenten der professionellen Handlungskompetenz. Dabei sind auch Kompetenzbeschreibungen zum Einsatz digitaler Medien von Interesse, wenngleich dort nicht zwingend dasselbe Verständnis zum Kompetenzbegriff vorliegen mag.

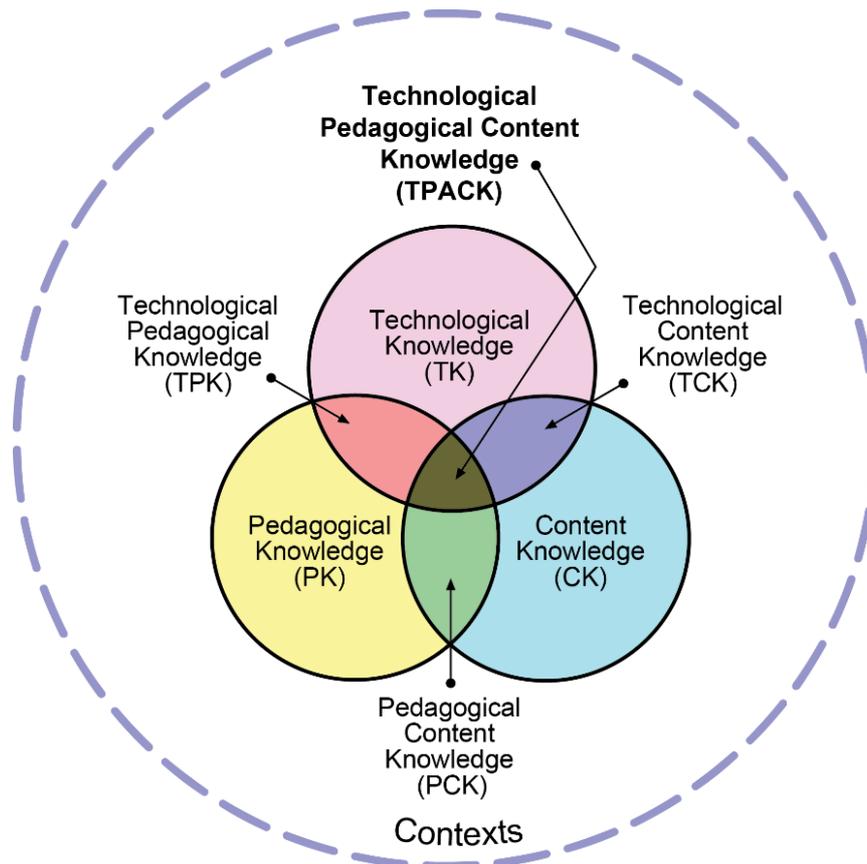


Abbildung 2-5: TPACK-Modell (Koehler & Mishra, 2009, S. 63; Mishra & Koehler, 2006)

Ähnlich wie bei den normativen Beschreibungen zu digitalen Kompetenzen fehlt es dem TPACK-Modell jedoch auch an Fachspezifität, da es übergreifende Kompetenzanforderungen oder Wissens Elemente beschreibt (Becker et al., 2020). Im TPACK-Modell wird das Fachwissen bzw. CK zwar explizit berücksichtigt, aber die Beschreibungen zu den einzelnen Wissensbereichen sind dennoch nicht für ein bestimmtes Fach spezifiziert und enthalten höchstens übergreifende Formulierungen. Für eine erfolgreiche Implementation digitaler Medien in den eigenen Fachunterricht scheint jedoch speziell die Inhalts- oder Fachspezifität von Bedeutung, wenn vor allem auch das fachliche Lernen mit dem Einsatz digitaler Medien unterstützt werden soll.

Ausgehend von diesem Mangel an Fachspezifität in Konzeptualisierungen digitaler Kompetenzen hat sich eine Arbeitsgruppe deutschsprachiger Didaktiker:innen in den naturwissenschaftlichen Fächern gebildet und den Orientierungsrahmen DiKoLAN zur Beschreibung digitaler Kompetenzen für das

Lehramt in den Naturwissenschaften entwickelt (Becker et al., 2020). In diesem Orientierungsrahmen werden digitale Basiskompetenzen in sieben Kompetenzbereichen sowie grundlegende rechtliche Rahmenbedingungen und technische Basiskompetenzen beschrieben. Vier der sieben Kompetenzbereiche beinhalten allgemeinere Kompetenzen (*Dokumentation, Präsentation, Kommunikation und Kollaboration* sowie *Recherche und Evaluation*). Die drei weiteren Kompetenzbereiche sind für die naturwissenschaftlichen Fächer eher fachspezifischere Bereiche (*Messwert- und Datenerfassung, Datenverarbeitung* sowie *Simulation und Modellierung*; Abbildung 2-6).



Abbildung 2-6: Struktur des DiKoLAN-Orientierungsrahmens für digitale Basiskompetenzen (<https://dikolan.de/>, letzter Zugriff am 20.02.2024)

Im Orientierungsrahmen werden für jeden der sieben Kompetenzbereiche in tabellarischer Form Kompetenzerwartungen beschrieben, die anhand drei Kompetenzniveaus (*Nennen, Beschreiben* und *Anwenden/Durchführung*) sowie nach vier Schwerpunkten differenziert werden. Die vier Schwerpunkte sind orientiert an den technologiebezogenen Wissenskomponenten des TPACK-Modells (Abbildung 2-5): *spezielle Technik* (Wissenskomponente TK), *fachwissenschaftlicher Kontext* (TCK), *Methodik und Digitalität* (TPK) sowie *Unterricht* (TPACK). Zur fachspezifischen Beschreibung des FDW zum Einsatz digitaler Medien bzw. der Facette (Digitale) Medien des FDW in Physik

sind demnach vor allem die Kompetenzerwartungen aus den drei fachspezifischeren Kompetenzbereichen (*Messwert- und Datenerfassung, Datenverarbeitung sowie Simulation und Modellierung*) mit dem Schwerpunkt *Unterricht* (TPACK) des DiKoLAN-Orientierungsrahmens von Bedeutung.

Die Fachspezifität ist zur Beschreibung des FDW zum Einsatz digitaler Medien im Fach Physik also besonders relevant, da aus fachdidaktischer Perspektive ein Ziel des Medieneinsatzes die Vermittlung zwischen fachlichen Inhalten, Lehrenden und Lernenden darstellt. Es sind aus dieser fachdidaktischen Sicht daher vor allem fachspezifische und fachtypische Medien von Relevanz, deren Gestaltung an fachlichen Lernzielen, Denk- und Arbeitsweisen des Fachs sowie der Veranschaulichung theoretischer Modelle oder Sachverhalte im Fach orientiert sind (Ropohl et al., 2018). Im Folgenden werden exemplarisch Medien oder Medieneinsätze beschrieben, die für das Fach Physik als fachtypisch oder -spezifisch zu nennen sind, wobei an einigen Stellen auf die fachspezifischen Kompetenzbeschreibungen im DiKoLAN-Orientierungsrahmen (Becker et al., 2020) Bezug genommen wird.

2.3.2 Fachtypische Einsatzmöglichkeiten digitaler Medien im Physikunterricht

So spielen Systeme zur **digitalen Messwerterfassung** bei physikalischen Experimenten eine wichtige Rolle für den Physikunterricht (z. B. Becker et al., 2020; Girwidz, 2020a, 2020c; Lampe et al., 2015). Mit Lehrmittelsystemen und Sensoren können Messdaten physikalischer Größen aufgenommen werden. Darstellung, Aufbereitung und Analyse der Messdaten können z. T. durch die Systeme selbst oder mittels Computereinsatzes erfolgen. Zunehmende Möglichkeiten bieten weiterhin mobile Systeme mit integrierten Sensoren z. B. in Smartphones und Tablets (z. B. Kuhn et al., 2015). Digitale Messwerterfassung und -analyse erweitert damit die experimentellen Möglichkeiten traditioneller Verfahren, wenn z. B. physikalische Größen erst durch die Nutzung bestimmter Sensoren messbar werden oder wenn durch die zeiteffiziente Aufnahme von Messdaten z. B. das Aufnehmen größerer Datenreihen oder Langzeitmessungen ermöglicht wird. Dies schafft außerdem Raum für die Fokussierung auf die Dateninterpretation und die physikalischen Zusammenhänge (z. B. Girwidz, 2020a; Hoyer & Girwidz, 2019; Lampe

et al., 2015). Weiterhin stellt Videoanalyse ein typisches Verfahren zur Analyse von Bewegungen im Physikunterricht dar und bietet sich vor allem in der Mechanik an (Wilhelm, 2020). Ebenso kann eine Videoanalyse am Computer oder an mobilen Endgeräten erfolgen (Becker et al., 2019; Wilhelm, 2020).

Simulationen stellen eine fachspezifische Arbeitsweise zum Erkenntnisgewinn dar, da in diesen virtuellen Experimenten z. B. physikalische Zusammenhänge und Hypothesen durch gezielte und einfache Variablenkontrolle untersucht werden können (Becker et al., 2020; Blake & Scanlon, 2007; Girwidz, 2020c). Auch Simulationen erweitern die experimentellen Möglichkeiten, wenn Realversuche im Klassenraum nicht praktikabel oder unmöglich sind. Sie können außerdem zum Testen physikalischer Vorstellungen oder zur Übung von experimentellen Arbeiten genutzt werden (Blake & Scanlon, 2007; Rutten et al., 2012; Wieman et al., 2010). Durch die Ortsunabhängigkeit ermöglichen Simulationen weiterhin ein virtuelles Experimentieren von zuhause. Auch interaktive Bildschirmexperimente bieten eine ortsunabhängige und ökonomische Möglichkeit zum Experimentieren. Sie beruhen auf Aufnahmen und Messdaten von Realexperimenten und weisen damit eine besondere Authentizität auf (Kirstein & Nordmeier, 2013). Dynamische Animationen sowie Verfahren im Bereich erweiterter Realität (*Augmented Reality*) erlauben die Visualisierung nicht sichtbarer oder nicht unmittelbar erfahrbare Sachverhalte (Girwidz, 2013). In Augmented-Reality-Experimenten können Realaufbauten gezielt durch virtuelle Darstellungen erweitert, um z. B. unsichtbare Konzepte wie Feldlinien, Ladungen oder Kraftpfeile zu visualisieren und mit dem Realaufbau zu verknüpfen (Teichrow & Erb, 2020). Dynamische Animationen können im Physikunterricht weiterhin z. B. zur Darstellung zeitlicher Abläufe oder abstrakter Konzepte genutzt werden, wodurch das konzeptionelle Verständnis der Lernenden unterstützt werden kann (Girwidz, 2013).

Erklärvideos bieten die Möglichkeit zum Erklären physikalischer Sachverhalte und können im Physikunterricht auf verschiedene Weise eingesetzt werden – variabel mit Lehrkräften oder Schülerinnen und Schülern als Produzierende oder Rezipierende (Wolf & Kulgemeyer, 2016). Die Videos enthalten dabei z. B. grafische Darstellungsformen, dynamische Animationen oder

auch Ausschnitte von Experimenten, um die gesprochenen Erklärungen visuell zu unterstützen (Kulgemeyer, 2018b). Zur Erstellung eines lernförderlichen Erklärvideos oder der Bewertung existierender Erklärvideos können verschiedene Qualitätskriterien zur Erklärqualität (z. B. minimales Erklären und Vermeidung von Exkursen, Verdeutlichung der Relevanz des Erklärten oder anschließende Lernaufgaben) berücksichtigt werden, um erfolgreiches Lernen mit den Videos zu fördern (Kulgemeyer, 2018a, 2018b). Erklärvideos sind im Vergleich zu den zuvor genannten Medien weniger fachspezifisch. Nichtsdestotrotz werden sie in der vorliegenden Arbeit als bedeutsames Medium für das Fach Physik angesehen, da sie sich für das Erklären komplexer Inhalte anbieten (Kulgemeyer, 2018a) – von denen es in der Physik einige gibt – und von Schülerinnen und Schülern auch vielfach selbstständig im Internet rezipiert werden (Wolf, 2018). Zudem gibt es bereits einige physikdidaktische Studien im Bereich der Erklärvideos (z. B. Finkenbergh, 2018; Kulgemeyer, 2018a, 2018b; Sterzing et al., 2019).

2.3.3 Bedingungen für einen Mehrwert des Medieneinsatzes

Beim Einsatz dieser vorgestellten digitalen Medien im Unterricht spielen zudem zwei grundlegende lernpsychologische Theorien eine wichtige Rolle: zum einen die *Cognitive Load Theory* (CLT; Sweller, 1994), welche die Begrenztheit des Arbeitsgedächtnisses und Folgen für das Lernen adressiert, und zum anderen die *Cognitive Theory of Multimedia Learning* (CMTL; Mayer, 2009), welche die aktive Verarbeitung von vielschichtigen Informationen auf zwei Kanälen (sprachbasiert und bildbasiert) beschreibt. Aus diesen Theorien ergeben sich Designprinzipien für multimediale Anwendungen oder Lernumgebungen (z. B. Mayer, 2009; Mayer & Moreno, 2003), die zu berücksichtigende Gesichtspunkte für einen lernförderlichen Einsatz des jeweiligen Mediums darstellen können. Diese fachunabhängigen Theorien sind daher auch für spezifische Situationen im Physikunterricht von Bedeutung, bei denen beispielsweise eines der in 2.3.2 genannten physiktypischen Medien eingesetzt wird.

FDW zum Einsatz digitaler Medien beinhaltet demnach Wissen über die inhaltsspezifische Nutzung sowie über Möglichkeiten und Anforderung solcher für den Physikunterricht spezifischen oder typischen digitalen Medien.

Dabei ist weiterhin bedeutsam, inwiefern die spezifischen Einsatzmöglichkeiten der digitalen Medien im Physikunterricht einen Mehrwert für das fachliche Lernen darstellen. In einer Metastudie zur Lernwirksamkeit beim Medieneinsatz im mathematisch-naturwissenschaftlichen Unterricht zeigte sich beispielsweise, dass digitale Medien vor allem in Kombination mit traditionellen Medien – und nicht als Ersatz – lernwirksam sind. Dabei hat insbesondere der kurzfristige Einsatz, die Zusammenarbeit mit Mitschülerinnen und Mitschülern sowie die Medienkompetenz der Lehrkräfte einen positiven Einfluss auf die Lernwirksamkeit (Hillmayr et al., 2020).

Aus inhaltlicher Perspektive lässt sich der Mehrwert eines digitalen Medieneinsatzes beispielsweise mit dem SAMR-Modell nach Puentedura (2006, 2012) beschreiben (Abbildung 2-7).

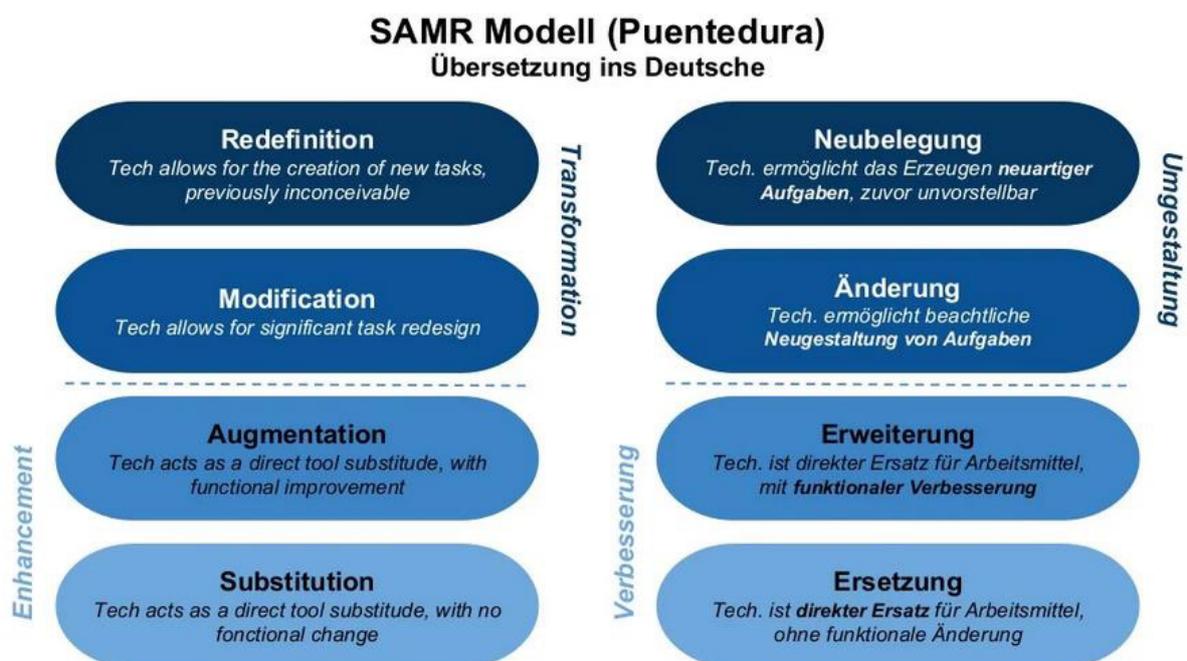


Abbildung 2-7: Das SAMR-Modell nach Puentedura (2012) (Übersetzung nach A. Wilke auf <http://homepages.uni-paderborn.de/wilke/blog/2016/01/06/SAMR-Puentedura-deutsch/>, letzter Zugriff am 25.04.2024).

In dem Modell werden die vier Ebenen *Ersatz (substitution)*, *Erweiterung (augmentation)*, *Änderung (modification)* und *Neubelegung (redefinition)* unterschieden. Ein bestimmter Einsatz digitaler Medien wird gegenüber traditionellem Vorgehen dahingehend reflektiert, ob dadurch überhaupt, nur funk-

tional oder auch ein inhaltlicher Mehrwert in Bezug auf die Unterrichtsaufgabe und das jeweilige Lernziel entsteht. Lehrkräfte sollten demnach die aufgeführten fachspezifischen oder -typischen digitalen Medien so in ihrem Unterricht einsetzen können, dass auch die beiden höheren Ebenen *Änderung* und *Neubelegung* des Modells erreicht werden und eine Umgestaltung im Sinne eines Mehrwerts für das fachliche Lernen resultieren kann (Abbildung 2-7).

Das bedeutet nicht, dass es im Unterricht nicht auch sinnvoll sein kann, Medien auf den beiden unteren Ebenen *Ersatz* und *Erweiterung* im Unterricht zu verwenden, wenn beispielsweise ein solcher Medieneinsatz zum angestrebten Lernziel passt (Hamilton et al., 2016). Nach Hamilton et al. (2016) sollte die Unterrichtspraxis weiterhin durch die Ziele der Lehrkraft und angestrebten Lernziele für die Lernenden bestimmt werden und nicht durch die Technologienutzung und das zwingende Erreichen höherer Ebenen im SAMR-Modell.

Das in diesem Kapitel beschriebene FDW zum Einsatz digitaler Medien bzw. dessen Entwicklung wird im Rahmen dieser Arbeit bei Physiklehramtsstudierenden untersucht, wozu ein entsprechendes Testinstrument zur Messung dieses FDW entwickelt und validiert wird. Im nachfolgenden Kapitel werden daher Grundlagen zur Kompetenzmessung in der Lehrkräftebildungsforschung beschrieben und konkrete Messverfahren von FDW (im Allgemeinen) sowie von TPACK (dem FDW zum Einsatz digitaler Medien ähnlichen Wissensbereich) vorgestellt.

2.4 Zusammenfassung theoretischer Grundlagen zum Fachdidaktischen Wissen zum Einsatz digitaler Medien

FDW stellt einen wichtigen Teil des Professionswissens und damit der professionellen Handlungskompetenz von (angehenden) Lehrkräften dar (2.1). Dieses fachspezifische Wissen wird nach Shulman (1986) von einer Lehrkraft benötigt, um fachliche Inhalte zu vermitteln und zu unterrichten (2.2). FDW wird in der Physik oder in den Naturwissenschaften hinsichtlich seiner inneren Struktur unterschiedlich modelliert, dabei meist unter der Berücksichtigung der Facetten zum Wissen über Schülervorstellungen sowie über Instruktionsstrategien (2.2.2). In einer umfassenden Modellierung nach

Gramzow et al. (2013) wurde die Facette (Digitale) Medien als eine von acht fachdidaktischen Facetten aufgegriffen (Abbildung 2-4), aber aufgrund anderer Schwerpunkte im entsprechenden Forschungsprojekt nicht operationalisiert. Diese Facette bzw. das FDW zum Einsatz digitaler Medien gewinnt zunehmend an Bedeutung (Kapitel 1) und stellt das zentrale Konstrukt der vorliegenden Arbeit dar.

Um eine spezifischere Beschreibung dieses FDW zum Einsatz digitaler Medien zu ermöglichen, werden weitere Strukturierungen zu digitalisierungsbezogenem Wissen oder Kompetenzen von Lehrkräften betrachtet (2.3.1). Das international verbreitete TPACK-Modell (Mishra & Koehler, 2006) erweitert die Professionswissensbereiche um technologiebezogene Wissensbereiche. Der zentrale Wissensbereich TPACK ähnelt dabei dem FDW zum Einsatz digitaler Medien, wobei das TPACK-Modell selbst keine Fachspezifität aufweist. Mithilfe des fachspezifischeren Orientierungsrahmens DiKoLAN (Becker et al., 2020) sowie der Fokussierung auf fachtypische und -spezifische Medien und Arbeitsweisen (Ropohl et al., 2018) kann eine spezifischere Beschreibung des FDW zum Einsatz digitaler Medien im Fach Physik erfolgen. Dabei spielt vor allem das Wissen über Möglichkeiten und Anforderungen bei der Nutzung physiktypischer und -spezifischer Medien wie digitale Messwertfassung, Simulationen, Animationen oder auch Erklärvideos eine wichtige Rolle (2.3.2).

3 Untersuchung von Fachdidaktischem Wissen und TPACK

In diesem Kapitel werden zu Beginn qualitative und quantitative Zugänge in der fachdidaktischen Forschung im Allgemeinen eingeführt (3.1). In den nachfolgenden Abschnitten wird der Fokus auf quantitative Kompetenzmessungen gelegt. Auch wenn in der vorliegenden Arbeit neben einem Leistungstest als quantitatives Kompetenzmessverfahren auch qualitative Interviews als Erhebungsmethode verwendet werden, stellt die Entwicklung des quantitativen Erhebungsinstruments (Leistungstest zum FDW zum Einsatz digitaler Medien) eine wesentlich bedeutsamere methodische Innovation der vorliegenden Arbeit dar. Dieses Theoriekapitel dient somit insbesondere als Grundlage für die Entwicklung des Leistungstests. Dazu werden in 3.2 Grundlagen zur Kompetenzmessung in der Lehrkräftebildungsforschung und in 3.3 Güte- und Validitätsaspekte dieser beschrieben. Anschließend werden Beispiele zu konkreten Messmodellen und Messverfahren zum FDW (3.4) sowie zum TPACK (3.5) vorgestellt. Schließlich wird in 3.6 eine Zusammenfassung dieses Kapitels gegeben.

3.1 Zugänge in der fachdidaktischen Forschung

Die fachdidaktische Forschung in Physik orientiert sich im Allgemeinen an Methoden aus den Sozialwissenschaften bzw. der empirischen Sozialforschung. Die dort bestehenden Verfahren oder Instrumente zur Datenerhebung und Auswertung müssen jedoch an die eigene Domäne und den jeweiligen Untersuchungskontext angepasst werden (z. B. Schecker et al., 2014). Dabei sind grundsätzlich in der fachdidaktischen Forschung (wie auch in der empirischen Sozialforschung) verschiedene Zugänge und Methoden denkbar, die qualitativ oder quantitativ ausgerichtet sein können oder eine Kombination davon darstellen (Baur & Blasius, 2014; Reinhold & Riese, 2020; Schecker et al., 2014). Im Vordergrund steht dabei nicht die Ausrichtung auf einen Forschungszugang, sondern die Beantwortung der Forschungsfrage mithilfe bestmöglich geeigneter (und möglicherweise verschiedener) Methoden für diese Fragestellung und den untersuchten Gegenstandsbereich (Baur & Blasius, 2014).

So zielen Verfahren quantitativer Forschung meist auf die Absicherung verallgemeinerbarer Aussagen ab und werden in der fachdidaktischen Forschung vor allem in Untersuchungen zur Wirksamkeit von Lehr- oder Lernkonzepten oder der Bestandsaufnahme von Lernständen größerer Stichproben angewendet (Schecker et al., 2014). Im Forschungsprozess werden die einzelnen Phasen von theoretischer Reflexion über Instrumentenentwicklung, Stichprobenbeziehung, Datenerhebung, -aufbereitung und -auswertung nacheinander und linear berücksichtigt. Die Erstellung oder Festlegung der Forschungsinstrumente und abzufragenden Variablen zu Beginn erfolgt jedoch bereits mit Blick auf die Datenauswertung, um geplante Auswertungen (die unter Umständen bestimmte Datenstrukturen erfordern) später auch durchführen zu können (Baur & Blasius, 2014). In der Regel werden die Erhebungsinstrumente vor der eigentlichen Untersuchung im Rahmen einer Pilotierung getestet, um Schwierigkeiten oder Mängel in den formulierten Fragen zu korrigieren. Ziel ist es, in allen Phasen und Arbeitsschritten im Forschungsprozess eine Objektivität, d. h. eine Unabhängigkeit von der Person der Forschenden zu erreichen (Baur & Blasius, 2014).

Hingegen wird in der qualitativen Forschung die Subjektivität der Forschenden nicht versucht auszuschalten, sondern stattdessen als relevante Forschungsressource angesehen. Objektivität bedeutet hier, dass Forschende ihr Vorgehen stets reflektieren und nachvollziehbar darstellen (Baur & Blasius, 2014). Die qualitative Forschung wird in der Regel genutzt, wenn ein neues Forschungsfeld explorativ erkundet und damit zur Theorienbildung beigetragen wird oder Hypothesen entwickelt oder aufgestellt werden sollen. In der fachdidaktischen Forschung bieten sich qualitative Zugänge daher an, wenn z. B. differenzierte Einblicke in Lernprozesse gewonnen, Typen oder Fälle charakterisiert oder Gründe für unterschiedliches (z. B. fachliches) Verständnis identifiziert werden sollen (Schecker et al., 2014). Der Forschungsprozess in der qualitativen Forschung ist zirkulär gestaltet, sodass sich die Phasen der Datenauswahl, -erhebung und -analyse wiederholt mit der theoretischen Reflexion abwechseln (Baur & Blasius, 2014). Meist werden wenige Fälle betrachtet und analysiert. Es gibt aber auch qualitative Methoden und Verfahren (z. B. die qualitative Inhaltsanalyse; Kuckartz, 2018;

7.2), die stärker der quantitativen, sequenziellen Logik folgen und bei denen z. T. sehr viel Material verarbeitet werden kann (Baur & Blasius, 2014).

Hinsichtlich der Datenerhebung stellt sowohl in der qualitativen als auch in der quantitativen Sozialforschung die Befragung das meistgenutzte Verfahren dar. Quantitative standardisierte Befragungen erfolgen über Fragebögen oder Tests und qualitative offene Befragungen über Interviews (Baur & Blasius, 2014). Entsprechend nutzt die fachdidaktische Forschung zur Datenerhebung diese Standardverfahren der Sozialforschung wie Tests, Fragebögen und Interviews sowie auch Beobachtungsverfahren oder Videoanalysen – z. B. zur Untersuchung der Unterrichtsqualität (Reinhold & Riese, 2020; Schecker et al., 2014). So werden beispielsweise fachspezifische Leistungstests entwickelt, um physikalisches Wissen bei Schülerinnen und Schülern oder die Wirksamkeit von unter physikdidaktischen Gesichtspunkten entwickelten Unterrichtskonzeptionen zu untersuchen (Reinhold & Riese, 2020). Fragebögen werden verwendet, wenn Interessen, Einstellungen oder die Selbsteinschätzung von Leistungen bei Schülerinnen und Schülern, Studierenden oder Lehrkräften erfasst werden sollen (Reinhold & Riese, 2020; Schecker et al., 2014). Während Antworten in Leistungstestaufgaben eindeutig als richtig oder falsch zu bewerten sind (3.2.1), gibt es bei Fragebögen keine richtigen oder falschen Antworten, sondern es werden meist mehrstufige Ratings (Likert-Skalen, z. B. Döring & Bortz, 2016b) zur Beantwortung (z. B. von ‚stimme gar nicht zu‘ bis ‚stimme vollkommen zu‘) verwendet (Schecker et al., 2014).

Als offene Befragung werden häufig Leitfaden- und Expert:innen-Interviews genutzt. Vor dem Interview wird eine Frageliste als Leitfaden erstellt, welche im Verlauf des Interviews abgearbeitet wird, aber auch im Interview um Nachfragen ergänzt werden kann (Baur & Blasius, 2014; Helfferich, 2011; 7.2). Klassische Leitfaden-Interviews fokussieren die persönliche Perspektive der befragten Personen und ihre Erfahrungen, Einstellungen oder Ansichten hinsichtlich des zu untersuchenden Gegenstands. Bei Expert:innen-Interviews wird den Interviewten – für den Kontext der Untersuchung – ein Expert:innenstatus zugesprochen und ein möglichst neutraler und breiter

Blick von ihnen erwartet (Baur & Blasius, 2014). In der fachdidaktischen Forschung können je nach Fragestellung beispielsweise Fachdidaktiker:innen, Hochschullehrende oder erfahrene Lehrkräfte Expert:innen darstellen.

Diese Unterscheidung in quantitativ gegenüber qualitativ ist insbesondere bei der Datenerhebung von Primärdaten sinnvoll; hinsichtlich der Datenauswertung greift sie jedoch zu kurz (Schecker et al., 2014). In der Regel verwendet man bei Texten (wie sie häufig in qualitativen Erhebungsverfahren wie Beobachtungen oder Interviews entstehen) qualitativ-inhaltsanalytische oder hermeneutische Verfahren, während man bei numerischen Daten (z. B. aus Punktzahlen bei Tests oder Ankreuzmustern bei Fragebögen) auf quantitativ-statistische Verfahren zurückgreift (Baur & Blasius, 2014; Schecker et al., 2014).

Teilweise werden in fachdidaktischen Studien außerdem mehrere Methoden kombiniert – auch durch Verbindung qualitativer und quantitativer Verfahren (Schecker et al., 2014). Dies kann im Sinne einer *Triangulation*, d. h. unter Einnahme unterschiedlicher Perspektiven bei der Beantwortung der Forschungsfrage (Flick, 2011) oder von *Mixed-Methods*, d. h. als kombinierter Einsatz und Integration von qualitativen und quantitativen Forschungsmethoden (Kuckartz, 2014) geschehen. Dabei wird der Begriff Mixed Methods häufig mit der Methodentriangulation (d. h. der Erfassung desgleichen Phänomens mit unterschiedlichen Methoden als häufigste Form der Triangulation) gleichgesetzt oder gemeinsam verwendet (Buchholtz, 2021; Kuckartz, 2014). Die beiden Konzepte weisen zwar Ähnlichkeiten auf, entstammen jedoch unterschiedlicher Grundansätze und sind nach Kuckartz (2014) weder gleichzusetzen, noch ist Mixed-Methods als Teilmenge der Triangulation anzusehen.

Triangulation bezog sich ursprünglich insbesondere auf die Validierung und wechselseitige Überprüfung von Ergebnissen; nach Flick (2011) liegt der Gewinn der Triangulation jedoch in der Erweiterung von Erkenntnismöglichkeiten durch die Betrachtung des untersuchten Gegenstands aus mehreren, möglichst gleichberechtigten Perspektiven. In der Regel wird diese Betrachtung aus zwei oder mehr Perspektiven durch die Verwendung verschiedener methodischer Zugänge realisiert (Methodentriangulation), sie kann jedoch auch auf Ebene der Daten, der Forschenden oder hinsichtlich theoretischer

Perspektiven erfolgen. Dadurch soll ein Erkenntniszuwachs ermöglicht werden, indem weiterreichende Erkenntnisse auf unterschiedlichen Ebenen gewonnen werden als bei der alleinigen Nutzung eines Zugangs (Flick, 2011). Betrachtet man die Ergebnisse einer Methodentriangulation oder -integration, so können diese (1) konvergent sein, d. h. übereinstimmen und damit Argumente für die Validität der Daten und Methoden liefern, (2) divergieren, d. h. sich widersprechen und somit Validitätsprobleme aufdecken oder (3) komplementär sein, d. h. sich wechselseitig ergänzen und damit ein umfassenderes Bild auf den Forschungsgegenstand geben (Flick, 2011; Kelle, 2014).

Im Mixed-Methods-Ansatz steht weniger die allgemeine Bereicherung von Perspektiven oder eine Validierung im Vordergrund, sondern eine konkrete Design- und Methodenwahl durch angemessene Kombination von qualitativen und quantitativen Methoden korrespondierend zum betrachteten Forschungsproblem oder zur betrachteten Forschungsfrage (Kuckartz, 2014). Es wird argumentiert, dass sich durch diese Kombination oder Integration unterschiedlicher Forschungsmethoden Stärken addieren oder Schwächen ausgleichen können (Buchholtz, 2021). In der fachdidaktischen Forschung bietet sich der Mixed-Methods-Ansatz vor allem im Bereich der Evaluationsforschung an, um sich hier weder auf das Berichten von Zahlen noch von verbalen Äußerungen zu beschränken, sondern beides nutzen und aufeinander beziehen zu können. Teilweise werden dabei komplexe Evaluationsdesigns aus Teilstudien angelegt, die parallel oder hintereinander den Evaluationsgegenstand über verschiedene Zugänge betrachten (Kuckartz, 2014).

Es gibt in der fachdidaktischen Forschung verschiedene Beispiele für die Kombination von qualitativen und quantitativen Methoden im Rahmen von Mixed-Methods-Designs oder als Triangulation, wobei beide Begriffe zum Teil gemeinsam verwendet und nicht unbedingt nach oben genanntem Verständnis getrennt werden. In der Arbeit von Ghassemi (2023) zur Evaluation eines Lehramtsstudiengangs für den Quereinstieg im Fach Physik wurden beispielsweise die Erkenntnisse aus einer quantitativen Teilstudie zur Entwicklung professioneller Handlungskompetenz mit denen einer qualitativen Interviewstudie zum Nutzen des Studiengangs aus individueller Perspektive integriert. Bernsteiner et al. (2024) evaluierten entwickelte fachdidaktische

Lernarrangements für Lehramtsstudierende mathematisch-naturwissenschaftlicher Fächer in einem Mixed-Methods-Ansatz unter Nutzung von standardisierten Fragebögen sowie offenen Reflexionsjournalen zu mehreren (Mess-)Zeitpunkten. In der Arbeit von Steinmetz (2021) zum kumulativen Lernen im Lehramtsstudium Physik evaluierte der Autor ebenfalls ein Lehrkonzept in insgesamt drei Teilstudien. In der ersten Teilstudie wurde der Fachwissenserwerb bei Physiklehramtsstudierenden (quantitativ) untersucht; in einer zweiten Teilstudie wurde die Untersuchung zum Fachwissen durch eine explorative Feldstudie ergänzt und in einer dritten Teilstudie wurde über qualitative Interviews der gewählte Lernzugang der Studierenden untersucht. Insgesamt ist auch dieses Design als Mixed-Methods zu bezeichnen, welches es ermöglichte, einen Zusammenhang zwischen Fachwissenserwerb und Lernzugang der Studierenden herzustellen. Steinmetz (2021) spricht dabei sowohl von Mixed-Methods als auch von einer Methodentriangulation. Im oben beschriebenen Verständnis trifft dies insbesondere auf die Kombination der ersten beiden Teilstudien zum Fachwissen zu, da somit die Validität der Einzelfalldiagnostik erhöht werden konnte (Steinmetz, 2021).

Zusammengefasst zeigt sich für die fachdidaktische Forschung eine Auswahl an Zugängen und Verfahren, die sich an denen der empirischen Sozialforschung orientiert. Dabei werden sowohl qualitative als auch quantitative Methoden sowie die Kombination dieser eingeschlossen – sei es als Ergebniserweiterung oder -validierung im Sinne einer Triangulation oder als Mixed-Methods-Ansatz z. B. in Untersuchungsdesigns zur Evaluation entwickelter fachdidaktischer Konzepte.

In den folgenden Abschnitten wird nun der Fokus auf die (quantitative) Messung von Kompetenzen in der Lehrkräftebildungsforschung gelegt. In diesem Zusammenhang werden Grundlagen zu verschiedenen Verfahren der Kompetenzmessung (3.2) sowie zur Güte und Validität solcher Messverfahren (3.3) beschrieben. Anknüpfend werden konkrete Messinstrumente zum FDW (3.4) sowie zum TPACK (3.5) exemplarisch vorgestellt.

3.2 Kompetenzmessung in der Lehrkräftebildungsforschung

Die (fachspezifische) Lehrkräftebildungsforschung stellt einen Teil fachdidaktischer Forschung dar und beschäftigt sich unter anderem mit der Messung professioneller Kompetenz von (angehenden) Lehrkräften. Allgemein können mithilfe von Kompetenzmessungen Kompetenzstrukturmodelle, wie beispielsweise in 2.1 vorgestellt, fundiert werden, d. h. hinsichtlich ihrer Struktur empirisch überprüft werden. Weiterhin dienen Kompetenzmessungen dazu, Lernstände in Bildungsprozessen für darauf aufbauende individuelle Förder- und Auswahlentscheidungen festzustellen oder Bildungsmaßnahmen zu evaluieren (Schaper, 2009). Kompetenzmessung nimmt somit eine Schlüsselfunktion für die Verbesserung von Bildungsprozessen und die Weiterentwicklung des Bildungswesens ein (Hartig & Jude, 2007; Klieme & Leutner, 2006). Dazu sind entsprechende Messverfahren bzw. -instrumente notwendig. Die verwendeten Kompetenzmessinstrumente sollten auf Kompetenzstrukturmodellen beruhen oder auf deren Grundlage entwickelt werden (Schaper, 2009). Die Erfassung von Kompetenzen ist dabei als Teil empirischer Sozialforschung auch an den Methoden der pädagogischen Psychologie und der Sozialwissenschaften orientiert (3.1). Schaper (2009) betont jedoch, dass Ansätze zur Kompetenzmessung sich von den herkömmlichen pädagogisch-psychologischen Ansätzen unterscheiden. Im Fokus steht hier die Erfassung komplexer, facettenreicher Leistungsdispositionen, die zur Bewältigung komplexer Anforderungssituationen notwendig sind und damit weniger klar definierte Konstrukte darstellen als solche in der pädagogischen Psychologie oder Diagnostik (z. B. Intelligenz oder Konzentrationsfähigkeit). Verfahren zur Kompetenzmessung orientieren sich demnach grundlegend an typischen Verfahren der pädagogischen Psychologie, berücksichtigen aber insbesondere die für die untersuchte Kompetenz relevanten Anforderungskontexte, um eine situations- und anforderungsbezogene Erfassung zu ermöglichen (Schaper, 2009).

3.2.1 Verfahren der Kompetenzmessung

Zur Erfassung von Kompetenz als latentes, d. h. nicht direkt beobachtbares Merkmal wird in der Regel auf quantitative Verfahren und Methoden (3.1) zurückgegriffen. Qualitative Verfahren, die z. B. auf Interviews, Beobachtungen

oder Portfolios zurückgreifen, eignen sich nach Schaper (2009) eher zur Beschreibung, Dokumentation und Förderung von Kompetenzen, jedoch nicht für eine solide Messung dieser.

Im Hinblick auf quantitative Messverfahren zur Erfassung latenter Merkmale lassen sich subjektive oder objektive Verfahren unterscheiden. Bei subjektiven Verfahren erfolgt die Einschätzung der Merkmalsausprägung (z. B. der Kompetenz) durch die befragte Person selbst (Kunter & Klusmann, 2010). Dazu zählen demnach Fragebögen oder Persönlichkeitstests auf Basis von Selbstauskünften, wobei zur Beurteilung häufig mehrstufige Ratingskalen verwendet werden (Döring & Bortz, 2016a, 2016b; Moosbrugger & Brandt, 2020; 3.1). Objektive Erfassung hingegen bedeutet, dass die Erhebung „von außen“ anhand von externen Kriterien erfolgt. Es wird dabei angenommen, dass das Messergebnis unabhängig von der jeweiligen beurteilenden Person ist (Kunter & Klusmann, 2010). Objektive Ansätze lassen sich weiterhin in distale und proximale Erfassung unterteilen. Distale Indikatoren werden auch als indirekte Indikatoren bezeichnet und können beispielsweise eine Note in der Schule, in der Ausbildung oder die Art des Abschlusses einer Person sein. Proximale Indikatoren erfassen hingegen direkt die kognitiven Merkmale, welche – auf Basis theoretischer Annahmen – ursächlich für das interessierende Handeln sind (Kunter & Klusmann, 2010). Zu proximalen Messverfahren zählen beispielsweise Wissens- oder Leistungstests, in denen das interessierende Merkmal über mehrere Testaufgaben erschlossen wird. Die Testaufgaben in solchen Leistungstests sind dabei richtig oder falsch zu lösen (3.1). Die Antworten können entsprechend dichotom oder auch mehrstufig bewertet und zu einem Testwert verrechnet werden, der schließlich die Ausprägung des interessierenden Merkmals darstellt (Döring & Bortz, 2016a; Moosbrugger & Kelava, 2020).

Zur Erfassung von Kompetenzen oder Kompetenzaspekten wird in der Forschung sowie im Bildungsbereich in der Regel auf Methoden zurückgegriffen, welche fragebogenbezogene Selbst- und Fremdeinschätzungsverfahren sowie Intelligenztests, aber insbesondere auch simulationsorientierte Formate wie szenarienbasierte Tests, Unterrichts- oder Arbeitsproben und Entwicklungsaufgaben einschließen (Frey, 2006; Schaper, 2009). Diese Methoden

lassen sich dabei den genannten Verfahren (subjektiv oder objektiv) zuordnen. Zur Erfassung kognitiver Kompetenzaspekte (wie dem FDW) bieten sich Wissens- oder Leistungstest an, die in passende Handlungs- oder Anforderungskontexte eingebettet sind (Schaper, 2009). Möchte man also die professionellen Kompetenzen von Lehrkräften empirisch untersuchen, wird man sich an solchen Methoden orientieren und gegebenenfalls neue Instrumente entwickeln oder bestehende adaptieren müssen (Frey, 2006). Zu beachten ist dabei, dass verschiedene Erfassungsmethoden auch verschiedene Gültigkeitsbereiche aufweisen (Kunter & Klusmann, 2010), welche es abzuwägen gilt.

So haben Selbstauskünfte als subjektive Verfahren den Vorteil, dass sie in kurzer Zeit durchgeführt werden können. Sie stellen daher eine zeitökonomische Erfassungsmethode dar – vor der Annahme, dass die jeweilige Person am besten eine Auskunft über sich geben kann (Amelang und Bartussek, 2001, zitiert nach Frey, 2006). Dies gilt jedoch nur unter der Voraussetzung, dass die befragte Person sich retrospektiv realistisch einschätzen kann und nicht etwa (unbewusste) fehlerhafte Einschätzungen gibt (Frey, 2006). Die Validität von Selbsteinschätzungen wird in Bezug auf die Kompetenzmessung aufgrund solcher Verzerrungen in Frage gestellt (Schaper, 2009; Seifert et al., 2009). Hartig und Jude (2007) argumentieren weiterhin, dass Instrumente auf Basis von Selbstauskünften eher das Selbstkonzept eigener Fähigkeiten, nicht aber die Kompetenz selbst messen. So zeigte sich z. B. im Rahmen der COACTIV-Studie zur Untersuchung der professionellen Kompetenz von Mathematiklehrkräften (2.1), dass nur ein geringer Zusammenhang zwischen selbsteingeschätztem und proximal gemessenem Fachwissen und gar kein entsprechender Zusammenhang für das FDW besteht. Die Befunde deuten darauf hin, dass Selbstaussagen nur begrenzt einen angemessenen Indikator für objektiv vorhandene Kompetenzen darstellen (Kunter & Klusmann, 2010). Die Problematik von Selbsteinschätzungen ist aufgrund deren vielfacher Nutzung insbesondere für die Untersuchungen zum TPACK-Modell (2.3.1) bedeutsam und wird in 3.5.3 nochmals aufgegriffen.

Mit Leistungstests wird das Wissens oder die Kompetenz der Befragten proximal erfasst, da es objektiv richtige und falsche Antworten auf die Testaufgaben gibt. Für Leistungstests, wie sie in der fachdidaktischen Forschung zur

Erfassung von Wissen und kognitiven Kompetenzkomponenten eingesetzt werden, ist dabei ein breites Spektrum an Leistungsanforderungen denkbar, welches von Faktenwissen bis zur Lösung komplexer Probleme reicht (Schecker et al., 2014). Um wirklich Kompetenzaspekte zu erfassen, die realitätsnahe Anforderungskontexte von Lehrkräften adressieren, sollten solche Testinstrumente und -aufgaben in einer „kontextualisierten Form“ verwendet werden, d. h. eingebettet in passende Handlungs- oder Problemkontexte (Schaper, 2009). Leistungstests bringen dadurch gegenüber Instrumenten zu Selbstauskünften auch entsprechende Herausforderungen mit sich. Sie erfordern meist eine höhere zeitliche und kognitive Belastung bei der Bearbeitung durch die Teilnehmenden. Zudem ist der Aufwand zur Erstellung des Testinstruments größer, da in der Regel umfangreichere Testaufgaben und Vorschriften zur Bewertung entwickelt werden müssen sowie häufig verschiedene Validierungsstudien mit der Testentwicklung einhergehen (z. B. Riese & Reinhold, 2014; Tepner & Dollny, 2014). Der Aufwand in der Entwicklung der einzelnen Leistungstestaufgaben unterscheidet sich zudem je nach Format der Aufgaben. Im Folgenden werden der Aufbau von Leistungstestaufgaben sowie verschiedene Antwortformate beschrieben.

3.2.2 Aufbau von Leistungstestaufgaben

Aufgaben in Leistungstests setzen sich grundlegend aus zwei Teilen zusammen: dem Aufgabenstamm, welcher die Aufgabe selbst als Stimulus sowie ggfs. zugehörige Informationen zum Aufgabenkontext enthält, und dem Antwortformat, in dem Vorgaben festgelegt werden, wie die Antwort durch die Testperson erfolgen soll (z. B. Brandt & Moosbrugger, 2020; Schaper, 2009). Bei Testinstrumenten zur Kompetenzmessung ist es demnach wichtig, über den Aufgabenstamm eine Einbettung der Aufgabenstellung in realitätsnahe Anwendungskontexte zu schaffen (Schaper, 2009). Zur Messung professioneller Kompetenzen von Lehrkräften bietet die Einbindung von Unterrichtsvignetten in den Aufgabenstamm als pädagogischer Kontext einen Ansatz zum näheren Messen am Unterrichtsgeschehen. Diese Vignetten sind (meist fiktive) prototypische Situationen aus dem Unterricht oder dem weiteren Alltag von Lehrkräften, die Probleme aufzeigen, zu deren Bewältigung die un-

tersuchten Kompetenzen gefordert sind. Durch die Verwendung von (Unterrichts-)Vignetten wird vermieden, „träges Wissen“ zu messen und hingegen für das berufliche Handeln erforderliche Kompetenzaspekte valider abzubilden (Rehm & Bölsterli, 2014; Tepner & Dollny, 2014). Unterrichtsvignetten reichen dabei von der schriftlichen Wiedergabe von Szenarien (z. B. Interaktionssequenzen zwischen Lehrkraft und Lernenden im Unterrichtsgeschehen) bis hin zur videogestützten Präsentation realer Situationen und Probleme (Aufschnaiter & Blömeke, 2010; Rehm & Bölsterli, 2014; Schaper, 2009). Die Testpersonen sollen in den entsprechenden Testaufgaben in der Regel zu den beschriebenen Situationen in den Vignetten Stellung nehmen oder Lösungsvorschläge (zur Fortführung) geben (Baer et al., 2007; Riese & Reinhold, 2008; Tepner & Dollny, 2014).

Antwortformate zur Beantwortung der Testaufgaben lassen sich in der Regel nach ihrer Offenheit und ihrem Strukturierungsgrad klassifizieren. Übergeordnet werden freie oder (halb-)offene Antwortformate von gebundenen oder geschlossenen Antwortformaten unterschieden (Döring & Bortz, 2016a; Moosbrugger & Brandt, 2020). Dabei wird beim freien oder (halb-)offenen Antwortformat die Antwort von der Testperson selbst formuliert, oft in einer vorgegebenen Struktur. Beim gebundenen oder geschlossenen Antwortformat werden Antworten oder Antwortalternativen vorgegeben, welche von der Testperson häufig als richtig oder falsch bewertet, ausgewählt (Auswahlaufgaben) oder nach bestimmten Vorgaben geordnet oder zugeordnet (Um- oder Zuordnungsaufgaben) werden. Auswahlaufgaben können dabei zwei (dichotome Auswahlaufgaben) oder mehrere (Multiple-Choice- bzw. Mehrfachwahlaufgaben) zur Auswahl stellen (Döring & Bortz, 2016a; Moosbrugger & Brandt, 2020).

Aufgaben mit freien bzw. halboffenen Antwortformaten bieten den Testpersonen mehr Spielraum in der Beantwortung der Aufgabe und vermeiden eine Einschränkung durch enge Vorgaben (Dollny, 2011; Döring & Bortz, 2016a). Freie Antwortformate bieten sich für die Kompetenzmessung an, da das Formulieren komplexerer Antworten oder Lösungsbeschreibungen den kompetenzorientierten Anforderungen (z. B. bei Aufgabenstellungen mit Unterrichtsvignetten) eher gerecht wird (Schaper, 2009). Demgegenüber stehen jedoch Nachteile freier Antwortformate im Hinblick auf den hohen Aufwand

bei der Beantwortung und vor allem bei der Bewertung der Testaufgaben. Seitens der Testleitung muss ein umfangreiches Kodiermanual entwickelt sowie die bewertenden Personen entsprechend geschult werden, um eine eindeutige objektive Bewertung zu ermöglichen.

Bei geschlossen Antwortformaten ist eine objektive Bewertung in der Regel trivial und auch zeitökonomisch (Dollny, 2011; Döring & Bortz, 2016a; Moosbrugger & Brandt, 2020). Dafür stellt sich im Sinne der professionellen Kompetenz die Frage, inwiefern aus der Lösung solcher Aufgaben mit vorgegebenen Antwortalternativen auf die kompetente Bewältigung realer Situationen geschlossen werden kann (Hartig & Jude, 2007). Weiterhin stellt die Formulierung angemessener Antwortalternativen bei der Konstruktion von geschlossenen Testaufgaben eine große Schwierigkeit dar. Dies gilt insbesondere für die Konstruktion der inkorrekten Alternativen (Distraktoren), welche einerseits eindeutig falsch, andererseits ausreichend plausibel sein sollen, sodass sie unwissende Testpersonen von der richtigen Antwort ablenken (Döring & Bortz, 2016a; Gronlund, 1974). Eine zusammenfassende Übersicht nach Hartig und Jude (2007) zu den verschiedenen Antwortformaten sowie ihren jeweiligen Vor- und Nachteilen zeigt Abbildung 3-1.

| halboffene/freie Antwortformate | geschlossene/gebundene Antwortformate |
|--|--|
| <ul style="list-style-type: none"> • Freitextantworten • ausformulierte Texte | <ul style="list-style-type: none"> • richtig / falsch • Um- / Zuordnungsaufgaben • Auswahlaufgaben (dichotom, Mehrfachwahl/Multiple-Choice) |
| <ul style="list-style-type: none"> + einfach zu konstruieren + hohe Inhaltsvalidität | <ul style="list-style-type: none"> – anspruchsvolle Konstruktion – fragliche Validität |
| <ul style="list-style-type: none"> – zeitaufwändige Beantwortung – aufwändige Auswertung – fragliche Objektivität | <ul style="list-style-type: none"> + schnell zu beantworten + einfache Auswertung + hohe Objektivität |

Abbildung 3-1: Übersicht zu offenen und geschlossenen Antwortformaten mit Beispielen sowie Vor- und Nachteilen in Anlehnung an Hartig und Jude (2007, S. 30)

Im folgenden Kapitel werden nun als Anforderungen an Leistungstests wissenschaftliche Gütekriterien für Testinstrumente beschrieben. Dabei werden vertieft verschiedene Validitätsaspekte vorgestellt und Möglichkeiten zu

deren Untersuchung und Gewährleistung erläutert. Insbesondere für Leistungstestaufgaben im Multiple-Choice-Format stellt die Gewährleistung der Validität eine Herausforderung dar.

3.3 Güte und Validitätsaspekte von Kompetenzmessungen

Leistungstests und weitere Verfahren zur Messung von Kompetenzen oder kognitiven Kompetenzaspekten wie dem FDW müssen zur angemessenen Interpretierbarkeit und Nutzung der Ergebnisse wissenschaftlichen Gütekriterien gerecht werden. Häufig werden in diesem Zusammenhang die drei Hauptgütekriterien Objektivität, Reliabilität und Validität als Testgütekriterien betrachtet (z. B. Döring & Bortz, 2016a; Moosbrugger & Kelava, 2020).

3.3.1 Hauptgütekriterien Objektivität, Reliabilität und Validität

Objektivität bedeutet, dass ein Testergebnis unabhängig von der anwendenden bzw. testleitenden Person ist. Dabei kann zwischen Durchführungs- (Unabhängigkeit von der Person, die den Test durchführt), Auswertungs- (Unabhängigkeit von der Person, die den Test bewertet) und Interpretationsobjektivität (Unabhängigkeit von der Person, die den Testwert interpretiert) unterschieden werden. Objektivität kann durch Testmanuale und Kodiermanuale mit genauen Vorgaben zur Testdurchführung und zur eindeutigen Bewertung der Testaufgaben sowie zum Vorgehen bei der Testinterpretation sichergestellt werden (Döring & Bortz, 2016a). Insbesondere die eindeutige Bewertung von Testaufgaben im Sinne der Auswertungsobjektivität kann je nach Aufgabenformat eine Herausforderung darstellen (3.2.2). Um eine angemessene Auswertungsobjektivität zu gewährleisten, bietet sich die Untersuchung der Beurteilendenübereinstimmung von zwei oder mehreren Rater:innen z. B. hinsichtlich der Bewertung der Testantworten an. Neben einfacher prozentualer Übereinstimmung stellt vor allem Cohens Kappa (Cohens κ ; Cohen, 1960) ein gängiges Maß für die Beurteilendenübereinstimmung (von zwei Urteilenden) dar, welches die prozentuale Übereinstimmung unter Berücksichtigung zufälliger Übereinstimmungen korrigiert (Döring & Bortz, 2016a). Cohens κ kann rechnerisch Werte zwischen -1 und +1 annehmen, wobei die Übereinstimmung umso besser ist, je höher der κ -Wert ist. Werte

ab 0,40 werden als mittelmäßig, ab 0,60 als gut und ab 0,75 als sehr gut eingestuft, wobei zur Interpretation des κ -Werts stets seine Abhängigkeit von der Anzahl der möglichen Kategorien (z. B. richtig oder falsch; 0, 1, 2 oder 3 Punkte etc.) berücksichtigt werden sollte (Döring & Bortz, 2016a).

Mit der Reliabilität eines Tests ist die Zuverlässigkeit oder auch Messgenauigkeit gemeint. Das Konzept der Reliabilität ist insbesondere in der klassischen Testtheorie (KTT: z. B. Moosbrugger et al., 2020) bedeutsam, in welcher eine Grundannahme ist, dass sich ein Testwert aus dem wahren Wert der Merkmalsausprägung und einem Messfehler zusammensetzt (Moosbrugger et al., 2020; Moosbrugger & Kelava, 2020). Der Testwert sollte die wahre Merkmalsausprägung präzise abbilden, das heißt ein Test sollte möglichst wenig Verzerrungen durch Messfehler aufweisen. Die Reliabilität kann in der KTT mit verschiedenen Methoden und Reliabilitätskoeffizienten abgeschätzt werden. Der häufig verwendete Reliabilitätskoeffizient Cronbachs Alpha (Cronbachs α ; Cronbach, 1951) betrachtet die Korrelationen aller einzelnen Testaufgaben miteinander als ein Maß für die interne Konsistenz der Testaufgaben (Döring & Bortz, 2016a; Moosbrugger & Kelava, 2020). Cronbachs α kann rechnerisch Werte zwischen 0 und 1 annehmen. In der Literatur werden häufig Werte ab 0,80 (Döring & Bortz, 2016a) oder 0,70 (Schmitt, 1996) als anzustrebende Schwellen genannt. Jedoch sollten diese Richtwerte gerade in der fachdidaktischen Forschung nicht als absolutes Hindernis für die Verwendung eines Testinstruments oder einer Skala gesehen werden (Schecker, 2014; Schmitt, 1996). Schmitt (1996) argumentiert, dass auch eine Messung mit geringer Reliabilität wie z. B. 0,49 nutzbar sein kann, wenn sie andere wünschenswerte Eigenschaften wie eine aussagekräftige inhaltliche Abdeckung eines Gebiets und angemessene Eindimensionalität aufweist. Diese Einschätzung ist für die Kompetenzmessung in der fachdidaktischen Forschung bedeutsam. Hier werden häufig Konstrukte adressiert, die in einer gewissen inhaltlichen Breite erfasst werden sollen und deren Operationalisierung anspruchsvoll ist. Sowohl daher als auch wegen der begrenzten Testzeit, die nicht das Abfragen beliebig vieler Testaufgaben ermöglicht, sind für diese fachdidaktisch inhaltlich anspruchsvollen Leistungstests niedrigere Werte als 0,70 für Cronbachs α nicht ungewöhnlich (Schecker, 2014).

Bei Tests auf Basis der probabilistischen Testtheorie bzw. Item-Response-Theorie (IRT: z. B. Kelava & Moosbrugger, 2020b) gibt es andere Verfahren zur Genauigkeitsbeurteilung der Testwerte. In der IRT wird die Grundannahme verfolgt, dass die Lösungswahrscheinlichkeit einer Aufgabe für eine Person von der Schwierigkeit der Aufgabe (Itemparameter) sowie der Fähigkeit der Person (Personenparameter) abhängt. Diese Item- und Personenparameter werden geschätzt, wobei es verschiedene Methoden und Schätzer gibt. Entsprechend ist auch die Bestimmung der Standardfehler und Reliabilität der Personenparameterschätzung je nach verwendetem Schätzer verschieden (Rose, 2020). Während die Genauigkeitsbeurteilung in der KTT also pauschal für alle Testpersonen gilt, erfolgt diese in der IRT testwertabhängig, d. h. unter Berücksichtigung der Merkmalsausprägung einer einzelnen Testperson (Moosbrugger & Kelava, 2020). Es gibt jedoch marginale, d. h. gemittelte Reliabilitätsmaße für einzelne Schätzer, die sich ähnlich der Reliabilitätsmaße der KTT wie Cronbachs α als Genauigkeitsmaß für das Testinstrument interpretieren lassen (Rose, 2020). Ein gebräuchliches marginales Reliabilitätsmaß stellt beispielsweise die EAP-Reliabilität dar, welche auf *Expected-a-posteriori*-Schätzer (EAP-Schätzer), d. h. Erwartungswerte der A-posteriori-Verteilung als Punktschätzer für die Personenparameter zurückgreift. Diese Verteilung berücksichtigt dabei einerseits Vorwissen in Form einer A-priori-Verteilung sowie andererseits empirische Informationen der Stichprobe (Rose, 2020).

Die Validität wird häufig als wichtigstes Gütekriterium eines Tests betrachtet und meint im engeren Sinne die Gültigkeit eines Tests; d. h., dass ein Test wirklich das Merkmal messen soll, welches er zu messen vorgibt (Moosbrugger & Kelava, 2020). Hohe Objektivität und hohe Reliabilität gelten als notwendige Voraussetzungen für die Validität eines Testinstruments, sie sind jedoch nicht hinreichend. Selbst ein objektiver und reliabler Test ist unbrauchbar, wenn er nicht das Konstrukt misst, welcher er zu messen beansprucht (Döring & Bortz, 2016a). Auch für die Validität können traditionell verschiedene Validitätsaspekte unterschieden werden, beispielsweise Augenschein- und Inhaltsvalidität, Konstruktvalidität und Kriteriumsvalidität (Döring & Bortz, 2016a; Moosbrugger & Kelava, 2020). Die Validität stellt ins-

besondere bei der Kompetenzmessung mithilfe geschlossener Aufgabenformate eine Herausforderung dar und ist für die vorliegende Arbeit daher von besonderer Bedeutung. Im Folgenden wird deshalb auf dieses Gütekriterium vertiefter eingegangen.

3.3.2 Argumentbasiertes Validitätsverständnis

In der Vergangenheit hat sich ein umfassenderes Validitätsverständnis etabliert, dass Validität nicht nur als Testgütekriterium und Eigenschaft eines Tests versteht, sondern vielmehr die Bedeutung und Interpretation der Testergebnisse sowie deren Verwendung in den Blick nimmt (z. B. Blömeke, 2013; Schaper, 2015). Dieses Verständnis wird vor allem durch den *argument based approach to validation* nach Kane (1992, 2001) und dem einheitlichen Konzept der Konstruktvalidität nach Messick (1995) geprägt. Nach diesem Validitätskonzept stellt Validität keine Eigenschaft des Tests dar, sondern vielmehr eine Eigenschaft der vorgeschlagenen Interpretationen und Verwendungen der Testergebnisse (Kane, 2013, S. 3). Das bedeutet insbesondere, dass die Interpretation der Ergebnisse des Testverfahrens sowie die beabsichtigten Schlussfolgerungen aus dieser Interpretation eine Rolle für die Bewertung der Validität spielen. Für die Kompetenzmessung in Ausbildungskontexten ist die Angemessenheit der praktischen Schlussfolgerungen aus den diagnostischen Ergebnissen der Kompetenzmessung von großer Bedeutung, weswegen klassische Validitätsbetrachtungen, die sich nur auf das Testverfahren als solches beziehen, für die Kompetenzmessung in diesen Kontexten zu kurz greifen (Schaper, 2015).

Zur Bewertung der Interpretation hinsichtlich der Validität ist es nach Kane (2001) zunächst notwendig, die vorgeschlagene Testwertinterpretation sowie den angestrebten Umgang damit darzulegen. Ohne eine konkrete Beschreibung der zu validierenden Testwertinterpretation ist die Validierung dieser praktisch unmöglich. Anknüpfend werden zur Validierung Argumente aus theoretischen Annahmen und empirischen Belegen gesammelt, um die Plausibilität der Interpretation und Schlussfolgerungen zu bewerten (Kane, 2001, 2013). Messick (1995) bezeichnet dieses umfassende Konzept der Validität mit dem Begriff der *Konstruktvalidität*. Zur Unterstützung einer ange-

messenen Testwertinterpretation können dabei verschiedene Arten von Argumenten aus unterschiedlichen Evidenzformen betrachtet werden. Messick (1995) unterscheidet dazu sechs Aspekte der Konstruktvalidität: (1) inhaltliche Validität, (2) kognitive Validität, (3) strukturelle Validität, (4) Verallgemeinerbarkeit, (5) externe Validität und (6) konsequentielle Validität (siehe auch Schaper, 2015). Diese Validitätsaspekte oder Evidenzformen helfen bei der Beurteilung der Angemessenheit und Aussagekraft der Testwertinterpretation und können somit als Qualitätskriterien für pädagogische Kompetenzmessung dienen (Messick, 1995). Im Folgenden werden die einzelnen Aspekte (1) bis (6) sowie Möglichkeiten ihrer Untersuchung beschrieben.

Der **inhaltliche Aspekt** der Konstruktvalidität (1) bezieht sich auf Evidenzen zur inhaltlichen Relevanz und Repräsentativität des Testinhalts für das interessierende Konstrukt (Messick, 1995). Eine Unterrepräsentation des zu messenden Konstrukts sollte vermieden werden, d. h. der Test darf keine relevanten Facetten oder Bereiche des Konstrukts auslassen. Gleichzeitig sollte auch keine konstruktirrelevante Varianz vorliegen. D. h., dass Eigenschaften des Testverfahrens, die für das Konstrukt nicht bedeutsam sind, für bestimmte Teilnehmende eine konstruktirrelevante Schwierigkeit oder Leichtigkeit bei der Beantwortung der Testaufgaben erzeugen (Messick, 1995; Schaper, 2015). Beispielsweise wäre die inhaltliche Validität in diesem Sinne verletzt, wenn ein Test zur Messung der mathematischen Fähigkeit im Bereich der Grundrechenarten eine komplexe Aufgabenstellung wie z. B. „Bestimme die Differenz der Zahl 90 und der nächstkleineren Primzahl.“ enthielte, welche insbesondere Fähigkeiten außerhalb des interessierenden Konstrukts adressiert. Würde dieser Test nur Aufgaben zur Subtraktion und Division berücksichtigen, aber keine zur Addition und Multiplikation, läge eine Unterrepräsentation des interessierenden Konstrukts vor. Der inhaltliche Aspekt der Validität spielt bereits vor und während der Testentwicklung eine Rolle, indem zunächst das zu erfassende Konstrukt möglichst exakt beschrieben wird, um dazu relevante und repräsentative Testaufgaben zu entwickeln (Schmiemann & Lücken, 2014). Die Bewertung des inhaltlichen Validitätsaspekts erfolgt dabei meist theoretisch gestützt und seltener durch empirische Untersuchungen (Hartig et al., 2012). Häufig werden zur Beurteilung

des Aspekts außerdem Expert:inneneinschätzungen eingeholt, um die Relevanz und Repräsentativität des Testinhalts bzw. der Aufgabeninhalte zu bewerten. Einschätzungen von Expert:innen sind dabei nicht nur aufgrund ihrer fachlichen Expertise, sondern ebenso aufgrund ihrer externen Perspektive wertvoll. Externe Personen, die bei der Entwicklung der Testaufgaben nicht direkt beteiligt waren, können Unstimmigkeiten in den Testaufgaben im Sinne mangelnder inhaltlicher Validität feststellen, die den beteiligten Testentwickelnden gegebenenfalls nicht mehr auffallen (Schmiemann & Lücken, 2014).

Die **kognitive Validität** ((2); „*substantive aspect*“ der Konstruktvalidität, vgl. Messick, 1995, S. 745) berücksichtigt empirische Evidenzen, dass theoretisch antizipierte Prozesse bei der Testbearbeitung von den Teilnehmenden tatsächlich stattfinden und zur Geltung kommen (Messick, 1995). Die Prozesse, die zur Lösung der Testaufgaben notwendig sind, sollen repräsentativ für das untersuchte Konstrukt sein (Schaper, 2015). Demnach stellt dieser Aspekt neben dem inhaltlichen Aspekt einen weiteren Teil der Repräsentativität dar (Messick, 1995). Über traditionelle professionelle Einschätzungen hinaus sollen dabei auch empirisch fundierte Argumente aus der Testbearbeitung ähnlichen Situationen gesammelt werden (Messick, 1995). Für solche Untersuchungen werden beispielsweise die Methode des Lauten Denkens bzw. *Think-Aloud*-Studien (Ericsson & Simon, 1980) oder Aufzeichnungen und Analysen von Augenbewegungen genutzt (Schaper, 2015). Gerade Lautes Denken stellt in der Kompetenzmodellierung und Testentwicklung eine häufig verwendete Methode dar, um zu untersuchen, wie die Testaufgaben von den Testpersonen wahrgenommen und verstanden werden (Sandmann, 2014). Die Teilnehmenden sprechen beim Lauten Denken ihre Gedanken zeitgleich während der Bearbeitung der Testaufgaben laut aus, sodass diese Verbalisierungen als Audio oder Video aufgenommen, protokolliert und anschließend analysiert werden können (Sandmann, 2014; Schmiemann & Lücken, 2014). Somit können Verständnisprobleme oder missverständliche Formulierungen in Aufgabenstellungen oder Antwortalternativen identifiziert und Hinweise zur Testoptimierung gewonnen werden (Sandmann, 2014). Im Sinne der kognitiven Validierung kann mit der Methode des Lauten Denkens aus den verbalisierten Überlegungen der Teilnehmenden bei der

Testbearbeitung untersucht werden, inwiefern tatsächlich der zu untersuchende Wissens- oder Fähigkeitsbereich bei der Bearbeitung einer Testaufgabe aktiviert wird oder ob womöglich andere Überlegungen Grund für die entsprechende Antwort sind (Schmiemann & Lücken, 2014). Gerade bei der Konstruktion von Mehrfachwahlaufgaben ist in diesem Sinne wichtig, dass richtige Antwortalternativen nicht beispielsweise allein durch sprachliche Hinweise identifiziert (z. B. Gronlund, 1974) werden sollten, sondern tatsächlich das zu untersuchende Wissen zur korrekten Lösung der Aufgabe benötigt wird (8.2).

Der **strukturelle Aspekt** der Konstruktvalidität (3) fordert die Übereinstimmung des einer Messung zugrundeliegenden Mess- und Bewertungsmodells mit der Struktur des Konstrukts (Messick, 1995; Schaper, 2015). Die Struktur der Testbewertung, d. h. auf welche Weise das Testverhalten zu einem Testwert kombiniert wird, sollte demnach mit der theoretisch angenommenen Struktur des Konstrukts übereinstimmen (Messick, 1995). Zur Untersuchung des strukturellen Aspekts werden in der Regel Messmodelle mit verschiedenen strukturellen oder dimensional angenommenen im Hinblick auf ihre Passung zu den empirischen Zusammenhängen im erhobenen Datensatz verglichen (Schaper, 2015). Es kann vorkommen, dass auf Basis entsprechender empirischer Ergebnisse das zugrundeliegende Modell anschließend verworfen oder angepasst werden muss. Für Kompetenzmodellierung und -messung unter Verwendung probabilistischer IRT-Modelle bezieht sich der strukturelle Aspekt der Konstruktvalidität nicht nur auf die Struktur der Testbewertung, sondern auch auf das gewählte probabilistische Messmodell (Schaper, 2015).

Der **externe Aspekt** der Konstruktvalidität (4) berücksichtigt Nachweise aus Vergleichen mit anderen Merkmalen oder Methoden, um zu untersuchen, inwiefern die Testwertergebnisse und andere Merkmale oder Messgrößen einen erwarteten (hohen oder niedrigen) Zusammenhang widerspiegeln (Messick, 1995). Dabei wird durch Korrelationsmuster der eigenen Messung mit anderen Messverfahren zum selben Zielkonstrukt konvergente Evidenz (konvergente Validierung) und durch Korrelationsmuster mit anderen, entfernten Konstrukten diskriminante Evidenz (diskriminante Validierung) gewonnen, wobei erstere möglichst hohe und letztere möglichst geringe Korrelationen

aufweisen sollte (Messick, 1995; Schaper, 2015). Beispielsweise deutet im Sinne der diskriminanten Validierung eine geringe Korrelation zwischen der Testleistung in einem fachdidaktischen Leistungstest und der letzten Deutschnote als Indikator für sprachliche Kompetenz darauf hin, dass sprachliche Fähigkeiten keinen erheblichen Einfluss auf die erfolgreiche Bearbeitung des FDW-Tests haben (Riese et al., 2017).

Der **Aspekt der Verallgemeinerbarkeit** der Konstruktvalidität (5) hat zum Anliegen, dass eine Messung und Bewertung der Kompetenz den Inhalt und die Prozesse eines Konstrukts repräsentativ abdecken sollen. Die Testbewertung sollte nicht nur auf die genutzten Testaufgaben beschränkt sein, sondern das interessierende Konstrukt im weiteren Sinne beschreiben (Messick, 1995). Nachweise zur Verallgemeinerbarkeit legen somit Grenzen für die Aussagekraft der Testbewertung und Testwertinterpretation fest (Messick, 1995). Für Verallgemeinerbarkeit ist eine breite Abdeckung des Konstrukts zur angemessenen Repräsentativität relevant, welche wiederum im Konflikt zu der Tiefe bzw. Genauigkeit der Messung steht (Messick, 1995; Schaper, 2015; siehe oben). Ein „perfektes“ Messverfahren, welches das Konstrukt ausreichend genau und spezifisch erfasst und dieses gleichzeitig breit genug abdeckt, um angemessene Verallgemeinerungen zu erlauben, ist bei zeitlich gebundenen Testsituationen letztendlich nicht zu erreichen (Schaper, 2015).

Der **konsequentielle Aspekt** der Konstruktvalidität (6) umfasst schließlich die Bewertung der möglichen oder tatsächlichen, beabsichtigten oder unbeabsichtigten Folgen durch die Nutzung und Interpretation von Testergebnissen im Hinblick auf Verzerrungen bei der Bewertung sowie auf Ungerechtigkeit bei der Testnutzung (Messick, 1995). Insbesondere sollen keine negativen Konsequenzen für Testteilnehmende entstehen, die aus Quellen der Validitätsbedrohung wie Unterrepräsentation oder konstruktirrelevanter Varianz (siehe oben) hervorgehen (Messick, 1995). Für Kompetenz- und Leistungsmessungen im Bildungsbereich spielt dieser Aspekt eine wichtige Rolle: Der zunehmende Einsatz von Kompetenzmessungen zur Leistungserhebung und -rückmeldung im Bildungssystem macht eine argumentative und empirische Prüfung der von den Messungen in Anspruch genommenen Ziele und Wirkungen umso bedeutsamer (Schaper, 2015). Leistungserhebungen

im Bildungsbereich können auch potenzielle Vorteile für das Lehren und Lernen mitbringen. Nach Messick (1995) sollten daher im Sinne der konsequentiellen Validierung empirische Belege für solche positiven Folgen sowie für das Ausbleiben von negativen Folgen erbracht werden. Der konsequentielle Validitätsaspekt ist jedoch auch umstritten, da sich dieser zum einen schwierig empirisch testen lässt und weniger in die Logik der Konstruktvalidität passt als die anderen Aspekte. Zum anderen fordert die konsequentielle Validierung die Haftung für jegliche spätere Nutzung des Tests und entsprechende Konsequenzen, was praktisch kaum möglich ist (Schaper, 2015).

Die vorgestellten Aspekte der Konstruktvalidität bieten zusammengekommen eine Möglichkeit zur Klärung von Fragen im Validierungsprozess, die zur Rechtfertigung der Interpretation und Verwendung der Testwerte notwendig sind. Der Validierungsprozess ist dabei sowohl empirisch wie theoretisch argumentativ (Messick, 1995). Diese Aspekte sind jedoch nicht als „Checkliste“ zu verstehen, die für jede Testwertinterpretation immer geprüft werden müssen. Kane (2001, 2013) betont, dass die zur Validierung erforderlichen Argumente einer Testwertinterpretation diese sind, die zur Bewertung der Schlussfolgerungen und Annahmen der Interpretation – bestimmt durch den Inhalt der jeweiligen Testwertinterpretation – am relevantesten sind. Für Validitätsargumente ist daher vor allem die Zielsetzung relevant, die mit dem verwendeten Testinstrument verfolgt wird (Kane, 2001, 2013). In der vorliegenden Arbeit werden verschiedene Validierungsstudien zu einem Teil der vorgestellten Validitätsaspekte beschrieben, um den Einsatz eines neu entwickelten Leistungstests zur Erfassung des FDW zum Einsatz digitaler Medien bei Lehramtsstudierenden im Fach Physik zu bewerten. Dabei werden insbesondere die Validitätsaspekte betrachtet, die für die angestrebte Nutzung des Tests bedeutsam sind. Die Auswahl der untersuchten Validitätsaspekte wird in 7.1 beschrieben, die Ergebnisse der einzelnen Validierungsstudien finden sich in Kapitel 9.

In den beiden nachfolgenden Abschnitten werden nun konkrete Beispiele aus der naturwissenschaftsdidaktischen Forschung zu Messverfahren von FDW sowie von TPACK beschrieben. Dabei wird auf das Vorgehen, untersuchte Gütekriterien und insbesondere auf Untersuchungen zur Validität im Rahmen der jeweiligen Testentwicklung und -erprobung eingegangen.

3.4 Messverfahren zum Fachdidaktischen Wissen

FDW ist als bedeutsamer Teilbereich des Professionswissens von (angehenden) Lehrkräften in den letzten Jahrzehnten in einigen Projekten der Lehrkräftebildungsforschung fokussiert worden. In diesem Abschnitt sollen exemplarisch Modelle und Verfahren zur Operationalisierung und Messung dieses Wissensbereichs aus verschiedenen nationalen Forschungsprojekten vorgestellt und erläutert werden; darunter auch das Modell zur Aufgabenentwicklung im Verbundprojekt ProfiLe-P (Kulgemeyer et al., 2012; Riese et al., 2017; 2.2.2), welches die Grundlage für die Aufgabenentwicklung in dieser Arbeit darstellt. Zum Vergleich werden jedoch auch andere Forschungsprojekte zum Professionswissen von (angehenden) Physiklehrkräften sowie deren Ansätze zur Testentwicklung erläutert. Manche dieser Projekte wurden in Folgephasen oder -projekten fortgeführt, deren Forschungsbeiträge auch für die vorliegende Arbeit von Bedeutung sind. In diesem Abschnitt werden jedoch zunächst nur die Projekte oder Projektphasen näher vorgestellt, in denen die Entwicklung und Validierung der Messinstrumente zum FDW erfolgte. Eine Beschreibung der entsprechenden Folgeprojekte sowie insbesondere relevanter Forschungsergebnisse daraus erfolgt in 4.1.1 und 4.1.2.

Zum Teil befassen sich die Forschungsprojekte zum Professionswissen mit der Erfassung dieses Wissens bei praktizierenden Lehrkräften. Im Verbundprojekt ProWiN („Professionswissen von Lehrkräften in den Naturwissenschaften“) wurde unter anderem ein Testinstrument zur Erfassung des FDW im Fach Physik entwickelt (Kirschner, 2013; Tepner et al., 2012). Zur Untersuchung der Validität wurden dabei entwickelte Testaufgaben mit Expert:innen diskutiert (inhaltliche Validierung) und die Testleistungen von Physiklehramtsstudierenden, Fachphysiker:innen sowie Lehrkräften anderer Fächer mit denen der Physiklehrkräfte verglichen (diskriminante Validierung). Dieser Vergleich ergab eine Überlegenheit der Physiklehrkräfte gegenüber allen anderen Gruppen und stellte somit ein Argument für die Konstruktvalidität dar (Kirschner et al., 2016). Im Rahmen des Forschungsverbunds FALKO („Fachspezifische Lehrerkompetenzen“) wurde im physikspezifischen Teilprojekt FALKO-P ebenfalls ein Testinstrument zur Messung des FDW von Physiklehrkräften der Sekundarstufe entwickelt (Schödl, 2017;

Schödl & Göhring, 2017). Im Rahmen der Validierung wurden auch hier Einschätzungen von Expert:innen zur Augenscheinvalidität bzw. Inhaltsvalidität der Testaufgaben eingeholt. Neben Physiklehrkräften wurden weiterhin Physiklehramtsstudierende, Fachphysiker:innen und Lehrkräfte anderer Fächer mit dem Instrument befragt, wobei sich auch hier erwartungskonform jeweils die Überlegenheit der Physiklehrkräfte gegenüber den anderen Gruppen zeigte (Schödl & Göhring, 2017). Da in den beiden Projekten der Fokus auf der Befragung von praktizierenden Physiklehrkräften lag, werden diese Projekte in dieser Arbeit nicht näher erläutert. Es soll hier lediglich herausgestellt werden, dass auch in diesen Projekten bei der Testentwicklung gängige Untersuchungen hinsichtlich verschiedener Validitätsaspekte (3.3.2) vorgenommen wurden. In der vorliegenden Arbeit wird der Fokus auf Studien oder Forschungsprojekte zur Untersuchung des FDW von *angehenden* Physiklehrkräften gelegt, welche auch die Entwicklung dieses Professionswissensbereichs im Lehramtsstudium adressieren.

So beschäftigte sich beispielsweise das Projekt KiL („Messung professioneller Kompetenzen in mathematischen und naturwissenschaftlichen Lehramtsstudiengängen“) mit der Erfassung und Entwicklung des Professionswissens bei Lehramtsstudierenden (Kröger et al., 2013, 2015). Die Entwicklung entsprechender Testinstrumente für das Fach Physik zur Messung der fachbezogenen Professionswissensbereiche Fachwissen und FDW erfolgte im Projekt auf Basis eines dreidimensionalen Itementwicklungsmodells, welches als Dimensionen fachdidaktische Inhalte, Fachinhalte und Wissensarten enthält (Kröger, 2019). Für das FDW werden dabei vier fachdidaktische Inhalte unterschieden: Schülerkognition, Instruktionsstrategien, Curriculum und Assessment. Weiterhin unterscheidet das Modell deklaratives, prozedurales sowie schematisches und strategisches Wissen als Wissensarten sowie acht fachphysikalische Inhaltsbereiche, um das Professionswissen in einer inhaltlichen Breite zu erfassen (Kröger et al., 2013). Die entwickelten fachdidaktischen Aufgaben sind dabei nicht zwingend inhaltsabhängig und einem spezifischen Fachinhalt zuzuordnen, sondern erlauben auch übergreifende Inhalte (Kröger, 2019).

Bei der Testentwicklung wurden verschiedene Validitätsaspekte (3.3.2) berücksichtigt, um möglichst valide Aussagen über die Ausprägung des FDW

bzw. Professionswissens von Physiklehramtsstudierenden machen zu können. Die Aufgabenentwicklung erfolgte unter Mithilfe von projektexternen Fachdidaktiker:innen, welche dafür eine Beschreibung des Itementwicklungsmodells sowie Erläuterungen zur Aufgabenkonstruktion und zu möglichen Aufgabenformaten erhielten. Die entwickelten Aufgaben wurden pilotiert und anschließend im Rahmen eines Meetings mit Expert:innen hinsichtlich ihrer inhaltlichen Validität untersucht und diskutiert. Zur kognitiven Validierung wurden Lehramtsstudierenden in begleiteten Interviews kritische Aufgaben vorgelegt, um unklare und unbekannte Begriffe sowie Lösungsbeschreibungen der Teilnehmenden zu identifizieren (Kröger et al., 2015; Sorge et al., 2019). Nach anschließender Überarbeitung und Aufgabenauswahl bestand das finale Testinstrument für die Hauptuntersuchung im Projekt aus 37 Testaufgaben, davon überwiegend Aufgaben im offenen Antwortformat sowie im Multiple-Choice-Format (Sorge et al., 2019). Die Bewertung der Aufgaben erfolgte durch geschulte Personen mithilfe einer Kodieranleitung. Dabei wurden in der Pilotierung die offenen Aufgaben mit einer Beurteiler:innenübereinstimmung von 98 % zweifach kodiert (Kröger, 2019). In der Hauptuntersuchung wurden 200 Physiklehramtsstudierende befragt, wobei sich eine Skalenreliabilität von $\alpha = 0,74$ ergab (Kröger, 2019).

Das im Projekt KiL entwickelte und erprobte Testinstrument wurde im anschließenden Folgeprojekt KeiLa („Kompetenzentwicklung in mathematischen und naturwissenschaftlichen Lehramtsstudiengängen“) weiterverwendet, in welchem unter anderem die Entwicklung des FDW von Physiklehramtsstudierenden untersucht wurde. Relevante Forschungsergebnisse aus dem Projekt KeiLa werden in 4.1.1 beschrieben.

Das Verbundprojekt ProfiLe-P zielte auf die Entwicklung eines Rahmenmodells zum Professionswissen angehender Physiklehrkräfte sowie der Entwicklung entsprechender Testinstrumente für die Bereiche FDW, Fachwissen und Erklärwissen (im Sinne einer prozeduralen Komponente des FDW) ab (Kulgemeyer et al., 2012). Zur Beschreibung der inneren Struktur des FDW in Physik wurde im Projekt ein differenziertes Modell entwickelt, welches acht fachdidaktische Facetten unterscheidet (Gramzow, 2015; Gramzow et al., 2013; Abbildung 2-4). Zur Operationalisierung des FDW wurde dieses differenzierte Modell zu einem Itementwicklungsmodell reduziert (Abbildung

3-2), um den Umfang eines resultierenden Testinstruments von 90 Minuten nicht zu überschreiten (Gramzow, 2015). Dabei wurden vier fachdidaktische Facetten ausgewählt: *Schülervorstellungen, Instruktionsstrategien, Experimente und Vermittlung eines angemessenen Wissenschaftsverständnisses* sowie *Fachdidaktische Konzepte*. Als physikalischer Inhalt wurde sich in der Reduktion auf den Inhalt Mechanik fokussiert, da zu diesem Inhaltsbereich bereits ab Studienbeginn Lerngelegenheiten bestehen (Riese et al., 2017; Riese et al., 2015). Dieser inhaltliche Fokus wurde im gesamten Verbundprojekt, auch für die Entwicklung der anderen Testinstrumente zum Professionswissen gewählt (Riese et al., 2015). Um unterschiedliche Anforderungsklassen für die Aufgaben beschreiben zu können, wurde im Itementwicklungsmodell eine dritte Dimension der kognitiven Anforderungen ergänzt, die in Anlehnung an Anderson und Krathwohl (2001) die Anforderungen *Reproduzieren, Anwenden* und *Analysieren* enthält. Diese Unterscheidung diente lediglich der Entwicklung von Items in verschiedenen Anforderungsklassen und strebte nicht den Anspruch an, schwierigkeitsbestimmende Klassen oder Kognitionsklassen abzubilden (Gramzow, 2015; Riese et al., 2017). In der Pilotierung der Testaufgaben zeigte sich jedoch, dass die Aufgaben der kognitiven Anforderung *Analysieren* für die Teilnehmenden signifikant schwieriger waren als die der anderen beiden Anforderungen (Gramzow, 2015). Das Itementwicklungsmodell ist mit seinen drei Dimensionen in Abbildung 3-2 dargestellt und bildet die Grundlage für das in der vorliegenden Arbeit verwendete Itementwicklungsmodell (8.2).

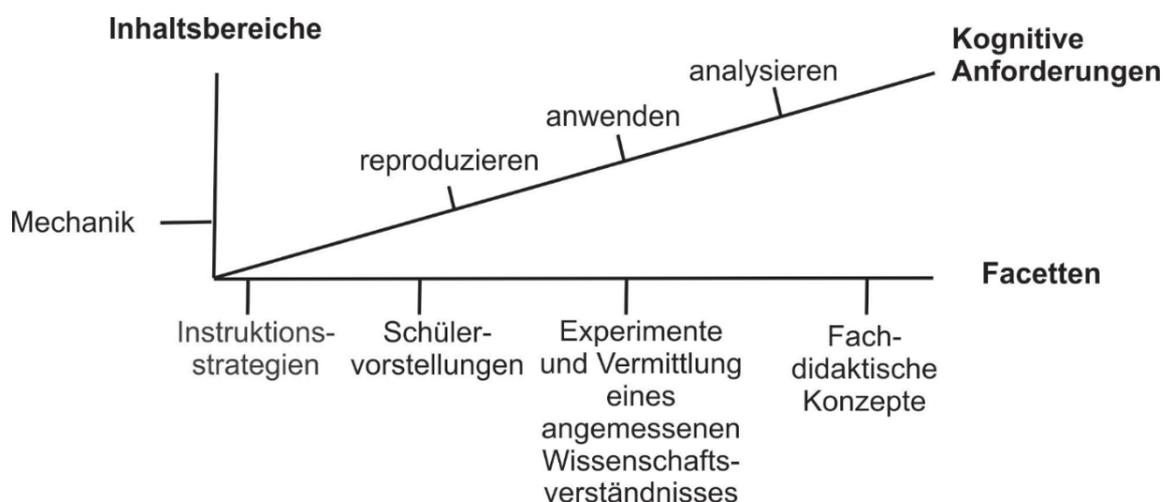


Abbildung 3-2: Itementwicklungsmodell für das FDW in Physik im Projekt ProfiLe-P (Gramzow, 2015, S. 104)

Gemäß einer deduktiven Testentwicklung wurden im Projekt die Facetten und kognitiven Anforderungen aus dem Modell ausdifferenziert, um mögliche Kerninhalte für einzelne Aufgaben herauszuarbeiten. Anknüpfend wurden zielgerichtet Aufgaben entwickelt, die bestimmte Kerninhalte adressieren und systematisch im Hinblick der kognitiven Anforderungen und fachdidaktischen Facetten variieren (Gramzow, 2015; Riese et al., 2015). Für die Pilotierung konnten somit 50 Aufgaben (überwiegend offenen Antwortformats und teilweise im Multiple-Choice-Format) entwickelt werden (Gramzow, 2015; Riese et al., 2015). Im Rahmen der Pilotierung wurden auch in diesem Projekt Untersuchungen zu verschiedenen Validitätsaspekten (3.3.2) durchgeführt, um Argumente zur Legitimation einer Testwertinterpretation zu sammeln, die auf die Erklärung von Leistungen abzielt (Riese et al., 2015). Zur Inhaltsvalidierung erfolgte eine Einschätzung der curricularen Passung der Testaufgaben über die Befragung der Lehrenden der Physikdidaktik an den im Verbundprojekt beteiligten Hochschulen, wobei sich im Mittel eine gute Passung ergab (Gramzow, 2015; Riese et al., 2015). Zur kognitiven Validierung (von den Autor:innen ebenfalls unter Inhaltsvalidierung gefasst) wurden 15 Lehramtsstudierende im Lauten Denken befragt, um Rückschlüsse darauf zu bekommen, welches Wissen zur Lösung der Testaufgaben herangezogen wird. Auf diese Weise konnten Aufgaben oder Antwortoptionen ausgeschlossen werden, die nicht hauptsächlich unter Nutzung von FDW beantwortet wurden, sondern beispielsweise ausschließlich mit physikalischem Fachwissen (Gramzow, 2015; Riese et al., 2015). Weiterhin wurde ebenfalls im Sinne der Inhaltsvalidierung die Passung der entwickelten Testaufgaben zu dem zugrundeliegenden Itementwicklungsmodell geprüft. Physikdidaktiker:innen ordneten dazu als Expert:innen die Aufgaben den kognitiven Anforderungen und fachdidaktischen Facetten zu, wobei sich jeweils zufriedenstellende Übereinstimmungen (Cohens $\kappa = 0,63$ bis $\kappa = 0,89$) ergaben (Gramzow, 2015; Riese et al., 2015).

Anhand dieser Untersuchungen zur Validität sowie weiteren Ergebnissen der Pilotierung wurden schließlich 26 Aufgaben für das Testheft zur Haupterhebung im Verbundprojekt ausgewählt, welches auf eine Bearbeitungszeit von ca. 75 Minuten (inklusive Demografie) ausgelegt ist (Gramzow, 2015; Riese et al., 2015). Zur Untersuchung der Auswertungsobjektivität wurden

die überwiegend offenen Aufgaben bei einem zufällig ausgewählten Teil der Testhefte von zwei unabhängigen Rater:innen bewertet, bei starken Abweichungen Fehlerquellen identifiziert und gegebenenfalls das Kodiermanual überarbeitet. Bei anschließender Rekodierung des gesamten Datensatzes (294 Physiklehramtsstudierende) ergab sich gute bis sehr gute Beurteilendenübereinstimmungen (Cohens $\kappa = 0,70$ bis $\kappa = 0,93$) in den einzelnen Aufgaben (Riese et al., 2017). Die EAP-Reliabilität² für den Gesamttest in der Hauptstudie war mit 0,79 ebenfalls zufriedenstellend. Zur Konstruktvalidierung wurde in der Hauptstudie weiterhin untersucht, inwiefern sich die angenommene Modellstruktur der vier fachdidaktischen Facetten empirisch unterstützen lässt, indem ein eindimensionales Raschmodell (IRT-Modell) mit einem vierdimensionalen Modell (gemäß den Facetten) hinsichtlich der Modellpassung zu den erhobenen Daten verglichen wurde. Dabei zeigte sich, dass für die gesamte Stichprobe kein Vorteil des vierdimensionalen Modells erkennbar wurde, aber bei der Eingrenzung auf fortgeschrittene Studierende ab dem dritten Fachsemester dieses eine bessere Modellpassung aufwies als das eindimensionale Modell. Dieses Ergebnis spricht dafür, dass sich Wissensstrukturen des FDW erst ab einer gewissen Basis ausbilden und empirisch auffindbar werden (Riese et al., 2017). Die EAP-Reliabilitäten dieser einzelnen Teilskalen lagen für die befragte Stichprobe im Bereich zwischen 0,55 und 0,72 (Riese et al., 2017). Das Verbundprojekt wurde unter dem Namen *ProfiLe-P+* in einer zweiten Förderphase fortgeführt und dabei unter anderem der beschriebene Leistungstest zum FDW in Physik genutzt, um die Entwicklung dieses Professionswissensbereichs im Verlauf des Lehramtsstudiums im Längsschnitt zu untersuchen (Vogelsang et al., 2019). Ergebnisse aus dieser zweiten Projektphase werden in 4.1.1 dargestellt.

In der Arbeit von Joswig-Käfer (2024) steht ebenfalls das FDW von Lehramtsstudierenden im Fach Physik im Fokus. Zur Messung des FDW wurde in diesem Projekt das zuvor beschriebene Testinstrument aus dem Projekt *ProfiLe-P* verwendet und um Aufgaben aus den Inhaltsbereichen Elektrizitätslehre und Optik erweitert (Joswig & Riese, 2019). Diese Testaufgaben zur in-

² Reliabilitätskoeffizient für die Analyse mittels IRT (3.3.1)

haltlichen Erweiterung wurden im Sinne der Verortung im Itementwicklungsmodell (Abbildung 3-2) strukturgleich zu den vorliegenden Testaufgaben im Inhaltsbereich Mechanik nach Gramzow (2015) entwickelt und anschließend pilotiert. Das erweiterte Testinstrument enthält 41 (offene und teilweise geschlossene) Testaufgaben bei einer Bearbeitungsdauer von ca. 105 Minuten (Joswig & Riese, 2019).

Die Bewertung der offenen Aufgaben erfolgte auch hier mit einem Kodiermanual, wobei auf Basis des existierenden Manuals adaptierte Kodiermanuale zu den Testaufgaben der Inhaltsbereiche Elektrizitätslehre und Optik entwickelt wurden (Joswig-Käfer, 2024). Eine Doppelkodierung aller Testhefte von zwei unabhängigen Rater:innen zur Gewährleistung der Auswertungsobjektivität ergab eine sehr gute Übereinstimmung von $\kappa = 0,85$ (Joswig-Käfer, 2024). Die Reliabilität des erweiterten Testinstruments (bezogen auf den Gesamttest) in der Hauptbefragung von $N = 38$ Studierenden (4.1.1) war ebenfalls zufriedenstellend (Cronbachs $\alpha = 0,72$). Bei Betrachtung einzelner Testteile ließ sich nur für die Testaufgaben zur Elektrizitätslehre eine zufriedenstellende Reliabilität feststellen ($\alpha = 0,70$), für die Testteile zur Mechanik und Optik ($\alpha < 0,50$) jedoch nicht (Joswig & Riese, 2019; Joswig-Käfer, 2024). In Bezug auf die Inhaltsvalidierung konnte sich auf die im Projekt Profile-P vorgenommenen Untersuchungen zur curricularen Passung und inhaltlichen Validität (Gramzow, 2015; Riese et al., 2015; siehe oben) der strukturgleichen Testaufgaben nach Gramzow (2015) gestützt werden. Zusätzlich wurden die neu entwickelten Aufgaben in Studien zum Lauten Denken mit Studierenden im Hinblick auf Verständnisschwierigkeiten sowie Hinweise zur Inhaltsvalidität erprobt (Joswig-Käfer, 2024). Mithilfe dieses erweiterten Leistungstests wurde in der Hauptstudie des Projekts die Entwicklung des FDW bei Masterstudierenden im Lehramt Physik über ein Semester untersucht. Weiterhin wurden qualitative Interviews geführt, um lernförderliche Elemente für die Entwicklung des FDW zu identifizieren (Joswig-Käfer, 2024; Joswig-Käfer & Riese, 2021). Die Ergebnisse dieser quantitativen und qualitativen Untersuchungen werden in 4.1.1 und 4.1.2 erläutert.

Zusammenfassend zeigt sich, dass die vorgestellten Studien zur Messung und Untersuchung des FDW in Physik verschiedene Schwerpunkte in ihrem Vorgehen und Untersuchungszielen aufweisen: in den Projekten KiL (Kröger

et al., 2015) und ProfiLe-P (Kulgemeyer et al., 2012; Riese et al., 2015) sowie bei Joswig-Käfer (2024) wurde das FDW von Lehramtsstudierenden untersucht, während die Projekte ProwiN (Tepner et al., 2012) und FALKO (Schödl & Göhring, 2017) das FDW praktizierender Lehrkräfte fokussierten. Der Erfassung des FDW in inhaltlicher Breite (KiL) steht die Fokussierung auf wenige Inhaltsbereiche (FALKO, ProwiN und Joswig-Käfer, 2024) oder einen Inhaltsbereich (ProfiLe-P) gegenüber. Zur Differenzierung wurden verschiedene Facetten des FDW in den Mess- oder Itementwicklungsmodellen unterschieden, wobei stets Schülervorstellungen bzw. Schülerkognition (FALKO, KiL, ProwiN, ProfiLe-P und Joswig-Käfer, 2024) und meist auch Instruktionsstrategien berücksichtigt wurden (KiL, ProfiLe-P und Joswig-Käfer, 2024). Weitere Facetten, die in den vorgestellten Messverfahren zum FDW operationalisiert wurden, sind Curriculum und Assessment (KiL) sowie Experimente und Fachdidaktische Konzepte (ProfiLe-P). Gleichzeitig wurden hinsichtlich der Testentwicklung sowie der Bewertung der Testgüte und Validität in den beschriebenen Projekten oftmals ähnliche Argumente verfolgt. Dabei wurde auf qualitative Befragungen zurückgegriffen – beispielsweise von Expert:innen zur Prüfung der Eignung der Aufgaben (FALKO, KiL, ProfiLe-P und ProwiN) oder von Lehramtsstudierenden zur Untersuchung der genutzten Wissensbestände sowie zur Identifizierung möglicher Verständnisschwierigkeiten bei der Aufgabenbearbeitung (KiL, ProfiLe-P und Joswig-Käfer, 2024). Durch statistische Analysen der Testergebnisse wurden Argumente zur Konstruktvalidität gewonnen, indem beispielsweise Vergleiche der Testleistungen unterschiedlicher Gruppen vorgenommen (FALKO, ProfiLe-P und ProwiN) oder Zusammenhänge mit anderen Wissensbereichen des Professionswissens (KiL, ProfiLe-P, ProwiN) untersucht wurden. Das Projekt ProfiLe-P analysierte erstmals auch die interne Struktur des FDW in Physik, indem die modellierten Facetten des FDW auf ihre empirische Trennbarkeit untersucht wurden (Riese et al., 2017).

Im folgenden Abschnitt werden nun analog exemplarisch Verfahren zur Messung des TPACK, d.h. des technologiebezogenen fachdidaktischen Wissens (2.3.1), vorgestellt – insbesondere in den naturwissenschaftlichen Fächern.

3.5 Messverfahren zum TPACK-Modell

Ausgehend von der theoretischen Modellierung des TPACK-Rahmens (2.3.1) nach Mishra und Koehler (2006) gab es in den vergangenen Jahren viele Studien z. B. hinsichtlich der Erfassung der sieben Wissensbereiche sowie ihrer Zusammenhänge und Trennbarkeit zur empirischen Prüfung des Modells (z. B. Kopcha et al., 2014; Scherer et al., 2017). Dabei wurden unterschiedliche Ansätze zur Erfassung der Wissensbereiche im TPACK-Modell verfolgt, welche im Folgenden anhand von Beispielen veranschaulicht werden. Aufgrund der Fokussierung dieser Arbeit auf das FDW zum Einsatz digitaler Medien und dessen Ähnlichkeit zum Wissensbereich TPACK des gleichnamigen Modells (2.3.1) wird hier insbesondere der Schwerpunkt auf die Messung dieses zentralen Wissensbereichs TPACK gelegt.

In ihrer Literaturübersicht untersucht Willermark (2018) empirische Studien, in denen das TPACK-Modell zur Erfassung jenes Wissens bei angehenden oder praktizierenden Lehrkräften genutzt wurde. Dabei unterscheidet sie zwei grundsätzliche Ansätze: die Messung über Selbstberichte bzw. Selbsteinschätzungen (zu allgemeinem, spezifischem oder auf Erfahrung basierendem TPACK) sowie die Messung über die Bewertung von Leistungen im unterrichtlichen Handeln (Planung, Durchführung oder Bewertung von Unterricht). Diese Ansätze unterscheiden sich demnach in ihrem Verfahren zur Erfassung von TPACK (subjektiv gegenüber objektiv; 3.2.1) und lassen sich nach Willermark (2018) auf einer Skala gemäß ihrem Verständnis von TPACK als Wissen (gemessen durch Selbsteinschätzungen) bis hin zu TPACK als Performanz (z. B. bei der Durchführung von Unterricht in authentischen Situationen) einordnen. Willermark bezeichnet Letzteres als *competence* und verwendet den Kompetenzbegriff damit im Sinne der Handlung selbst, wohingegen in der vorliegenden Arbeit Kompetenz als kognitive Disposition für unterrichtliches Handeln verstanden wird (2.1). Neben diesen Unterschieden im Begriffsverständnis wird außerdem deutlich, dass der in der vorliegenden Arbeit gewählte Ansatz eines standardisierten Leistungs- oder Kompetenztests – wie er zur Messung von FDW etabliert ist (3.4) – gar nicht in Willermarks (2018) Übersicht auftritt. Als häufigster Zugang zur Messung von TPACK stellte sich in der Übersicht die Nutzung von Selbsteinschätzungen heraus (72 %), insbesondere durch Fragebögen zum allgemeinen TPACK

(52 %), in denen die Teilnehmenden Aussagen allgemeiner Natur und ohne bestimmten Kontext auf einer Ratingskala dahingehend beantworten, wie sie ihr Wissen dazu einschätzen (Willermark, 2018). Die Bewertung der Leistungen im unterrichtlichen Handeln trat deutlich seltener auf (28 %), davon am häufigsten mit Bezug zur Planung von Unterricht (16 %).

Eine weitere Literaturübersicht von Wang et al. (2018) untersuchte empirische Studien, die die Entwicklung des TPACK von angehenden Lehrkräften analysierten. Hier wurden fünf verschiedene Messmethoden unterschieden: Selbsteinschätzungen (über Ratingskalen), offene Fragebögen (mit schriftlichen Antworten), Leistungsbewertungen (z. B. Raster zur Bewertung von Unterrichtsplanungen), Interviews und Beobachtungen (z. B. über Unterrichtsvideografien). Auch hier stellten sich Selbsteinschätzungen als am häufigsten genutzte Messmethode heraus (Wang et al., 2018), allerdings weniger deutlich als in Willermarks Übersicht (48 %). Leistungsbewertungen traten ebenfalls häufiger auf (40 %) und beinhalteten z. B. die Nutzung von Produkten wie Unterrichtsplanungen oder problembasierten und situationsspezifischen Aufgabenstellungen (Wang et al., 2018). Beide vorgestellten Literaturübersichten stellten jedoch auch heraus, dass einige Studien mehrere Methoden zur Erfassung des TPACK verwendeten, um die Ergebnisse zu triangulieren (Wang et al., 2018; Willermark, 2018).

Im Folgenden werden nun einige Instrumente vorgestellt, die den Kategorien Selbsteinschätzungen und Leistungsbewertungen zuzuordnen sind. Dabei werden Selbsteinschätzungen berücksichtigt, da sie derzeit das am häufigsten verwendete Erhebungsverfahren für die Messung von TPACK darstellen (Wang et al., 2018; Willermark, 2018). Leistungsbewertungen werden einbezogen, da sie den etablierten Verfahren für die Messung von FDW (3.4) am ähnlichsten sind und somit auch dem Ansatz zur Messung des FDW zum Einsatz digitaler Medien in der vorliegenden Arbeit. Weitere Verfahren zur Erfassung von TPACK wie (qualitative) Interviews, Fragebögen oder Beobachtung von Unterrichtshandeln sind weniger vergleichbar mit dem Ansatz eines standardisierten Tests zur Messung von Kompetenz als Disposition in dieser Arbeit. Daher und da diese Verfahren zudem seltener vorkommen, werden sie hier nicht näher thematisiert. Die meisten der nachfolgend beschrieben

Instrumente werden (unter anderem) im deutschsprachigen Raum in der naturwissenschaftsdidaktischen Forschung zum TPACK von angehenden Lehrkräften genutzt. Diese Studien haben besondere Relevanz für die vorliegende Arbeit und werden zum Teil in 4.2.1 im Hinblick auf ihre Ergebnisse zur Entwicklung des TPACK in naturwissenschaftlichen Lehramtsstudiengängen erneut aufgegriffen.

3.5.1 Selbsteinschätzungsinstrumente zu TPACK

Ein häufig verwendetes Instrument zur Erfassung des selbsteingeschätzten TPACK angehender Lehrkräfte wurde von Schmidt et al. (2009) entwickelt (Willermark, 2018). Ursprünglich konzipiert wurde der Fragebogen für Lehramtsstudierende im Grundschulbereich oder der frühkindlichen Erziehung (Schmidt et al., 2009). Er wurde bislang in einigen weiteren Studien verwendet und auf den jeweiligen Untersuchungskontext angepasst (Wang et al., 2018; Willermark, 2018). Zur Entwicklung des Fragebogens wurden zunächst passende Literatur und bestehende Instrumente in diesem Forschungsfeld gesichtet und insbesondere für Informationen zum Erhebungsansatz herangezogen (Schmidt et al., 2009). Auf dieser Basis entwickelte Items wurden zur inhaltlichen Validierung drei Expert:innen zur Bewertung der Items hinsichtlich ihrer Passung zu einem TPACK-Bereich vorgelegt und anschließend überarbeitet. Eine erste Version des Fragebogens aus 75 Items wurde pilotiert, im Hinblick auf die Reliabilitäten der Subskalen untersucht und anschließend um problematische Items reduziert. Der finale Fragebogen besteht aus 47 Items in den sieben TPACK-Bereichen (davon 8 Items zu TPACK) und wies gute Kennwerte für die Reliabilität einzelner Skalen auf ($\alpha > 0,72$ für alle Skalen; $\alpha_{\text{TPACK}} = 0,92$). Die Items sind als Aussagen wie z. B. „*Ich kann Technologien auswählen, die den Inhalt einer Unterrichtsstunde bereichern.*“³ (TPACK) formuliert, worauf die Teilnehmenden auf einer fünfstufigen Likert-Skala ihre Zustimmung äußern (Schmidt et al., 2009). Nach Willermarks (2018) Kategorisierung handelt es sich bei dem Fragebogen um ein

³ Originalitem: *“I can choose technologies that enhance the content for a lesson.”*

Selbsteinschätzungsinstrument zum allgemeinen TPACK, da es die Einschätzung unspezifischer Aussagen zu den eigenen Fähigkeiten in den TPACK-Bereichen situations- und kontextunabhängig erfordert.

Das TPACK-Instrument nach Schmidt et al. (2009) ist von einigen Autoren adaptiert und weiterentwickelt worden (Wang et al., 2018; Willermark, 2018). So nutzten beispielsweise Chai et al. (2011) für die Entwicklung ihres Erhebungsinstruments einen großen Teil dieser Items (alle TPACK-Bereiche bis auf PK bzw. pädagogisches Wissen) und ergänzten diese um 13 alternative Items zum pedagogical knowledge for meaningful learning (PKML), um den pädagogischen Fokus ihres untersuchten Technologiekurses zu adressieren (Chai et al., 2011). Weiterhin wurden fünf Items zu webbasierten Kompetenzen im Bereich TK ergänzt und das Antwortformat auf eine siebenstufige Likert-Skala adaptiert. Dieser adaptierte Fragebogen enthält final 46 Items (Chai et al., 2011). 2013 wurde dieses Instrument erneut weiterentwickelt sowie auf den entsprechenden Studienkontext angepasst und enthält in dieser Version 36 Items (Chai et al., 2013b). Die Instrumente von Chai et al. (2011) bzw. Chai et al. (2013b) wurden ebenfalls in anderen Studien zur Untersuchung von TPACK genutzt (Willermark, 2018). So verwendeten im deutschsprachigen Raum beispielsweise Stinken-Rösner et al. (2023) das Instrument von Chai et al. (2013b), um die Entwicklung der selbst eingeschätzten TPACK-Bereiche von Lehramtsstudierenden der naturwissenschaftlichen Fächer im Verlauf von zwei Lerngelegenheiten zu untersuchen (4.2.1). Dazu wurden die Originalitems übersetzt und für den naturwissenschaftlichen Unterricht adaptiert. Ein Beispiel zum Bereich TPACK lautet: *„Ich kann Lernumgebungen gestalten, die Schüler*innen dabei unterstützen naturwissenschaftliche Inhalte auf verschiedene Weise mit digitalen Medien darzustellen.“* (Stinken-Rösner, 2021a, S. 182).

Die verwendeten Selbsteinschätzungsinstrumente wie solche basierend auf dem Fragebogen nach Schmidt et al. (2009) beinhalten in der Regel vergleichsweise generische sowie fach- und kontextunabhängige Items (Willermark, 2018) – sofern sie nicht konkret z. B. im Hinblick auf die Spezifität des untersuchten Fachs adaptiert werden. In vielen Studien wird jedoch entwe-

der kein Fach spezifiziert oder es werden zwar (angehende) Lehrkräfte verschiedener Fächer befragt, aber nicht miteinander verglichen (Willermark, 2018).

Es gibt jedoch auch andere Ansätze zur Erfassung von TPACK, in denen konkret die naturwissenschaftlichen Fächer adressiert werden. So wurde beispielsweise zum Orientierungsrahmen DiKoLAN, der Kompetenzerwartungen für das Lehramt in den Naturwissenschaften für die technologiebezogenen TPACK-Bereiche formuliert (2.3.1), ein entsprechendes Selbsteinschätzungsinstrument entwickelt (Kotzebue et al., 2021). Für die sieben Kompetenzbereiche im DiKoLAN-Orientierungsrahmen⁴ wurden entsprechende Items für die Kompetenzerwartungen zugehörig zu den vier Schwerpunkten des Rahmens (2.3.1) *Unterricht* (TPACK), *Methodik und Digitalität* (TPK), *fachwissenschaftlicher Kontext* (TCK) und *spezielle Technik* (TK) erstellt. Auf einer achtstufigen Likert-Skala geben Teilnehmende an, inwiefern sie den Aussagen bzgl. ihrer Fähigkeit in der entsprechenden Kompetenzbeschreibung zustimmen (Henne et al., 2022; Kotzebue et al., 2021). Das Instrument wurde bei Biologielehramtsstudierenden an zwei deutschen Universitäten für zwei der sieben Kompetenzbereiche (Bereich Präsentation (P) bei 118 Teilnehmenden; Bereich Recherche und Bewertung (RB) bei 90 Teilnehmenden) eingesetzt. Die Reliabilitäten für die acht Einzelskalen je betrachteten TPACK- und Kompetenzbereich ergaben gute Kennwerte ($\alpha_{\text{TPACK_P}} = 0,91$ und $\alpha_{\text{TPACK_RB}} = 0,94$; Kotzebue et al., 2021). Darüber hinaus wurden jedoch bislang keine Ergebnisse zur Betrachtung der Güte und Validität dieses Selbsteinschätzungsinstruments berichtet.

Die bis hierhin vorgestellten Instrumente zur Erfassung des TPACK beruhen auf Selbsteinschätzungen, in denen die Teilnehmenden vorgegebene Aussagen dahingehend einschätzen, ob sie über das entsprechende Wissen oder die Fähigkeiten verfügen. Auch wenn diese Erhebungsmethode zur Erfassung von TPACK sehr verbreitet ist (Scherer et al., 2017; Wang et al., 2018; Willermark, 2018), stehen solche Instrumente beruhend auf Selbsteinschätzungen zunehmend hinsichtlich ihrer Validität in Kritik (Akyuz, 2018; Kopcha

⁴ Die sieben Kompetenzbereiche des DiKoLAN-Orientierungsrahmens sind *Dokumentation, Präsentation, Kommunikation und Kollaboration, Recherche und Bewertung, Messwert- und Datenerfassung, Datenverarbeitung* und *Simulation und Modellierung* (2.3.1).

et al., 2014; Kotzebue, 2022a, 2022b; Lachner et al., 2019). Eine akkurate Einschätzung der eigenen Fähigkeiten in solchen Instrumenten ist schwierig, da Teilnehmende sich ihres fehlenden Wissens gar nicht unbedingt bewusst sein müssen und somit ihr Wissen oder ihre Fähigkeiten überschätzen können. Besonders schwierig ist eine realistische Einschätzung, wenn die zu bewertenden Aussagen im Instrument sehr allgemein formuliert sind und einen großen Interpretationsspielraum erlauben – wie es bei einigen verwendeten Selbsteinschätzungsinstrumenten zur Erfassung von TPACK der Fall ist (Willermark, 2018). Kritisiert wird daher, dass die Selbsteinschätzungen der Teilnehmenden keinen Indikator für tatsächliches Wissen darstellen und eher Selbstwahrnehmung oder Selbstwirksamkeit erfassen (Backfisch et al., 2020b; Kotzebue, 2022a; Lachner et al., 2019; Scherer et al., 2017; Willermark, 2018). In einigen Studien werden dementsprechend zwar weiterhin solche Selbsteinschätzungsinstrumente genutzt, aber dabei eindeutig benannt, dass damit die TPACK-Selbstwirksamkeit und nicht das Wissen untersucht wird (z. B. Bilici et al., 2013; Henne et al., 2022; Zimmermann, 2021).

3.5.2 Leistungsbasierte Messungen zu TPACK

Diese Kritik zu Selbsteinschätzungsinstrumenten adressierend, gibt es verschiedene Ansätze, TPACK als Wissen möglichst objektiv und proximal (3.2.1) über Leistungsbewertungen zu messen, in denen die Leistung der Teilnehmenden bei vorgegebenen Aufgaben direkt untersucht wird. Dazu zählen z. B. Leistungstests mit szenario- und problembasierten Aufgaben oder das Erstellen von Artefakten wie Portfolios oder Reflexionsjournale (Koehler et al., 2012; Wang et al., 2018). In vielen Studien wird weiterhin auf Unterrichtsplanungen als erstellte Produkte der Teilnehmenden zurückgegriffen (Wang et al., 2018; Willermark, 2018).

Harris et al. (2010) entwickelten beispielsweise ein Raster zur Bewertung von TPACK auf Basis von Unterrichtsplanungen angehender Lehrkräfte. In dem Raster werden vier Kriterien zur Beurteilung herangezogen, die sich drei der Wissensbereiche des TPACK-Modells zuordnen lassen. Die Kriterien sind Lehrplanziele und Technologien (TCK), Instruktionsstrategien und Technologien (TPK), Technologieauswahl (TPACK) und Passung (TPACK) und erlauben jeweils eine Bewertung auf einem vierstufigen Ranking (Harris et al., 2010).

Zur Erprobung des Instruments wurden Expert:innen zur Bewertung der Augenschein- und Konstruktvalidität des Rasters befragt sowie die Analyse von Unterrichtsplanungen mithilfe des Rasters im Hinblick auf die Beurteilendenübereinstimmung (Koeffizient von 0,86) und auf die Reliabilität ($\alpha = 0,91$) untersucht (Harris et al., 2010). Weitere Validitätsbetrachtungen wurden in dieser Studie nicht berichtet. In einer Studie von Kopcha et al. (2014) wurde das Raster jedoch in Kombination mit einer Selbsteinschätzung zum TPACK eingesetzt und die Ergebnisse hinsichtlich diskriminanter und konvergenter Validität verglichen (3.5.3).

Ebenfalls nutzte Akyuz (2018) Unterrichtsplanungen zur Bewertung von TPACK in seiner Studie mit angehenden Mathematiklehrkräften, die einen technologiebezogenen Mathekurs zur Nutzung von dynamischer Geometrie-Software (DGS) absolvierten. Die teilnehmenden Studierenden erstellten mathematische Aktivitäten mit der DGS sowie zugehörige Unterrichtsplanungen und Arbeitsmaterialien, welche mithilfe eines Kriterienrasters⁵ bewertet wurden. Dieses Raster berücksichtigt die sieben TPACK-Bereiche mit jeweils drei Items und beinhaltet somit 21 Items zur Bewertung der mathematischen Unterrichtsaktivität mit der DGS im Hinblick auf das TPACK der Teilnehmenden (Akyuz, 2018). Beispielhafte Items lauten „*Potenzielle technische Antworten und/oder Schwierigkeiten der Schülerinnen und Schüler werden antizipiert.*“ (TPACK) oder „*Die Unterrichtsplanung/die Aktivität enthält Fragen, die dazu auffordern, ihre Lösungen mit der DGS zu erklären.*“ (TPK)⁶ (Akyuz, 2018). In der Studie wurden in über zehn Durchgängen insgesamt 138 Lehramtsstudierende befragt, für die in der Regel jeweils die Unterrichtsplanung, die DGS-Aktivität sowie ein Arbeitsblatt zur Aktivität zu bewerten waren, sodass zur Datenanalyse 1458 Dokumente mithilfe des entwickelten Rasters bewertet wurden (Akyuz, 2018). Ein Teil der Daten wurde durch eine externe

⁵ Der Autor spricht zwar von einem *assessment instrument*, jedoch wird im Verständnis dieser Arbeit dieses Verfahren als Kriterienraster bezeichnet und nicht als Testinstrument im Sinne eines Leistungstests (3.2.1), weil damit Produkte der Teilnehmenden anhand von Kriterien bewertet werden und keine Testaufgaben gestellt werden, die die Teilnehmenden beantworten müssen.

⁶ Originalitems: „*Student's potential technological answers and/or difficulties are anticipated*“ (TPACK); „*The lesson plan/activity sheet includes questions prompted to explain their solutions with DGS*“ (TPK)

Person mit einer Übereinstimmung von 80 % doppelkodiert, woraufhin Diskrepanzen gemeinsam besprochen und Definitionen der Items überarbeitet wurden (Akyuz, 2018). Aufgrund dieser zeitaufwändigen Auswertung empfiehlt Akyuz (2018) den Einsatz des Bewertungsinstruments insbesondere für kleine Untersuchungsgruppen bis zu 20 Personen. Für passende Anwendungsszenarien, in denen ähnliche Produkte zur Unterrichtsplanung bewertet werden, stellt das Instrument jedoch einen kontextspezifischen Ansatz für die Erfassung von TPACK dar, welcher nach Akyuz (2018) auch Adaption auf andere Technologien erlaubt.

Weiterhin wurden auch in naturwissenschaftsdidaktischen Studien im deutschsprachigen Raum Unterrichtsplanungen zur Erfassung einzelner TPACK-Bereiche genutzt (Stinken-Rösner et al., 2023; Zimmermann, 2021). In der Studie von Stinken-Rösner et al. (2023) wurden beispielsweise Unterrichtsplanungen von Studierendengruppen im Lehramt in naturwissenschaftlichen Fächern im Hinblick auf die Quantität und Qualität des Medieneinsatzes bewertet. Die Qualitätsanalyse berücksichtigte dabei die Bewertung der Technologieintegration anhand des SAMR-Modells (Puentedura, 2006, 2012) auf den vier Stufen *Ersetzung*, *Erweiterung*, *Änderung* und *Neubelegung* (2.3.3) als Indikator für TPK sowie die kognitive Aktivierung der Lernenden in den vier Stufen passiv, aktiv, konstruktiv und interaktiv des ICAP-Modells nach Chi und Wylie (2014) als Indikator für PCK bzw. FDW. Als Indikator für TPACK wurde der Mittelwert aus den Werten für TPK und PCK gebildet (Stinken-Rösner et al., 2023).

In der Arbeit von Zimmermann (2021) wurde ein Seminar zur Förderung digitalisierungsbezogener Kompetenzen von Chemielehramtsstudierenden entwickelt und evaluiert (4.2.1). Dabei wurde u. a. die Fähigkeit der Seminarteilnehmenden, digitale Werkzeuge in die Unterrichtsplanung zu integrieren, analysiert, indem sie eine Aufgabe zur Planung einer Unterrichtsstunde bearbeiteten. Anknüpfend wurden leitfadenorientierte Interviews mit den Studierenden zur Erläuterung und Begründung ihrer Planungsentscheidungen geführt. Die Unterrichtsplanungen wurden mithilfe eines Kodiermanuals analysiert, welches sieben zentrale Unterrichtsschritte berücksichtigt: Unterrichtseinstieg, Hypothesengenerierung, Erarbeitung, Diskussion der Lernprodukte, Sicherung und Vernetzung, Transfer und Anwendung sowie

Berücksichtigung der Heterogenität (Zimmermann, 2021). Für jeden Unterrichtsschritt wird eine Bewertung für die vier TPACK-Bereiche PCK, TK, TCK und TPK auf einer fünfstufigen Skala vergeben, wobei das Kodiermanual jeweils Indikatoren zur Bewertung der insgesamt 28 Variablen vorgibt. Sowohl die Planungsskizze als auch das zugehörige Interviewtranskript werden bei der Bewertung einbezogen (Zimmermann, 2021). Eine Doppelkodierung eines Teils der Daten mithilfe des Kodiermanuals ergab eine gute Übereinstimmung mit Koeffizienten über 0,90 für alle Variablen PCK, TK, TCK und TPK. Um eine angemessene Validität zu gewährleisten, wurden das Kodiermanual sowie das Testinstrument zur Planungsaufgabe literaturbasiert und mithilfe begleitender Expert:innenrunden entwickelt sowie in einer Präpilotierung vor dem Einsatz in Vor- und Hauptuntersuchung der Arbeit erprobt (Zimmermann, 2021). Dieser Ansatz ermöglichte einen facettenreichen Einblick hinsichtlich der Kompetenzen der Studierenden, erforderte jedoch viel Aufwand in der Testentwicklung, Datenerhebung und Auswertung (Zimmermann, 2021).

Neben der Analyse von Unterrichtsplanungen gibt es Studien mit weiteren proximalen Ansätzen zur Erfassung von TPACK, in welchem fachspezifische Leistungstests entwickelt werden (z. B. Kotzebue, 2022b; Max et al., 2020). So entwickelten Max et al. (2020) einen (biologiespezifischen) vignettenbasierten Leistungstest zur Erfassung von TPACK auf zwei Ebenen: als deklaratives Wissen über TPACK sowie als Anwendung von TPACK auf eine Unterrichtssituation. Die Entwicklung der Testaufgaben erfolgte auf Basis einer Literaturrecherche zu Einsatzszenarien für digitale Technologien im naturwissenschaftlichen und insbesondere im Biologieunterricht. Daraus wurden unterschiedliche fachdidaktische, kognitionspsychologische und mediendidaktische Kriterien für die Bewertung der TPACK-Bereiche PCK, TPK und TPACK ausgewählt (Max et al., 2020). Für die Dimension des PCK bzw. FDW zählen darunter kognitive Aktivierung und inhaltliche Strukturierung als Qualitätskriterien für naturwissenschaftlichen Unterricht (Steffensky & Neuhaus, 2018) sowie die CLT (Sweller, 1994; 2.3.3). Für TPK waren es das SAMR-Modell (Puentedura, 2006, 2012; 2.3.3) und die CTML (Mayer, 2009; 2.3.3) und für TPACK wird die Begründung in einer Aufgabe auf Grundlage der zuvor genannten Kriterien bewertet (Max et al., 2020). Bezüglich digitaler

Technologien wurde auf Basis des DiKoLAN-Orientierungsrahmens (2.3.1) die Nutzung von Erklärvideos, von Messwerterfassungssystemen und von Animationen und Simulationen als bedeutsam für den Biologieunterricht identifiziert (Max et al., 2020, 2022). Für jede dieser drei digitalen Technologien wurde eine Aufgabe entwickelt, welche sich aus einem Aufgabenstamm mit der Beschreibung einer Unterrichtssituation sowie drei Teilaufgaben zusammensetzt. Die Teilaufgaben unterscheiden dabei auf lernpsychologischer⁷ Ebene drei Anforderungen in Anlehnung an Anderson und Krathwohl (2001). In der ersten Teilaufgabe soll Wissen über die digitale Technologie reproduziert (Anforderung *Reproduzieren/Erinnern*), in der zweiten Teilaufgabe Wissen in der konkreten Unterrichtssituation angewendet und begründet (*Anwenden*) und in der dritten Teilaufgabe Potenziale und Anforderungen für die Technologie diskutiert werden (*Diskutieren/Verstehen*). Die erste und dritte Teilaufgabe zählen zu deklarativem TPACK und die zweite Teilaufgabe zur Anwendung von TPACK (Max et al., 2020, 2022).

Der Leistungstest beinhaltet demnach drei Testaufgaben mit je drei Teilaufgaben und hat eine Bearbeitungszeit von 60 Minuten (Max et al., 2020). Die Bewertung erfolgt anhand eines Kodierleitfadens, welcher – ebenso wie die entwickelten Aufgaben – mit Expert:innen aus mehreren Fachdisziplinen der Naturwissenschaften und Psychologie diskutiert wurde. In dieser Diskussion wurden Qualitätskriterien fokussiert, wie z. B. die Nachvollziehbarkeit und Vollständigkeit der Aufgaben, die inhaltliche Validität sowie die sprachliche und strukturelle Qualität (Max et al., 2020). Die daraufhin überarbeiteten Aufgaben wurden einer kleinen Anzahl von Studierenden zur Bearbeitung vorgelegt, um den Kodierleitfaden weiterentwickeln zu können und eine ergänzende Validierung zu ermöglichen. Anschließend wurde das Testinstrument in einer Gruppe von 45 Biologielehramtsstudierenden pilotiert (Max et al., 2020). Die Bewertung mithilfe des Kodierleitfadens wurde in der Hauptstudie (4.2.1) von drei Rater:innen vorgenommen, welche eine gute Übereinstimmung von Kennwerten zwischen 0,82 bis 0,85 für die einzelnen

⁷ In der Originalquelle wird dies als *pedagogical domain* bezeichnet. In der vorliegenden Arbeit wird diese Variation nach Anforderungen jedoch als lernpsychologisch verstanden.

Testaufgaben und 0,96 für den Gesamttest erzielten (Max et al., 2022). Darüber hinaus wurden aus dem Projekt jedoch keine Untersuchungen zur Validität oder zu weiteren Gütekriterien berichtet.

Auch Kotzebue (2022b) entwickelte einen biologiespezifischen Leistungstest, fokussierte dabei jedoch die TPACK-Bereiche TK, TCK, TPK und TPACK sowie vor allem die Kompetenzbereiche Präsentation, Messwert- und Datenerfassung sowie Datenaufbereitung des DiKoLAN-Orientierungsrahmens (2.3.1). Die Aufgabenentwicklung erfolgte in Anlehnung an etablierte Instrumente zum FDW (Kotzebue, 2022b). Zur Unterstützung der Aufgabenentwicklung wurden sechs Expert:inneninterviews mit Fachdidaktiker:innen und erfahrenen Lehrkräften geführt, um ein valides, objektives und kontextorientiertes Instrument entwickeln zu können. Die Aufgaben zum TPACK sind situationsspezifisch gestaltet, indem der Aufgabenstamm die Vorstellung einer Unterrichtssituation in einer bestimmten Jahrgangsstufe beinhaltet, worauf eine fachspezifische und medienbezogene Aufgabenstellung folgt (Kotzebue, 2022b). Der finale Leistungstest enthält 16 (überwiegend offene) Aufgaben zum Fachinhalt Honigbienen (davon vier Aufgaben zum Bereich TPACK). Die Bewertung der Aufgaben erfolgt mithilfe eines Kodiermanuals, wobei eine Doppelkodierung von 10 % der Testhefte zur Beurteilendenübereinstimmung gute Kennwerte von $\kappa = 0,90$ bis $\kappa = 0,97$ ergab. Die Reliabilität für die einzelnen Bereiche lag zwischen $\alpha = 0,59$ (TPACK) und $\alpha = 0,90$ (Kotzebue, 2022b). Neben den Expert:innenbefragungen im Hinblick auf die Inhaltsvalidität wurden im Projekt zudem Untersuchungen zur konvergenten und diskriminanten Validität (3.3.2) vorgenommen. Dazu wurde die leistungstestbasierte Messung der vier TPACK-Bereiche mit einem TPACK-Selbsteinschätzungsinstrument in diesen Bereichen verglichen (Kotzebue, 2022b). Im folgenden Abschnitt werden diese sowie weitere Studien beschrieben, die in ähnlicher Weise Selbsteinschätzungen und Leistungsmessungen zu TPACK in Kombination einsetzen und miteinander vergleichen.

3.5.3 Kombination von Selbsteinschätzung und Leistungsmessung

In den vorausgehenden Abschnitten wurden einige Verfahren zur Untersuchung des TPACK vorgestellt, welche einerseits objektive proximale Zugänge (3.2.1), die auf der Analyse von Unterrichtsplanungsdokumenten oder auf

vignettenbasierten Leistungstestsaufgaben beruhen, oder andererseits subjektive Zugänge (3.2.1) basierend auf Selbsteinschätzungen darstellen. Diese Zugänge unterscheiden sich insbesondere hinsichtlich des Aufwands in der Konstruktion, der Bearbeitung und der Bewertung der Testaufgaben. Selbsteinschätzungsinstrumente beruhen auf Likert-Skalen, deren Beantwortung in der Regel wenig Zeit einfordert. Sie lassen sich ebenso zeitökonomisch und trivial auswerten. Leistungstestbasierte Messungen erfordern hingegen für die Teilnehmenden mehr Zeit in der Bearbeitung, wenn freie Antworten oder die Planung von Unterricht gefordert werden. Zur objektiven Bewertung ist ein Kodiermanual mit eindeutigen Bewertungsregeln erforderlich, dessen Qualität mittels einer (teilweisen) Doppelkodierung zur Untersuchung der Beurteilendenübereinstimmung sichergestellt werden muss (3.2.2; 3.3.1).

Dass sich dieser Mehraufwand der leistungsbasierten Messung lohnen kann, zeigen einige Studien, die verschiedene Messverfahren in Bezug auf TPACK miteinander kombinieren (z. B. Akyuz, 2018; Kopcha et al., 2014; Kotzebue, 2022b; Max et al., 2022; Stinken-Rösner et al., 2023; Zimmermann, 2021). Durch diese Kombination können Diskrepanzen oder Übereinstimmungen zwischen den Ergebnissen der genutzten Instrumente identifiziert werden, welche Hinweise zur konvergenten und diskriminanten Validität (3.3.2) liefern.

In der Studie von Kopcha et al. (2014) wurden beispielsweise 27 Lehramtsstudierende (Grundschullehramt) mithilfe des Selbsteinschätzungsinstruments von Schmidt et al. (2009; 3.5.1) zu ihrem TPACK befragt und zusätzlich Unterrichtsplanungen der Teilnehmenden mithilfe des Rasters nach Harris et al. (2010; 3.5.2) analysiert. Die Analysen zeigten höhere Zusammenhänge zwischen unterschiedlichen TPACK-Bereichen innerhalb der einzelnen Messverfahren und keine signifikanten Zusammenhänge im Vergleich der beiden unterschiedlichen Messverfahren zu den gleichen TPACK-Bereichen (Kopcha et al., 2014). Ein Vergleich in Bezug auf Einzelfälle zeigte etwa für ein Viertel der Fälle deutliche Diskrepanzen zwischen den TPACK-Ergebnissen aus der Selbsteinschätzung und denen der Bewertung der Unterrichtsplanungen mittels des Rasters. In diesen Fällen schätzten die Teilnehmenden ihr TPACK überwiegend mit am besten ein, wobei ihre Ergebnisse aus der leis-

tungsbasierten Rasterbewertung mit am schlechtesten ausfielen. Eine zusätzliche Sichtung und Analyse von erstellten Portfolios der Teilnehmenden deutete an, dass die Bewertung mithilfe des Rasters das TPACK der angehenden Lehrkräfte im Vergleich angemessener widerspiegelte und die Fragebogenergebnisse durch die Selbsteinschätzung eher abbildeten, was die Teilnehmenden glaubten über TPACK zu wissen (Kopcha et al., 2014).

Ähnliche Unstimmigkeiten identifizierte Akyuz (2018) in seiner Studie mit 138 angehenden Mathematiklehrkräften. Hier wurde zusätzlich zum vorgestellten Instrument zur Bewertung der Aktivitäten mit der DGS und zugehörigen Unterrichtsplanungen hinsichtlich TPACK (3.5.2) ebenfalls das Selbsteinschätzungsinstrument nach Schmidt et al. (2009; 3.5.1) zum Vergleich der Ergebnisse eingesetzt (Akyuz, 2018). Die Ergebnisse der Leistungsbewertung mithilfe des entwickelten leistungsorientierten Instruments waren hier durchweg niedriger als die der Selbsteinschätzung (bis auf den Bereich TCK). Die Diskrepanzen waren dabei für die pädagogikbezogenen Bereiche PK, TPK und TPACK am größten. Die Lehramtsstudierenden schienen ihr pädagogisches Wissen (isoliert sowie in Kombination mit technologischem Wissen) eher zu überschätzen, während sie die inhaltsbezogenen Wissensbereiche realistischer einschätzten (Akyuz, 2018).

Max et al. (2022) untersuchten in ihrer Studie ebenfalls die Konsistenz einer Selbsteinschätzungsmessung zu TPACK im Vergleich zur proximalen Messung mit dem im Projekt entwickelten Leistungstest (3.5.2) zur Erfassung des TPACK von Biologielehramtsstudierenden (Max et al., 2020). Für das Selbsteinschätzungsinstrument wurden unter anderem Items von Schmidt et al. (2009) verwendet (3.5.1). Auch hier wurden Diskrepanzen im Vergleich der Ergebnisse aus beiden Messverfahren deutlich. Zwar konnten mit beiden Verfahren Veränderungen in TPACK im Verlauf der untersuchten Intervention (4.2.1) gemessen werden, jedoch tendierten auch die 149 Teilnehmenden dieser Studie dazu, ihr Wissen zu überschätzen. Dabei überschätzten leistungsschwächere Studierende (in Bezug auf das Ergebnis im Leistungstest) ihr Wissen stärker als leistungsstarke oder mittelstarke Studierende (Max et al., 2022). Die Abweichungen der beiden Messverfahren variierten außerdem je nach Anforderung in den Leistungstestaufgaben. Das Reproduzieren von Fakten oder die Erläuterung genereller Vor- oder Nachteile eines

digitalen Mediums konnte durch die Selbsteinschätzung besser abgebildet werden. Jedoch zeigte sich für die Aufgaben, die die Anwendung von TPACK in einer konkreten Unterrichtssituation adressierten, eine starke Diskrepanz (Max et al., 2022). Die Ergebnisse verdeutlichen die Limitationen von Selbsteinschätzungen, insbesondere wenn es sich um fach- und kontextspezifische Anforderungen handelt und die Teilnehmenden geringes untersuchungsrelevantes Wissen haben (Max et al., 2022).

In ihrer Studie verglich Kotzebue (2022b) ebenfalls die Ergebnisse des entwickelten Leistungstests zu den TPACK-Bereichen TK, TCK, TPK und TPACK (3.5.2) von 106 Biologielehramtsstudierenden in Österreich mit deren Ergebnissen in einer TPACK-Selbsteinschätzung, um Gemeinsamkeiten und Diskrepanzen der Messverfahren und damit Indizien zur konvergenten und diskriminanten Validität (3.3.2) zu gewinnen. Für die Selbsteinschätzung wurde der Fragebogen nach Schmidt et al. (2009) erweitert (3.5.1) und für den biologischen Kontext adaptiert (Kotzebue, 2022b). Die Korrelationsmaße für die verschiedenen TPACK-Bereiche innerhalb der einzelnen Messverfahren zeigten mittlere bis starke Zusammenhänge für die Selbsteinschätzungen und vergleichsweise geringere Zusammenhänge im kleinen bis mittleren Bereich für den Leistungstest. Die Untersuchung der Konvergenz gleicher Konstrukte (d. h. TPACK-Bereiche) zwischen den beiden Messverfahren ergab geringe signifikante Korrelationen für die Bereiche TK, TCK und TPK, jedoch keinen signifikanten Zusammenhang für TPACK (Kotzebue, 2022b). Als mögliche Erklärung für diese eher mäßigen Ergebnisse zur konvergenten und diskriminanten Validität nennt Kotzebue (2022b) den unterschiedlichen Grad der Spezifität beider Instrumente bzw. der dort enthaltenen Aufgaben. Der Leistungstest ist themenspezifisch und bezieht sich auf konkrete Anwendungen, während das Selbsteinschätzungsinstrument zwar für das Fach Biologie angepasst, aber dennoch unspezifisch formuliert ist. In einer anknüpfenden Studie untersuchte Kotzebue (2022a) den Einfluss des TPACK auf die Qualität von Unterrichtsplanungen bei 82 Biologielehramtsstudierenden. Dabei wurde TPACK (als Wissensbereich) sowohl mithilfe des Selbsteinschätzungsinstruments (Schmidt et al., 2009) als auch über die vier Leistungstestaufgaben zum Bereich TPACK (Kotzebue, 2022b) erfasst. In den Ergebnissen

zeigte sich nur die TPACK-Messung über den Leistungstest als positiver Prädiktor für die Qualität von Unterrichtsplanungen, nicht jedoch die TPACK-Selbsteinschätzung, woraus Kotzebue (2022a) schlussfolgert, dass Selbsteinschätzungen zu TPACK allein kein geeignetes Instrument zur Vorhersage guter Unterrichtsqualität sind. Für die proximale TPACK-Messung stellt dieses Ergebnis hingegen einen Hinweis zur Konstruktvalidität dar.

Die vorgestellten Ergebnisse zeigen dahingehend Unstimmigkeiten in der Messung von TPACK auf, dass mit Instrumenten beruhend auf Selbsteinschätzungen und proximalen Leistungstests oder Rastern zur Bewertung von Unterrichtsplanungen scheinbar nicht dasselbe gemessen wird. Dies unterstützt die Kritik, dass mit Selbsteinschätzungen eher Selbstwirksamkeitsüberzeugungen erfasst werden und nicht das tatsächliche Wissen (Backfisch et al., 2020b; Lachner et al., 2019; Willermark, 2018). Nach Backfisch et al. (2020b) liegt in diesem Fall das Problem einer *jingle-jangle-fallacy* vor (Gonzalez et al., 2021). Dabei handelt es sich um zwei Trugschlüsse (*fallacies*) hinsichtlich der konvergenten Validität. Einerseits werden zwei Messungen unter demselben Begriff gefasst, obwohl sie eigentlich unterschiedliche Konstrukte abbilden; hier, wenn man davon ausgeht, dass selbsteingeschätzte und leistungsbasierte Messung von TPACK dasselbe Konstrukt erfassen. Andererseits werden zwei Messungen unterschiedlich bezeichnet, obwohl sie eigentlich dasselbe darstellen; hier selbst eingeschätztes TPACK und Selbstwirksamkeitserwartungen in Bezug auf digitalgestützten Unterricht (Backfisch et al., 2020b). Diese Trugschlüsse können dann problematisch werden, wenn Ergebnisse bezüglich einer TPACK-Selbsteinschätzung interpretiert und daraus Implikationen gezogen werden (Backfisch et al., 2020b).

Auch in der Kompetenzmessung im Allgemeinen liegt eine ähnliche Kritik an Selbsteinschätzungen vor (3.2.1). In Bezug auf die Messung von TPACK ist diese Kritik aufgrund der häufigen Nutzung solcher Instrumente jedoch von besonderer Bedeutung und wurde hier deshalb vertiefter adressiert. Nichtsdestotrotz können Selbsteinschätzungen zu TPACK auch nützlich sein, da sie zwar scheinbar nicht das tatsächliche Wissen, aber dennoch ein wichtiges Merkmal abbilden (Kotzebue, 2022a). Nach Scherer et al. (2017) bieten sie leicht zugängliche, ökonomische und zuverlässige Indikatoren für Selbstüberzeugungen und hängen damit auch mit den Absichten der (angehenden)

Lehrkräfte zusammen, digitale Medien einzusetzen. Beachtet werden sollte in der Forschung jedoch die zutreffende Bezeichnung und Reflexion dessen, was mit TPACK-Selbsteinschätzungen gemessen (und nicht gemessen) wird, um im Sinne einer validen Testwertinterpretation (3.3.2) keine unangemessenen Schlussfolgerungen zu ziehen.

3.6 Zusammenfassung und Schlussfolgerungen für die Messung des Fachdidaktischen Wissens zum Einsatz digitaler Medien

Ausgehend von allgemeinen Zugängen in der fachdidaktischen Forschung sowie Verfahren und Qualitätsanforderungen der Kompetenzmessung in der Lehrkräftebildungsforschung wurden in diesem Kapitel beispielhafte Messverfahren vorgestellt, die auf die Erfassung des FDW bzw. TPACK angehender Lehrkräfte in den naturwissenschaftlichen Fächern abzielen.

Allgemein sind in der fachdidaktischen Forschung verschiedene Zugänge zur Untersuchung einer fachdidaktischen Fragestellung möglich, die qualitativ oder quantitativ sein können oder eine Kombination daraus darstellen (3.1). Die Kompetenzmessung als Teil der Lehrkräftebildungsforschung verwendet dabei in der Regel quantitative Verfahren wie Selbsteinschätzungen oder Leistungstests (3.2.1), welche in der Vergangenheit auch zur Messung des FDW oder des TPACK genutzt wurden. Während Selbsteinschätzungen in der Regel mit Ratingskalen beantwortet werden, können Leistungstests verschiedene freie/halboffene oder gebundene/geschlossene Antwortformate aufweisen, die jeweils Vor- und Nachteile mit sich bringen. Im Kontext der fachdidaktischen Forschung und der Kompetenzmessung bietet es sich zudem an, in den Leistungstestaufgaben Unterrichtsvignetten zu nutzen, um möglichst handlungsrelevantes und kein träges Wissen zu erfassen (3.2.2). Die zur Kompetenzmessung verwendeten Messverfahren sollten weiterhin bestimmten Qualitätskriterien wie z. B. der Objektivität, Reliabilität und Validität gerecht werden (3.3.1). Insbesondere bei der Messung von FDW oder TPACK mit geschlossenen Aufgaben stellt die Sicherstellung der Validität eine besondere Herausforderung dar. In dieser Arbeit wird sich in diesem Zusammenhang auf ein umfassendes Validitätsverständnis im Sinne der Bewertung einer Testwertinterpretation (Kane, 1992, 2001; Messick, 1995) bezogen. Dabei können unterschiedliche Validitätsaspekte über verschiedene

Untersuchungsverfahren Hinweise und Evidenzen für eine angemessene Testwertinterpretation liefern (3.3.2).

In den Abschnitten 3.4 und 3.5 wurden schließlich beispielhaft Messinstrumente zum FDW und TPACK sowie deren Testentwicklung beschrieben. Das FDW in Physik wurde bereits in mehreren nationalen Studien mithilfe proximaler und inhaltspezifischer Testinstrumente untersucht (3.4). Die dazu vorgestellten Messverfahren weisen einerseits unterschiedliche Schwerpunkte oder Eigenschaften auf, z. B. im Hinblick auf die physikalisch-inhaltliche Breite oder die damit zu untersuchende Zielgruppe. Andererseits zeigen sich auch Gemeinsamkeiten der Studien, insbesondere im Vorgehen zur Aufgabenentwicklung sowie der Untersuchung verschiedener Validitätsaspekte. Weiterhin werden zwar teilweise unterschiedliche fachdidaktische Facetten in den entwickelten Testinstrumenten eingeschlossen, es gibt jedoch noch kein Instrument zum FDW in Physik, das eine Facette zu digitalen Medien berücksichtigt (3.4).

In der Forschung zum TPACK-Modell wird hingegen genau das Wissen zu digitalen Medien oder Technologien im Professionswissen einer Lehrkraft in den Vordergrund gestellt. Die bisher entwickelten Instrumente zur Messung von TPACK beruhen jedoch meist auf Selbsteinschätzungen und sind häufig generisch und nicht fach- oder inhaltspezifisch formuliert (3.5.1). Während fachspezifische Leistungstests in der Forschung zum FDW also bereits etabliert sind, wird hier gerade erst begonnen, statt oder neben Selbsteinschätzungen leistungsbasierte Testinstrumente zu entwickeln und einzusetzen (3.5.2). Das TPACK-Modell und dazu verwendete Messverfahren werden daher im Hinblick auf mangelnde Fachspezifität und fehlende Evidenz zur Validität kritisiert (z. B. Kotzebue, 2022a; 3.5.1; 3.5.3). Dies scheint insofern nachvollziehbar, da im Vergleich zu den vorgestellten Messverfahren zum FDW (3.4) die Validierungsuntersuchungen in der Entwicklung von TPACK-Instrumenten auch weniger elaboriert und mehrperspektivisch ausfallen. Hauptsächlich werden hier Expert:innenbefragungen zur Inhaltsvalidierung einbezogen, aber darüber hinaus nur wenig Validitätsbetrachtungen vorgenommen, um konkrete Hinweise darauf zu geben, dass mit dem entwickelten Instrumenten tatsächlich TPACK gemessen werden kann. Bei Untersuchungen mit gemeinsamen Einsatz leistungsbasierter Testinstrumente zu TPACK

sowie TPACK-Selbstberichten deuten Unstimmigkeiten zwischen beiden Messverfahren auf mangelnde konvergente und diskriminante Validität hin (z. B. Akyuz, 2018; Kopcha et al., 2014; Kotzebue, 2022b; Max et al., 2022; 3.5.3). Im Hinblick auf die Validität von TPACK-Selbstberichten wird daher zunehmend kritisiert, inwiefern damit tatsächlich Wissen gemessen wird oder ob diese Instrumente nicht eher die Selbstwirksamkeit der Teilnehmenden bezüglich des TPACK messen (z. B. Backfisch et al., 2020b; Kotzebue, 2022a; Lachner et al., 2021). Diese Kritik adressierend gibt es einzelne Ansätze der leistungsorientierten Messung des TPACK von Lehramtsstudierenden (3.5.2). Häufig werden dabei Unterrichtsplanungen der Studierenden untersucht und mithilfe eines Rasters bewertet (z. B. Akyuz, 2018; Harris et al., 2010; Zimmermann, 2021). Diese Verfahren sind jedoch aufwändig in der Bearbeitung sowie insbesondere in der Bewertung. Manche der vorgestellten Studien nutzen spezifische Leistungstests, um fachspezifisches TPACK zu messen (Kotzebue, 2022b; Max et al., 2020). Auch für diese leistungsorientierten Messungen über Unterrichtsplanungen oder Leistungstests wurden vorrangig Expert:innenbefragungen zur inhaltlichen Validierung genutzt. In den Arbeiten von Kotzebue (2022a, 2022b) wurden zudem konvergente und diskriminante Aspekte der Konstruktvalidität untersucht sowie Hinweise gewonnen, dass die Leistungstestergebnisse im TPACK die Qualität von Unterrichtsplanungen von Studierenden positiv vorhersagen (3.5.3). Für das Fach Physik existiert jedoch bislang kein leistungsorientiertes Messinstrument (3.5.2).

Insgesamt zeigt sich, dass es zwar etablierte Leistungstests zur Messung des FDW in Physik sowie auch erste Leistungstests zur Messung von TPACK in naturwissenschaftlichen Fächern gibt; ein Instrument, welches fachspezifisches TPACK oder FDW zum Einsatz digitaler Medien im Fach Physik erfasst, gibt es jedoch bisher nicht. Ausgehend von den beschriebenen Messverfahren scheint es zur Messung dieses FDW zum Einsatz digitaler Medien in Physik sinnvoll, sich der Entwicklung eines Leistungstests zu widmen, da Selbsteinschätzungen kein ausreichend valides Verfahren zur Erfassung von kognitiven Kompetenzaspekten wie dem FDW oder TPACK darstellen (3.5.3). In der Erstellung und im Einsatz eines solchen Testinstruments sind Validie-

rungsbetrachtungen bedeutsam, die über Expert:innenbefragungen zur inhaltlichen Validität hinausgehen und stattdessen mehrere Validitätsaspekte (3.3.2) beleuchten. Um diesen Mangel an Validität sowie an Fachspezifität zu vermeiden, den einige Instrumente zum TPACK bisher aufweisen, sollte die Entwicklung eines solchen Testinstruments an den Verfahren der Testentwicklung und -validierung fachspezifischer Leistungstests orientiert sein, wie sie in der Forschung zum FDW in Physik bereits etabliert sind (3.4). Um gleichzeitig den Technologie- oder Medienbezug sowie die fachdidaktische Perspektive zu berücksichtigen, ist die Orientierung an fachspezifischen Technologie- und Medieneinsätzen für den Unterricht sinnvoll (2.3.2), wie sie in manchen Leistungstests zum TPACK bereits erfolgt (z. B. Kotzebue, 2022b; Max et al., 2022; 3.5.2).

Im nachfolgenden Kapitel werden schließlich für diese Arbeit relevante Forschungsergebnisse hinsichtlich der Entwicklung des Professionswissens von Lehramtsstudierenden erläutert. Dabei werden empirische Studien betrachtet, die einerseits die Entwicklung des FDW von Lehramtsstudierenden in Physik und andererseits die Entwicklung von TPACK im Lehramt in den Naturwissenschaften untersuchen. Die thematisierten Studien greifen dabei auf Messverfahren zum FDW oder TPACK zurück, die in diesem Kapitel vorgestellt wurden.

4 Forschungsergebnisse zum Fachdidaktischen Wissen und TPACK

In diesem Kapitel werden bisherige Forschungsergebnisse zum FDW sowie zum TPACK bei Lehramtsstudierenden im deutschsprachigen Raum dargestellt. Dabei widmen sich die Abschnitte in 4.1 ausgewählten Ergebnissen zum FDW im Fach Physik und die Abschnitte in 4.2 den Ergebnissen zum TPACK in naturwissenschaftlichen Fächern. Es werden jeweils Forschungsergebnisse aus Studien zur Entwicklung des jeweiligen Wissens im Lehramtsstudium oder über einzelne Lehrveranstaltungen (4.1.1 bzw. 4.2.1) sowie aus (qualitativen) Untersuchungen zu Zusammenhängen zwischen den Wissensbereichen und Lerngelegenheiten (4.1.2 bzw. 4.2.2) vorgestellt. In 4.3 folgt abschließend eine kurze Zusammenfassung dieses Kapitels.

4.1 Ausgewählte Forschungsergebnisse zum Fachdidaktischen Wissen

Das Professionswissen von (angehenden) Lehrkräften, d. h. Fachwissen, pädagogisches Wissen sowie FDW, gehört zum Kern professioneller Kompetenz und seine Erforschung ist entsprechend bedeutsam für die Lehrkräftebildungsforschung (z. B. Kaiser et al., 2020; 2.1). In den naturwissenschaftlichen Fächern sowie der Mathematik wurde in der Vergangenheit vor allem die innere Struktur des Professionswissens, die Entwicklung der Professionswissensbereiche sowie deren Bedeutung für die erfolgreiche Handlungs- und Unterrichtsqualität empirisch untersucht (z. B. Baumert et al., 2010; Blömeke et al., 2022; Förtsch et al., 2016; Riese et al., 2022b). Als wesentlicher Kernbereich des Professionswissens wurde auch das FDW in diesen Studien erforscht und dessen praktische Relevanz durch empirische Ergebnisse gestützt.

Neben einzelnen abweichenden Ergebnissen, die keinen signifikanten Zusammenhang zwischen dem FDW der Lehrkräfte und den Lernergebnissen der Schülerinnen und Schüler oder der Unterrichtsqualität feststellen konnten (Cauet et al., 2015), zeigten sich in mehreren Studien unterschiedlicher Fächer derartige Zusammenhänge für das FDW. So konnte im COACTIV-Projekt im Fach Mathematik (Baumert et al., 2010; 2.1) ein positiver Effekt des

FDW der befragten Lehrkräfte auf den Lernzuwachs ihrer Schülerinnen und Schüler mediiert durch die Bereitstellung kognitiver Aktivierung und individueller Lernunterstützung (als Maße für die Unterrichtsqualität) festgestellt werden. Das FDW zeigte einen maßgeblichen Einfluss auf die Unterrichtsqualität und die größte Vorhersagekraft für den Lernfortschritt der Lernenden (Baumert et al., 2010). Im Fach Physik stellten Keller et al. (2017) ebenfalls fest, dass das FDW der Lehrkräfte die kognitive Aktivierung im Unterricht positiv beeinflusst und diese wiederum den Lernfortschritt der Schülerinnen und Schüler. Ein solcher indirekter Zusammenhang wurde bei Förtsch et al. (2016) ebenso für das Fach Biologie deutlich. Ähnlich zeigte sich bei Blömeke et al. (2022) ein indirekter Zusammenhang zwischen dem FDW befragter Mathematiklehrkräfte und dem Lernzuwachs der Schülerinnen und Schüler, welcher hier über die fachspezifischen Handlungsfähigkeiten der Lehrkräfte (im Sinne des Kontinuumsmodells zur Kompetenz: Blömeke et al., 2015; 2.1) mediiert wurde. Im Rahmen des Projekts ProfiLe-P+ (Vogelsang et al., 2019; 3.4) wurde die Relevanz des FDW in Physik zudem durch Zusammenhänge mit der Entwicklung prototypischer Handlungsfähigkeiten wie der Unterrichtsplanungsfähigkeit (Riese et al., 2022b; Schröder et al., 2020) sowie der Erklärfähigkeit (Kulgemeyer et al., 2020; Kulgemeyer & Riese, 2018) empirisch gestützt.

Insgesamt legen diese Befunde einen relevanten Einfluss des FDW auf das Unterrichtshandeln sowie die Unterrichtsqualität und z. T. indirekt auf die Lernleistungen der Schülerinnen und Schüler nahe. Der Erwerb von FDW scheint daher für angehende Lehrkräfte bedeutsam und sollte im Rahmen der Lehrkräftebildung gefördert werden. Der folgende Abschnitt beschreibt Forschungsergebnisse zur Entwicklung des FDW im Lehramtsstudium Physik, welche Indizien für die Wirksamkeit der universitären Lehrkräftebildung bzw. einzelner Lehrveranstaltungen darstellen. Dabei werden Studien dargestellt, deren verwendete Testinstrumente zur Messung des FDW in 3.4 vorgestellt wurden.

4.1.1 Entwicklung des Fachdidaktischen Wissens im Lehramt Physik

In den Projekten zum Professionswissen sowie zum FDW in Physik wurden z. T. Querschnittsuntersuchungen bei Physiklehramtsstudierenden unterschiedlichen Studienfortschritts vorgenommen, um erste Hinweise für die Lernwirksamkeit der universitären (fachdidaktischen) Lehrkräftebildung zu erhalten. So zeigten sowohl die Untersuchungen des Projekts KiL (Sorge et al., 2019) als auch des Projekts ProfiLe-P (Riese et al., 2017), dass fortgeschrittenere Studierende gegenüber Studierenden zu Studienbeginn höhere Ausprägungen im FDW in Physik aufweisen. Um eine aussagekräftigere Untersuchung der Wissensentwicklung zu ermöglichen, wurden jeweils in anknüpfenden Folgeprojekten auch Längsschnittuntersuchungen zur Entwicklung des FDW in Physik über das Lehramtsstudium durchgeführt. So wurde im Projekt KeiLa (Nachfolgeprojekt von KiL) eine Längsschnittstudie an 20 deutschen Hochschulen zur Untersuchung der Entwicklung des FDW (und des Fachwissens) in Physik durchgeführt (Sorge et al., 2018). Dabei konnten 68 Lehramtsstudierende im Fach Physik zu zwei Messzeitpunkten im Abstand von einem Jahr befragt werden, wobei das im Vorgängerprojekt KiL entwickelte Testinstrument zum FDW (Kröger et al., 2015; 3.4) genutzt wurde. Die Studie zeigte, dass die Fähigkeit im FDW zum ersten Messzeitpunkt die Fähigkeit zum zweiten Messzeitpunkt positiv voraussagt. Studierende, die bereits im ersten Messzeitpunkt hohe Ausprägungen im FDW besaßen, wiesen demnach auch im zweiten Messzeitpunkt ein hohes FDW auf (Sorge et al., 2018).

Im Projekt ProfiLe-P+ (zweite Projektphase von ProfiLe-P) wurde ebenfalls das Professionswissen angehender Physiklehrkräfte im längsschnittlichen Verlauf untersucht, einerseits im Verlauf der ersten beiden Studienjahre im Bachelorstudium sowie andererseits im Verlauf eines Praxissemesters (Riese et al., 2022a; Vogelsang et al., 2019). In der Bachelor-Kohorte des Projekts wurden die Studierenden zu drei Messzeitpunkten – zu Beginn des ersten, dritten und fünften Semesters – an zehn Hochschulen in Deutschland und Österreich befragt, wobei die erprobten Testinstrumente zum FDW (3.4), Fachwissen und pädagogischen Wissen aus der ersten Projektphase genutzt wurden (Vogelsang et al., 2019). In der Kohorte zum Praxissemester wurden

Masterstudierende an vier Hochschulen vor und nach einem ca. fünfmonatigen Praxissemester ebenfalls mit den Leistungstests zu den drei Professionswissensbereichen befragt (Riese et al., 2022a; Vogelsang et al., 2019). Aus diesen Erhebungen konnten Längsschnittdaten von 109 Studierenden im ersten Studienjahr (Messzeitpunkte erstes und drittes Semester), 45 Studierenden im zweiten Studienjahr (Messzeitpunkte drittes und fünftes Semester) und 69 Studierenden im Praxissemester zum FDW in Physik gewonnen werden (Riese et al., 2022a). Für alle betrachteten Längsschnittphasen wurden dabei signifikante Zuwächse im FDW mit hohen Effektstärken im ersten und zweiten Studienjahr im Bachelorstudium ($d = 1,05$ und $d = 0,86$) und mittlerer Effektstärke über das Praxissemester ($d = 0,44$) festgestellt (Riese et al., 2022a). Bei Betrachtung unterschiedlicher Leistungsgruppen bezüglich der Fähigkeit im FDW zum jeweils ersten Messzeitpunkt zeigte sich (ebenfalls für alle betrachteten Phasen), dass der Effekt für das leistungsschwächste Drittel am größten und für das leistungsstärkste Drittel am geringsten ist. Im Verlauf des Praxissemesters konnte für das zu Beginn leistungsstärkste Drittel sogar kein Zuwachs, sondern nur eine Stagnierung festgestellt werden. Diese Ergebnisse deuten darauf hin, dass sich die Leistungsschere für das FDW im Verlauf des Lehramtsstudiums schließt (Riese et al., 2022a).

Auch in der Arbeit von Joswig-Käfer (2024) erfolgte eine längsschnittliche Untersuchung des FDW von Physiklehramtsstudierenden. Hier wurden 38 Masterstudierende im Verlauf eines Semesters (Vorbereitungssemester zum Praxissemester) zu ihrem FDW in Physik in Prä-Post-Erhebungen befragt (Joswig-Käfer & Riese, 2021). Dazu nutzte Joswig-Käfer (2024) einen im Projekt erweiterten Leistungstest zum FDW, welcher die Testaufgaben aus dem Projekt ProfiLe-P zum Inhaltsbereich Mechanik (Gramzow, 2015) um strukturgleiche Testaufgaben zu den Inhaltsbereichen Elektrizitätslehre und Optik ergänzt (3.4; Joswig & Riese, 2019). Diese inhaltliche Erweiterung zielte darauf ab, die Rolle des physikalischen Inhaltsbereiches bei der Entwicklung des FDW zu untersuchen. Die befragten Studierenden besuchten im untersuchten Semester zwei physikdidaktische Seminare, in denen fachdidaktische Inhalte und (Unterrichts-)Beispiele aus den Bereichen Elektrizitätslehre und Optik thematisiert wurden, nicht aber aus dem Bereich Mechanik. Die Ergebnisse zeigen, dass die Teilnehmenden einen signifikanten Zuwachs im

FDW mittlerer Effektstärke über das untersuchte Semester erfuhren ($d = 0,65$; bezogen auf den Gesamttest). In Bezug auf die einzelnen Inhaltsbereiche war für die Leistungen in den FDW-Testteilen zur Elektrizitätslehre und Optik ebenfalls ein signifikanter Zuwachs mittlerer bzw. kleiner Effektstärke ($d = 0,59$ bzw. $d = 0,41$) vermerkbar, nicht aber für den Inhaltsbereich Mechanik. Dieser Befund – passend zur inhaltlichen Ausrichtung der Seminare – deutet an, dass sich FDW inhaltspezifisch je nach Lerngelegenheit und nicht als Ganzes entwickelt (Joswig-Käfer & Riese, 2021).

Die vorgestellten Studien verdeutlichen, dass das Lehramtsstudium Physik und an der Universität implementierte physikdidaktische Lerngelegenheiten in der Lage sind, das FDW der Studierenden zu fördern. In der bisherigen Forschung wurde dabei auch untersucht, welche Personenmerkmale die Ausprägung oder Entwicklung des FDW von Lehramtsstudierenden beeinflussen.

So zeigt sich in der bisherigen Forschung, dass die Ausprägung des FDW der Physiklehramtsstudierenden unabhängig vom Geschlecht (Riese et al., 2017; Riese & Reinhold, 2012; Sorge et al., 2019) oder sprachlichen Fähigkeiten (Riese et al., 2017) zu sein scheint, jedoch signifikant von der Abiturnote abhängt (Riese et al., 2017; Riese & Reinhold, 2012; Schödl & Göhring, 2017; Sorge et al., 2019), welche sich als Indikator für allgemeine kognitive Leistungsfähigkeit etabliert hat (Riese et al., 2017; Riese & Reinhold, 2012). Zudem scheint hohes Fachwissen eine Voraussetzung für hohes FDW zu sein (Riese & Reinhold, 2012). Sorge et al. (2019) konnten weiterhin feststellen, dass Studierende zu Beginn ihres Studiums einen engeren Zusammenhang des FDW mit ihrem pädagogischen Wissen aufweisen, während fortgeschrittene Studierende einen größeren Zusammenhang mit Fachwissen zeigen. Hinsichtlich schulpraktischer Vorerfahrungen der Lehramtsstudierenden ergaben sich verschiedene Befunde: das Hospitieren von mehr als 20 Unterrichtsstunden im Physikunterricht zeigte sich im Projekt KiL als signifikanter Einflussfaktor auf das FDW, die eigene Lehrerfahrung von mehr als 10 eigenen Unterrichtsstunden hingegen nicht (Sorge et al., 2019). Für Studierende in der Studieneingangsphase erwies sich eine umfangreiche Lehrerfahrung im Projekt ProfiLe-P sogar als tendenziell kontraproduktiv; womöglich, da diese ohne bislang explizite fachdidaktische Lerngelegenheiten nicht unter

fachdidaktischen Gesichtspunkten reflektiert werden kann (Riese et al., 2017).

Weiterhin wurden häufig Zusammenhänge des FDW der Physiklehramtsstudierenden mit Kovariaten untersucht, die als Indikatoren für formale Lerngelegenheiten dienten (z. B. Fachsemester oder Semesterwochenstunden in der Fachdidaktik). Solche Zusammenhänge mit Lerngelegenheiten unterstützen zusätzlich die Wirksamkeit der (fachdidaktischen) Ausbildung im Lehramtsstudium und werden im folgenden Abschnitt genauer erläutert.

4.1.2 Zusammenhänge von Lerngelegenheiten und Fachdidaktischen Wissen

In bisherigen Studien zeigte sich mehrfach ein signifikanter Zusammenhang im FDW mit der Studiendauer bzw. dem Fachsemester der Physiklehramtsstudierenden (Riese et al., 2017; Riese & Reinhold, 2012; Schödl & Göhring, 2017; Sorge et al., 2019) sowie mit besuchten Semesterwochenstunden (SWS) oder erreichten Studienleistungspunkten (*credit points*, CP) in der Physikdidaktik (Riese et al., 2017; Riese & Reinhold, 2012). Diese Zusammenhänge legen nahe, dass das Lehramtsstudium Physik förderlich für den Erwerb von FDW ist und auch einzelne fachdidaktische Lerngelegenheiten zum Ausbildungserfolg beitragen. Riese et al. (2017) konnten zudem den Einfluss erworbener CP in physikdidaktischen Lehrveranstaltungen auf die vier erhobenen Facetten des FDW (3.4; Abbildung 3-2) untersuchen, wobei sich ein mittlerer Effekt bei den Facetten Schülervorstellungen und fachdidaktische Konzepte und ein kleiner Effekt bei der Facette Instruktionsstrategien zeigte. Für die Facette Experimente zeigte sich kein signifikanter Effekt durch erworbene fachdidaktische CP, hingegen jedoch durch CP, die in physikalischen Laborpraktika erworben wurden (Riese et al., 2017).

In der Studie von Schiering et al. (2021) wurde weiterführend untersucht, inwiefern gewisse Studienstrukturen im Lehramtsstudium Physik sich auf das FDW (und Fachwissen) angehender Physiklehrkräfte auswirken. Auf Basis der Curricula von 20 deutschen Hochschulen konnten mithilfe einer Clusteranalyse vier prototypische Ausbildungsprofile für das Lehramtsstudium Physik (in den ersten sechs Semestern) mit geringerem bzw. umfangreichem

Fach- und Fachdidaktikanteil identifiziert werden. Zudem wurde zur Triangulation (3.1) auf Daten zum FDW und Fachwissen von befragten Physiklehr-
amtsstudierenden aus dem Projekt KeiLa (Sorge et al., 2019; 4.1.1) zurückge-
griffen. Während sich Ausbildungsprofile mit umfangreichem Fachanteil sig-
nifikant besser auf das Fachwissen der Studierenden auswirkten, zeigte sich
kein solcher Effekt für Profile mit umfangreichem Fachdidaktikanteil und das
FDW (Schiering et al., 2021). Dabei ist zu berücksichtigen, dass der mittlere
Umfang der fachdidaktischen Ausbildung nur etwa ein Fünftel des mittleren
Umfangs der fachlichen Ausbildung in den ersten sechs Semestern im Lehr-
amtsstudium Physik ausmacht. Selbst für die Profile mit umfangreichem
Fachdidaktikanteil fiel die mittlere Anzahl an SWS in der Fachdidaktik (ca. 13
SWS) deutlich geringer aus als die mittleren SWS im Fach bei Profilen mit
geringem Fachanteil (ca. 32 SWS). Aufgrund dieses geringen Ausmaßes der
fachdidaktischen Ausbildung in den ersten sechs Semestern war es in der
Studie nicht möglich, eine „typische“ inhaltliche Abfolge der fachdidakti-
schen Ausbildung zu beschreiben (Schiering et al., 2021). Auch wenn die Zu-
gehörigkeit zu einem Ausbildungsprofil allein keinen Effekt auf das FDW der
Studierenden hatte, zeigte sich jedoch ein signifikanter Interaktionseffekt
von Profilen mit geringem fachdidaktischen Anteil und dem Ausmaß des
Fachwissens auf das FDW, da sich in diesen Fällen die Höhe des Fachwis-
sens der Studierenden signifikant auf deren FDW auswirkte – für den Fall von
vielen fachdidaktischen Veranstaltungen jedoch nicht. Schiering et al. (2021)
folgerten daraus, dass in Ausbildungsprofilen mit wenig fachdidaktischen
Veranstaltungen bis zum Ende des sechsten Semesters scheinbar die Höhe
des Fachwissens ausschlaggebend für das FDW der Studierenden ist; und
dass womöglich bei Ausbildungsprofilen mit vielen fachdidaktischen Veran-
staltungen mehr Lernzeit verfügbar ist, um grundlegende fachdidaktische
Konzepte unabhängig vom Fachwissen zu vermitteln.

Mit dieser Studie konnten bedeutsame kategorielle Zusammenhänge von
Lerngelegenheiten und der Kompetenzentwicklung im FDW (und Fachwis-
sen) in Physik aufgezeigt werden, die über die zuvor bekannten Zusammen-
hänge von Fachsemester und SWS mit dem FDW (siehe oben) hinausgehen.
Die methodische Fokussierung auf Lehrveranstaltungen in den ersten sechs

Semestern des Lehramtsstudiums Physik hat jedoch Einschränkungen hinsichtlich der Interpretation der Ergebnisse zur fachdidaktischen Ausbildung sowie dem Erwerb von FDW zur Folge, da bis zum Ende des sechsten Semesters sehr wenige fachdidaktische Lerngelegenheiten (im Vergleich zu fachlichen) stattfinden (Schiering et al., 2021).

In der Arbeit von Joswig-Käfer (2024) wurde hingegen das FDW von Physiklehramtsstudierenden im Verlauf eines Semesters im Masterstudium untersucht (4.1.1). Weiterführend erfolgten mit 24 der befragten Studierenden qualitative Interviews anknüpfend an die Prä-Post-Erhebung zum FDW, um Ursachen für veränderte Testantworten zu erkunden und lernförderliche Lerngelegenheiten zur Entwicklung des FDW zu identifizieren. Diese qualitative Studie zeigte, dass die Studierenden häufig Wissen aus universitären (insbesondere physikdidaktischen, zum Teil auch fachlichen) Lerngelegenheiten zur Beantwortung der Testantworten heranzogen und dieses Wissen auch eine erfolgreiche Beantwortung der Testaufgaben zum FDW begünstigte. Weiterhin sind die Studierendenantworten auch durch das Testinstrument selbst (z. B. Formulierung der Aufgabenstellungen, Hinweise durch andere Testaufgaben) oder durch verändertes Zeitmanagement bedingt (Joswig-Käfer, 2024). Als genannte Ursachen für Antwortveränderungen, die deutlich häufiger mit verbessertem als mit verschlechtertem Testscore im FDW einhergingen, wurden insbesondere die behandelten Themen in den beiden untersuchten physikdidaktischen Seminaren (wie „Schülervorstellungen“, „Experimente“ und „Lernzirkel“) deutlich. Diese Ergebnisse unterstützten die erwartete Wirksamkeitsannahme, dass die untersuchten physikdidaktischen Lerngelegenheiten die Entwicklung des FDW der Studierenden positiv beeinflussten (Joswig-Käfer, 2024; Joswig-Käfer & Riese, 2021). Weiterhin trat die Vorgehensweise, sich in die Lehrperson oder die Lernenden hineinzusetzen, sowie eine fachdidaktisch geänderte Sichtweise etwas häufiger in Kombination mit einem verbesserten Testscore auf. Möglicherweise führte der Erwerb von FDW im Untersuchungszeitraum bei den Studierenden dazu, dass sie diese Vorgehensweise erfolgreich anwenden konnten oder ihre Sichtweise hinsichtlich fachdidaktischer Gesichtspunkte veränderten. Die Nutzung physikalischen Fachwissens stellte sich als zum Teil

erfolgreich für die Beantwortung der FDW-Aufgaben heraus, da diese genannte Ursache etwas häufiger im Zusammenhang mit verbesserten, aber auch bei verschlechterten Testscore vorkam. Zu einem Teil ließen sich die veränderten Testscores auch mit der Motivation der Studierenden oder mit der Sorgfalt bei der Aufgabenbearbeitung in Verbindung bringen (Joswig-Käfer, 2024).

Mit dieser qualitativen Interviewstudie vertiefend zur Prä-Post-Erhebung des FDW über ein Semester konnte die Wirksamkeit der dort untersuchten physikdidaktischen Seminare gestützt werden. Während der gemessene Zuwachs im FDW (4.1.1) nicht ohne Weiteres auf den Besuch der beiden Seminare zurückgeführt werden kann, bestärken die Ergebnisse der Interviewstudie diese Vermutung, da in den Seminaren behandelte Themen häufig als Ursache für veränderte Testantworten genannt wurden und dabei öfter mit einer verbesserten Testbearbeitung einhergingen (Joswig-Käfer, 2024).

Im folgenden Abschnitt werden nun analog ausgewählte Forschungsergebnisse zum TPACK – zur Entwicklung des Wissens über Lehrveranstaltungen sowie zu Zusammenhängen mit Lerngelegenheiten – dargestellt.

4.2 Ausgewählte Forschungsergebnisse zum TPACK

Gemäß der Konzeption im TPACK-Modell (2.3.1) nach Mishra und Koehler (2006) bzw. Koehler et al. (2013) sowie weiteren Forschungsarbeiten zu diesem Modell wird der Wissensbereich TPACK als bedeutsam für einen umfassenden didaktischen Einsatz digitaler Technologien oder Medien gesehen (z. B. Chai et al., 2013a; Schmid et al., 2020; Voogt et al., 2013). Hingegen konnte in der Studie von Schmid et al. (2021) kein Zusammenhang zwischen der Ausprägung in den TPACK-Wissensbereichen und dem geplanten Einsatz digitaler Medien (erfasst über Unterrichtsplanungen) festgestellt werden. Lediglich bei Berücksichtigung der Unterrichtsfächer zeigten Studierende in MINT-Fächern mit höherem TPACK (über alle Wissensbereiche) auch ein höheres Maß an Integration digitaler Medien in ihren Unterrichtsplanungen (Schmid et al., 2021). Als möglichen Grund für die fehlenden Zusammenhänge diskutieren Schmid et al. (2021) die Validität ihrer TPACK-Messung über Selbsteinschätzungen. Dafür spricht auch die Studie von Kotzebue

(2022a) im Fach Biologie, in der sich für die verwendete TPACK-Selbsteinschätzung kein signifikanter Zusammenhang mit der Qualität der Unterrichtsplanungen (als Maß für die Unterrichtsqualität) zeigte, während das Ergebnis des TPACK-Leistungstests einen positiven Effekt auf die Unterrichtsqualität hatte (Kotzebue, 2022a; 3.5.3).

Zudem werfen Schmid et al. (2021) als weiteren möglichen Grund für die fehlenden Zusammenhänge in ihrer Studie die Frage auf, ob TPACK überhaupt der entscheidende Aspekt für die Integration digitaler Medien in den Unterricht ist. Verschiedene Studien betrachten daher auch motivationale und affektive Aspekte im Zusammenhang mit dem Lernen und Lehren mit digitalen Medien (Backfisch et al., 2021; Joo et al., 2018; Kotzebue, 2022a; Lachner et al., 2021; Max et al., 2023; Mayer & Girwidz, 2019; Stinken-Rösner, 2021a; Walpert & Wodzinski, 2023; Zimmermann, 2021). So stellten sich z. B. Überzeugungen zum Lernen mit digitalen Medien (Kotzebue, 2022a) sowie motivationale Bedingungen wie die TPACK-Selbstwirksamkeit und der wahrgenommene Nutzwert digitaler Medien (Backfisch et al., 2020a; Backfisch et al., 2021) als weitere relevante Einflussfaktoren neben TPACK auf die Qualität digital-gestützten Unterrichts bzw. von Unterrichtsplanungen heraus. Zudem wird TPACK im Kontext des Technologieakzeptanz-Modells (Davis et al., 1989) untersucht, inwiefern das Wissen die Verhaltensabsicht, digitale Medien zu nutzen, beeinflusst. Dabei zeigen sich indirekte Zusammenhänge beispielsweise über die wahrgenommene Leichtigkeit der Mediennutzung (Joo et al., 2018; Mayer & Girwidz, 2019), die wahrgenommene persönliche Nützlichkeit sowie die Selbstwirksamkeit der Lehrkräfte (Joo et al., 2018). Max et al. (2023) stellten fest, dass TPACK mit der Technologieakzeptanz sowie der intrinsischen Motivation, digitale Medien zu nutzen, zusammenhängt und dass ein höheres TPACK nach ihrer untersuchten Intervention einen positiven Effekt auf die Verhaltensabsicht hat, digitale Medien zu nutzen. In ähnlicher Weise sprechen signifikante Korrelationen zwischen TPACK und Variablen wie den Einstellungen, der Selbstwirksamkeitserwartung sowie der motivationalen Orientierung gegenüber dem Einsatz digitaler Medien bei Stinken-Rösner (2021a) im Sinne der *Theory of Planned Behaviour* (Ajzen, 1991) dafür, dass TPACK einen Einfluss auf die Verhaltensabsichten hinsichtlich der Nutzung digitaler Medien hat.

Die konzeptuelle Rahmung zum TPACK-Modell sowie die beschriebenen Ergebnisse legen nahe, dass neben anderen relevanten Faktoren wie Überzeugungen, motivationalen Bedingungen und wahrgenommener Nützlichkeit auch das TPACK eine bedeutsame Rolle für den unterrichtlichen Einsatz digitaler Medien zu spielen scheint. Dementsprechend ist es wichtig, die Entwicklung von TPACK in der Lehrkräftebildung zu fördern (z. B. Max et al., 2023). Im Folgenden werden nun für diese Arbeit relevante Forschungsergebnisse aufgeführt, welche diese Entwicklung von TPACK im Lehramtsstudium bzw. einzelnen Lehrveranstaltungen in den naturwissenschaftlichen Fächern (4.2.1) sowie Zusammenhänge solcher Lerngelegenheiten mit der Entwicklung des TPACK (4.2.2) untersuchen. Die in diesen Studien verwendeten Testverfahren wurden in 3.5 vorgestellt.

4.2.1 Entwicklung des TPACK im Lehramt für naturwissenschaftliche Fächer

Im Hinblick auf Studien, die die Entwicklung von TPACK im Lehramtsstudium im deutschsprachigen Raum untersuchten, sind zum Zeitpunkt der Verfassung dieser Arbeit keine physikspezifischen Studien mit passender Ausrichtung bekannt. Zwar gibt es physikspezifische Studien in dem Forschungsbereich; diese fokussieren in ihren Erhebungen jedoch andere Schwerpunkte wie z. B. die Einstellungen in Bezug auf die Vermittlung digitaler Kompetenzen (Walpert & Wodzinski, 2023) oder den Einsatz digitaler Medien im Unterricht von Physiklehramtsstudierenden (Schaber & Friege, 2023). Daher wird im Folgenden auf Ergebnisse aus Studien zur Untersuchung des TPACK bei Lehramtsstudierenden in den naturwissenschaftlichen Fächern eingegangen, welche aufgrund ihrer Nähe zum Fach Physik ebenfalls als relevant anzusehen sind. In diesen Studien wurden Lerngelegenheiten zur Förderung digitalisierungsbezogener Kompetenzen bei Lehramtsstudierenden evaluiert und dabei unter anderem die Entwicklung des TPACK bzw. von TPACK-Bereichen der Teilnehmenden in Prä-Post-Erhebungen untersucht.

In der Studie von Henne et al. (2022) wurde ein Seminar zur Förderung digitaler Kompetenzen von Lehramtsstudierenden der Naturwissenschaften in

Anlehnung an den DiKoLAN-Orientierungsrahmen entwickelt (2.3.1). Das Seminar setzte sich aus mehreren Phasen zusammen. Nach einer Einführung zum Lernen mit digitalen Medien in den Naturwissenschaften folgten in der ersten Phase wöchentliche Module zu jedem der sieben Kompetenzbereiche im Orientierungsrahmen. In der zweiten (asynchronen) Phase entwickelten die Teilnehmenden Unterrichtsszenarien, die sie in der dritten Phase präsentierten und in einer Erprobung implementierten (Henne et al., 2022). Für die einzelnen Module zu den Kompetenzbereichen wurden Kompetenzerwartungen aus dem Orientierungsrahmen als Hauptlernziele und Nebelernziele identifiziert, welche in dem Modul gefördert werden sollten. Die Evaluation der Lehrveranstaltung erfolgte im Hinblick auf die Selbstwirksamkeitserwartungen bzgl. TPACK mithilfe des Selbsteinschätzungsbogens zu den in DiKoLAN formulierten Kompetenzanforderungen nach Kotzebue et al. ((2021); 3.5.1). Dazu wurden wöchentliche Prä-Post-Erhebungen für die einzelnen Module zu den sieben Kompetenzbereichen durchgeführt, wobei jeweils nur die Items für den in der Woche adressierten Kompetenzbereich erhoben wurden. Insgesamt konnten 13 an der Lehrveranstaltung teilnehmende Lehramtsstudierende der naturwissenschaftlichen Fächer befragt werden, davon eine Person mit dem Fach Physik (Henne et al., 2022). Erwartungskonform zeigten sich über alle Kompetenzbereiche signifikante Zuwächse (großer bzw. mittlerer Effektstärke) hinsichtlich der TPACK-Selbstwirksamkeit für die in Haupt- bzw. Nebelernzielen adressierten Kompetenzerwartungen sowie keine Zuwächse in den Kompetenzerwartungen, die nicht gefördert wurden (Henne et al., 2022). Bezüglich der beiden fachspezifischeren Kompetenzbereiche Messwert- und Datenerfassung sowie Simulation und Modellierung wurden dabei kaum Haupt- oder Nebelernziele im Bereich TPACK adressiert (Henne et al., 2022). Dies reduziert die Relevanz der Befundlage in dieser Studie für die vorliegende Arbeit, da diese Bereiche dafür am relevantesten wären. Weiterhin ist anzumerken, dass die Intervention und das genutzte Erhebungsinstrument sehr eng aufeinander abgestimmt waren und somit die großen Effekte und erwartungstreuen Ergebnisse wenig überraschend sind.

Anders als in der zuvor vorgestellten Studie, wurde in der Studie von Stinken-Rösner et al. (2023) keine Lehrveranstaltung zur Förderung digitaler Kompetenzen neu entwickelt, sondern bestehende Lehrveranstaltungen im Lehramt in Naturwissenschaften im Hinblick auf den Einsatz digitaler Medien im Unterricht erweitert. Konkret reicherten Stinken-Rösner et al. (2023) zwei fachdidaktische Module, die Lehramtsstudierende der naturwissenschaftlichen Fächer gemeinsam besuchen, durch „digitale Ergänzungen“ an. So wurde im ersten Modul, das zentrale Themen der Naturwissenschaftsdidaktiken behandelt, beispielsweise das Thema Erklären um den Aspekt Erklärvideos oder das Thema Experimente um interne und externe Sensoren sowie interaktive Simulationen ergänzt (Stinken-Rösner, 2021b; Stinken-Rösner et al., 2023). Das zweite Modul adressierte die praktische Anwendung von digitalen Medien, indem die Seminaraufgabe, eine Unterrichtseinheit zu planen, durchzuführen und zu reflektieren, um die Integration digitaler Medien in diese Planung erweitert wurde (Stinken-Rösner, 2021b). Die Evaluation dieser Anreicherung der Module erfolgte in zwei Kohorten mit insgesamt 133 Studierenden, von denen niemand angab, Physik zu studieren (Stinken-Rösner et al., 2023). Dabei wurde zu drei Zeitpunkten das selbsteingeschätzte TPACK mithilfe des für den Kontext angepassten Instruments nach Chai et al. (2013b) erhoben (Stinken-Rösner, 2021a; 3.5.1). Aus der Stichprobe nahmen 102 Studierende am Prätest (vor dem ersten Modul), 76 am Re-Test (vor dem zweiten Modul) und 55 am Posttest (nach dem zweiten Modul) teil. Die Entwicklung des selbsteingeschätzten TPACK über die beiden Module ist über alle TPACK-Bereiche durchweg positiv mit mittleren bis großen Effektstärken. Für die Selbsteinschätzung zum Wissensbereich TPACK zeigt sich jeweils ein signifikanter Anstieg über den Besuch des ersten ($r = 0,34$) und des zweiten ($r = 0,61$) Moduls sowie ein signifikanter Anstieg über beide Module ($r = 0,80$). Insbesondere scheint der Besuch des zweiten Moduls zur Vertiefung in der Praxis hilfreich (Stinken-Rösner et al., 2023).

Die Studien von Max et al. (2022) und Max et al. (2023) beschreiben die Evaluation einer Lerngelegenheit zur Förderung des TPACK angehender Lehrkräfte auf Basis eines pädagogischen Makerspace. In diesem Makerspace wurde eine medienbasierte Projektarbeit durchgeführt, in welcher die Teilnehmenden kollaborativ und kreativ arbeiteten, um Projektprodukte wie

z. B. multimediale und interaktive Lehr-Lernmaterialien zu entwickeln. Einen weiteren Teil der untersuchten Intervention bildete vor der Projektarbeit im Makerspace eine theoretische Einführung zu fachdidaktischen und medienpädagogischen Grundlagen (CLT: Sweller, 1994; CTML: Mayer, 2009; SAMR-Modell: Puentedura, 2006; 2.3.3) sowie zu Einsatzmöglichkeiten digitaler Medien im naturwissenschaftlichen Unterricht (Max et al., 2022). Zur Evaluation dieser Lerngelegenheit wurde einerseits ein TPACK-Selbsteinschätzungsinstrument basierend auf Schmidt et al. (2009) sowie ein entwickelter Leistungstest zur proximalen Messung von TPACK (3.5.2) verwendet. Die Testaufgaben des (biologiespezifischen) Leistungstests berücksichtigen dabei auf technologischer Ebene mit je einer Aufgabe die drei Technologien Erklärvideos, digitale Messwerterfassung und Simulationen sowie auf lernpsychologischer Ebene in jeweils drei Teilaufgaben die Reproduktion von Wissen, die Anwendung auf eine bestimmte Situation sowie die Diskussion von Vor- und Nachteilen (Max et al., 2020, 2022; 3.5.2).

Aus einer Stichprobe von 221 Teilnehmenden zählten 149 zur Interventionsgruppe im Makerspace und 72 zu einer Kontrollgruppe, die hinsichtlich ihres Studienfortschritts vergleichbar war, aber weder die theoretische Einführung noch die Projektarbeit im Makerspace absolvierten. Die Evaluationsergebnisse zeigten, dass nur die Lehramtsstudierenden der Interventionsgruppe einen signifikanten Zuwachs hinsichtlich des selbsteingeschätzten TPACK vermerkten (Max et al., 2022). In einer weiterführenden Studie konnte weiterhin die Vermutung bestätigt werden, dass die Einschätzung im Prätest die Entwicklung des selbsteingeschätzten TPACK über die Intervention beeinflusst. Studierende, die sich vor dem Kurs besser bzgl. ihres TPACK einschätzten, hatten scheinbar auch den Eindruck, mehr von dem Kurs zu profitieren (Max et al., 2023). Für den Leistungstest konnten 77 vollständige Prä-Post-Datensätze in der Interventionsgruppe erhoben werden. Hier zeigten die Ergebnisse einen signifikanten Zuwachs mittlerer Effektstärke in den beiden Teilaufgaben zum *Reproduzieren* und *Diskutieren* von Vor- und Nachteilen sowie in der Gesamtleistung, jedoch keine signifikante Veränderung in der Teilaufgabe zum *Anwenden* von TPACK (Max, 2022; Max et al., 2022).

Zimmermann (2021) entwickelte in ihrer Arbeit ein Seminar zur Förderung digitaler Kompetenzen und zur Vorbereitung auf den Einsatz digitaler Werkzeuge für Masterstudierende im Lehramt Chemie. Dabei handelte es sich um ein Vorbereitungsseminar zum Praxissemester, welches demnach ähnlich verortet ist wie die untersuchten physikdidaktischen Lerngelegenheiten in der Arbeit von Joswig-Käfer (2024; 4.1.1). Das chemiedidaktische Seminar beinhaltete Themenblöcke zu Grundlagen des Einsatzes sowie der unterrichtspraktischen Implementation digitaler Werkzeuge, zum Experimentieren mit digitalen Werkzeugen und zu methodischen Grundlagen für die Implementation (Zimmermann, 2021). Insgesamt wurden 26 Lehramtsstudierende der Chemie vor und nach dem Seminar sowie nach dem Praxissemester (Follow-up) befragt. Eine Gruppe von 18 Chemielehramtsstudierenden einer anderen Universität, die ein Vorbereitungsseminar zum Praxissemester ohne Bezug zu digitalen Medien besuchten, diente als Vergleichsgruppe. Erhoben wurde unter anderem die Selbstwirksamkeit der Studierenden hinsichtlich ihres TPACK mit dem Selbsteinschätzungsinstrument nach Schmidt et al. (2009; 3.5.1). Weiterhin wurde die Fähigkeit, digitale Werkzeuge in die Unterrichtsplanung zu integrieren, über die Bewertung von Unterrichtsplanungen der Studierenden mithilfe eines Kodiermanuals untersucht, das Indikatoren zu sieben Unterrichtsschritten jeweils für die Bereiche PCK, TK, TCK und TPK berücksichtigt (Zimmermann, 2021; 3.5.2). Die Studie zeigte, dass das entwickelte Seminar zu einer unmittelbaren sowie langfristigen Steigerung der TPACK-Selbstwirksamkeit der Teilnehmenden führte, da die Interventionsgruppe dort signifikante Zuwächse großer Effektstärke vermerkte, während in der Vergleichsgruppe keine Veränderung auftrat. Ebenfalls zeigte sich über das Seminar eine signifikante Steigerung in der Fähigkeit, digitale Werkzeuge in die Unterrichtsplanung zu integrieren hinsichtlich der untersuchten Wissensbereiche PCK, TK, TCK und TPK – ebenfalls mit großen Effektstärken (Zimmermann, 2021).

Weiterhin wurde die Wirksamkeit des entwickelten Seminarkonzepts in zwei veränderten Kontexten untersucht (Zimmermann & Melle, 2023; Zimmermann et al., 2021). Zum einen implementierten Zimmermann et al. (2021) das Seminar für eine andere Zielgruppe, und zwar für Bachelorstudierende im

Lehramt Chemie, die keine Grundkenntnisse aus vorausgehenden fachdidaktischen Lerngelegenheiten mitbrachten und kein an das Seminar anknüpfendes Schulpraktikum oder Praxissemester besuchten. Die Ergebnisse dieses Vergleichs zeigten, dass auch die Bachelorstudierenden hinsichtlich ihrer TPACK-Selbstwirksamkeit sowie ihren Fähigkeiten in der Unterrichtsplanung bezogen auf PCK, TK, TCK und TPK von dem Seminar profitierten, jedoch signifikant weniger als die Interventionsgruppe der Masterstudierenden. Das entwickelte Seminar scheint demnach effektiver zu sein, wenn die Teilnehmenden bereits gewisse Grundkenntnisse im Professionswissen aufweisen. Weiterhin scheint es auch gewinnbringend, das Seminar vorausgehend einer unterrichtspraktischen Erprobung einzubetten, um die Theorie-Praxis-Verknüpfung im Seminar zu verstärken (Zimmermann, 2021; Zimmermann et al., 2021). Zum anderen wurde das Seminar pandemiebedingt in einem Durchgang als Online-Seminar adaptiert und mit den regulären Durchgängen aus der Hauptstudie verglichen. Hier zeigte sich, dass auch über das Online-Seminar die TPACK-Selbstwirksamkeit sowie die Fähigkeiten in der Unterrichtsplanung in Bezug auf TK, TCK und TPK gesteigert werden konnten (jedoch nicht für PCK). Das Seminarkonzept eignete sich demnach auch im Online-Format zur Förderung des TPACK der angehenden Lehrkräfte. Da das Präsenzformat noch lernwirksamer war, ist dieses dennoch zu bevorzugen (Zimmermann, 2021; Zimmermann & Melle, 2023).

Insgesamt zeigt sich durch die vorgestellten Studienbeispiele, dass durch fachdidaktische Lerngelegenheiten an der Universität TPACK bzw. TPACK-Selbstwirksamkeit gefördert werden kann. In diesen Studien werden dabei zwei unterschiedliche Ansätze zur Förderung von TPACK deutlich: einerseits als eigenständiges Seminar, das sich mit dem Einsatz digitaler Medien im Unterricht beschäftigt und im Rahmen der Studien neu entwickelt wurde (z. B. Henne et al., 2022; Max et al., 2022; Zimmermann, 2021) und andererseits als Anreicherung bestehender fachdidaktischer Lerngelegenheiten im Hinblick auf digitale Medien (Stinken-Rösner et al., 2023).

In Bezug auf die Entwicklung des TPACK stellt sich weiterhin die Frage, welche Merkmale das TPACK von (angehenden) Lehrkräften beeinflussen. In der Studie von Max et al. (2023) beeinflussten die intrinsische Motivation digitale Medien zu nutzen sowie die Technologieakzeptanz (selbsteingeschätztes)

TPACK vor der untersuchten Intervention, welches wiederum einen relevanten Prädiktor für das TPACK nach der Intervention darstellte. Zimmermann (2021) konnte in ihrer Arbeit einen Einfluss des Geschlechts im zweiten Messzeitpunkt (d. h. nach Besuch des untersuchten Seminars) feststellen. Weibliche Studierende zeigten im Posttest signifikant höhere Fähigkeiten in der Unterrichtsplanung unter Einbindung digitaler Medien in den technologiebezogenen Bereichen TK, TCK und TPK. Zudem konnte ihre TPACK-Selbstwirksamkeit über das Seminar signifikant mehr gesteigert werden als die ihrer männlichen Kommilitonen (Zimmermann, 2021). Weiterhin zeigte sich ein Zusammenhang zwischen den Vorerfahrungen der Studierenden im Umgang mit digitalen Medien und der TPACK-Selbstwirksamkeit. Für die Fähigkeiten in der Unterrichtsplanung zeigte sich ein solcher Zusammenhang nur für den Bereich TCK, nicht aber für TK, TPK und PCK. Ebenso wenig hing die TPACK-Selbstwirksamkeit mit der Fähigkeit zur Unterrichtsplanung (in allen Bereichen) zusammen (Zimmermann, 2021). Kotzebue (2022a) untersuchte in ihrer Studie den Einfluss von Überzeugungen zum Lernen mit digitalen Medien sowie von studienbezogenen Einflussfaktoren wie dem Fachsemester und bisheriger Lehrerfahrung auf das selbsteingeschätzte sowie proximal gemessene TPACK. Dabei zeigte das Fachsemester der Studierenden zwar einen Einfluss auf die TPACK-Selbsteinschätzung, nicht aber auf das Ergebnis im Leistungstest. Sowohl die bisherige Lehrerfahrung als auch die Überzeugungen hatten auf beide Arten der TPACK-Ausprägung keinen Einfluss (Kotzebue, 2022a).

Es zeigen sich insgesamt verschiedene Merkmale, die die TPACK-Ausprägung einer Person möglicherweise beeinflussen. Allerdings scheinen die Einflussfaktoren unterschiedlich, je nachdem, ob es sich um leistungstestbasierte Erfassung von TPACK-Bereichen handelt oder um eine Selbsteinschätzung von TPACK (bzw. TPACK-Selbstwirksamkeit). Dies stützt wiederum die Annahme der *jingle-jangle-fallacy*, dass mit den beiden Messverfahren – leistungstestbasiert oder über Selbsteinschätzungen – nicht dasselbe Konstrukt erfasst wird (3.5.3).

Neben Personenmerkmalen oder globalen Studienmerkmalen, die die Ausprägung oder Entwicklung von TPACK beeinflussen, ist für die Lehrkräftebil-

dungsforschung zudem von Interesse, welche Lerngelegenheiten oder Merkmale von Lerngelegenheiten die Entwicklung von TPACK begünstigen. Im folgenden Abschnitt werden daher Zusammenhänge zwischen Lerngelegenheiten und der Entwicklung hinsichtlich des TPACK beschrieben.

4.2.2 Zusammenhänge von Lerngelegenheiten und TPACK

Die Lerngelegenheiten der zuvor beschriebenen Studien (4.2.1) integrieren meist einen Bezug zur Schulpraxis, indem die Studierenden z. B. Unterricht unter Einbettung digitaler Medien planen und erproben (Henne et al., 2022; Max et al., 2022; Stinken-Rösner et al., 2023; Zimmermann, 2021). Aufgrund positiver Evaluationsergebnisse wird diesem Praxisbezug daher eine wichtige Rolle zur Förderung von TPACK bzw. FDW zum Einsatz digitaler Medien zugeschrieben (Max et al., 2022; Stinken-Rösner et al., 2023; Zimmermann, 2021). Gleichzeitig können Ursachen für über bestimmte Lehrveranstaltungen gemessene Wissensveränderungen nicht unbedingt auf konkrete Lehrelemente zurückgeführt werden.

In der Arbeit von Zimmermann (2021) spricht die Vergleichsstudie mit einer anderen (unerfahreneren) Zielgruppe zumindest dafür, dass ein gewisses Grundlagenwissen in den Professionswissensbereichen Fachwissen, pädagogisches Wissen und FDW sowie die Anknüpfung einer Praxisphase an die Lehrveranstaltung sinnvoll scheint, um effektiv davon zu profitieren (Zimmermann, 2021; Zimmermann et al., 2021). Ebenso ist nach den Ergebnissen einer weiteren Vergleichsstudie ein Präsenzformat gegenüber eines Online-Formats der Lerngelegenheit zu bevorzugen (Zimmermann, 2021; Zimmermann & Melle, 2023). Weiterhin empfiehlt Zimmermann (2021) die bewusste Verknüpfung fachdidaktischer und digitalisierungsbezogener Inhalte in Lerngelegenheiten zum Einsatz digitaler Medien (anstelle von reinen Technologiekursen), da in ihrer Studie neben der positiven Entwicklung in den technologiebezogenen Bereichen TK, TCK und TPK auch das PCK bzw. FDW der Studierenden gesteigert werden konnte (Zimmermann, 2021; Zimmermann et al., 2021). Ähnlich betonen Schmid et al. (2020), dass technologisches Wissen nicht isoliert gefördert werden sollte. Stattdessen sollte das Ziel sein, ein integriertes Wissen aufzubauen, indem bestehende fachdidaktische Aspekte berücksichtigt und um eine technologische Perspektive erweitert werden

(Schmid et al., 2020). Ebenso fassen Backfisch et al. (2021) aus ihren Studien mit (angehenden) Lehrkräften zusammen, dass vorwiegend (fach-)didaktisches Wissen und weniger das technologische Wissen eine Voraussetzung für die Verfügbarkeit von TPACK darstellt.

In der Synthese von Tondeur et al. (2012) wurden 19 verschiedene qualitative Studien ausgewertet und untersucht, die Strategien zur Vorbereitung angehender Lehrkräfte auf den Einsatz digitaler Medien im Unterricht fokussierten. Dazu zählten Studien, die überwiegend angehende Lehrkräfte untersuchten und verschiedene Methoden wie qualitative Interviews, Fallstudien, Beobachtungen und Inhaltsanalysen von Dokumenten nutzten. Als Synthese zentraler Studienergebnisse wurden verschiedene Schlüsselaspekte auf Ebene der Lehrkonzeptgestaltung sowie institutioneller Rahmenbedingungen herausgestellt und im *Synthesis of Qualitative Evidence* (SQD) Modell zusammengefasst (Tondeur et al., 2012). Die Schlüsselaspekte auf Lehrkonzeptebene geben damit Hinweise darauf, welche Elemente von Lerngelegenheiten nützlich zur Förderung von TPACK sind.

So stellte sich in 14 der 19 Studien die Relevanz einer gelungenen Einbettung der Theorie in die Praxis heraus, damit angehende Lehrkräfte Gründe für den Einsatz digitaler Medien im Unterricht verstehen. Gemäß der Studienergebnisse zeigten Lehramtsstudierende wenig Interesse an stark theoretischen Kursen in Bezug auf digitale Medien, sondern waren vielmehr daran interessiert, Möglichkeiten für den Einsatz bestimmter Technologien im Unterricht aufgezeigt zu bekommen (Tondeur et al., 2012). Die Bedeutsamkeit solcher Demonstrationen spiegelte sich im zweiten Schlüsselaspekt wider, da 13 von 19 Studien nahelegten, dass die Lehrperson der untersuchten Kurse den Teilnehmenden als Vorbild hinsichtlich der Nutzung digitaler Medien diene. Die Beobachtung der Mediennutzung durch die Kursleitung führte dann zur Adaption und eigenen Nutzung der beobachteten Medien und Strategien (Tondeur et al., 2012). Ebenso stellte sich die Reflexion hinsichtlich der Rolle digitaler Medien für das Lehren und Lernen als wiederkehrender Aspekt in 13 Studien heraus. Dabei waren Austausch und Diskussionen unter den Teilnehmenden und mit der Lehrperson sowie Beobachtungen und schriftliche Reflexionen hilfreich, um über die eigenen Einstellungen diesbezüglich re-

flektieren zu können. In sechs der ausgewählten Studien war zudem die eigene Gestaltung von Unterricht und Erstellung von Materialien ein relevantes Thema und wurde von den Studierenden als gewinnbringend beschrieben (Tondeur et al., 2012). In diesem Zusammenhang deutete sich auch die Zusammenarbeit mit Mitstudierenden, z. B. durch gegenseitiges Vorstellen und Bewerten erstellter Materialien (Jang, 2008, zitiert nach Tondeur et al., 2012) oder gegenseitiges Unterstützen, als förderlich an, was wiederum insgesamt in neun Studien vorkam und einen weiteren Schlüsselaspekt darstellte. Die Anwendung ihres technologiebezogenen Wissens in authentischen Erfahrungen wurde in 13 der 19 Studien als relevanter Aspekt aus Sicht der Studierenden deutlich. So führte diese Anwendung des Wissens im Unterricht oder unterrichtlichen Situationen zu Erfolgserlebnissen bei den Teilnehmenden (Barton und Haydn, 2006, zitiert nach Tondeur et al., 2012). Dabei stellten sich Unterstützungsmaßen z. B. zur Planung und Vorbereitung als bedeutsam für solche Erfolge heraus (Brush et al., 2003, zitiert nach Tondeur et al., 2012). Der letzte Schlüsselaspekt (auf Lehrkonzeptebene) greift die in neun Studien thematisierte fehlende Passung zwischen traditionellen Bewertungen technologiebezogener Kompetenzen und notwendigen Aspekten zum erfolgreichen Medieneinsatz im Unterricht auf. Empfohlen wird stattdessen, die Fähigkeiten und Einstellungen der angehenden Lehrkräfte bezüglich der Nutzung digitaler Technologien kontinuierlich und prozessbezogen zu evaluieren, z. B. durch Diskussionen, Beobachtungen, Interviews oder Portfolios (Tondeur et al., 2012).

Die beschriebenen qualitativen Studienergebnisse haben Tondeur et al. (2012) als Schlüsselaspekte im SQD-Modell zusammengefasst. Diese Schlüsselaspekte stellen relevante Elemente für Lerngelegenheiten dar, in denen angehende Lehrkräfte auf den Einsatz digitaler Medien im (zukünftigen) Unterricht vorbereitet werden und im Hinblick auf ihr TPACK gefördert werden sollen. Das Modell ist somit eine hilfreiche (methodische) Grundlage für die Entwicklung entsprechender Lerngelegenheiten im Lehramtsstudium und wurde auch für die Entwicklung der in dieser Arbeit untersuchten Lehrveranstaltungen herangezogen (5.2.1). Das Modell ist jedoch fachunspezifisch und benennt keine inhaltlichen Schwerpunkte oder fachspezifische Beson-

derheiten, da die zugrundeliegenden Studien aus verschiedenen Fachbereichen oder fachübergreifenden Kontexten stammen. Konkrete Hinweise auf fachspezifischer Ebene, wie TPACK oder FDW zum Einsatz digitaler Medien bei Physiklehramtsstudierenden gefördert werden kann, sind demnach noch offen.

Im folgenden Abschnitt werden nun die Erläuterungen des Forschungsstands zum FDW und TPACK aus diesem Kapitel zusammengefasst.

4.3 Zusammenfassung des relevanten Forschungsstands zum FDW und TPACK

In diesem Kapitel wurde der für diese Arbeit zur Untersuchung des FDW zum Einsatz digitaler Medien relevante Forschungsstand beschrieben. So wurden einerseits ausgewählte Forschungsergebnisse zum (allgemeinen) FDW (4.1) sowie andererseits zum – mit dem FDW zum Einsatz digitaler Medien vergleichbaren – (digitalisierungsbezogenen) TPACK (4.2) erläutert. Deutlich zeigt sich, dass sowohl FDW als auch TPACK bedeutsam für die Professionalisierung von Lehrkräften sind und in diesem Feld vielfältig beforscht werden. So wird in der bisherigen Forschung die Rolle des FDW für das Unterrichtshandeln und Unterrichtsqualität sowie zum Teil für den Lernfortschritt der Schülerinnen und Schüler aufgezeigt (z. B. Baumert et al., 2010; Blömeke et al., 2022; Riese et al., 2022b; 4.1). Auch Studien zum TPACK stellen den Einfluss des Wissens auf den Unterrichtseinsatz digitaler Medien dar (4.2). Dabei werden vor allem auch motivationale Aspekte einbezogen und TPACK als (indirekter) Einflussfaktor auf die Motivation und Bereitschaft digitale Medien zu nutzen untersucht (z. B. Mayer & Girwidz, 2019; Stinken-Rösner, 2021a).

Der jeweils beschriebene Forschungsstand zur Entwicklung von FDW bzw. TPACK im Lehramtsstudium Physik (bzw. in den Naturwissenschaften) verdeutlicht, dass mit den vorgestellten Messverfahren Wissensveränderungen über das Lehramtsstudium oder einzelnen Lehrveranstaltungen messbar sind und Lehramtsstudierende ihr FDW bzw. TPACK über entsprechende Lerngelegenheiten verbessern können (4.1.1 bzw. 4.2.1). Während manche Untersuchungen zum FDW bereits in Längsschnitten über größere Zeitspannen erfolgen (Riese et al., 2022a; Sorge et al., 2018; 4.1.1), zeigt sich in der

Forschung zum TPACK noch die Tendenz, einzelne Lerngelegenheiten mit dem Schwerpunkt auf digitale Medien über ein Semester zu evaluieren (Henne et al., 2022; Max et al., 2022; Zimmermann, 2021; 4.2.1). Weiterhin haben sich bereits mehrere Forschungsprojekte der fachspezifischen Untersuchung des FDW von (angehenden) Physiklehrkräften gewidmet, von denen in 4.1.1 drei exemplarisch aufgegriffen wurden (KiL/KeiLa: Sorge et al., 2018; Sorge et al., 2019; ProfiLe-P/P+: Riese et al., 2017; Riese et al., 2022a und Joswig-Käfer, 2024). Eine fachspezifische Untersuchung zur Entwicklung des TPACK von Physiklehramtsstudierenden (im deutschsprachigen Raum) konnte hingegen nicht gefunden werden, sodass an dieser Stelle der Fokus auf Forschungsergebnisse aus den naturwissenschaftlichen Fächern ausgeweitet werden musste (4.2.1).

Die gemessenen Wissensveränderungen über das Lehramtsstudium oder einzelne Lerngelegenheiten zu FDW oder TPACK deuten zwar auf eine entsprechende Wirksamkeit hin, jedoch können Ursachen für die Veränderungen durch quantitative Befragungen nicht immer auf konkrete Lehrelemente zurückgeführt werden. Daher sind (qualitative) Ergänzungsstudien, die Zusammenhänge zwischen Lerngelegenheiten und der Entwicklung der Wissensbereiche erkunden, von besonderem Interesse, um die Wirksamkeit der Lerngelegenheiten zu stützen oder lernwirksame Elemente herauszustellen. So konnten Schiering et al. (2021) mit ihrer Studie bedeutsame kategorielle Zusammenhänge von Ausbildungsprofilen und der Kompetenzentwicklung im FDW (und Fachwissen) in Physik aufzeigen, die über zuvor bekannte globale Zusammenhänge mit dem Fachsemester hinausgehen. Da die Studie auf die ersten sechs Semester beschränkt ist, sind die Erkenntnisse für die fachdidaktische Ausbildung und das FDW jedoch begrenzt (4.1.2). Joswig-Käfer (2024) identifizierte in ihrer Arbeit lernförderliche Lernelemente im Hinblick auf die Entwicklung des FDW über zwei physikdidaktische Masterseminare. Hier erwiesen sich die Seminarthemen wie „Schülervorstellungen“, „Experimente“ oder „Lernzirkel“ sowie die Vorgehensweise, sich in die Lehrperson oder die Lernenden hineinzusetzen, als Elemente, die häufig mit Verbesserungen im Test einhergingen (4.1.2); jedoch lag der Fokus in der Arbeit nicht auf dem Einsatz digitaler Medien. Für Lerngelegenheiten zum TPACK scheint ein Praxisbezug (z. B. Max et al., 2022; Stinken-Rösner et al., 2023;

Tondeur et al., 2012; Zimmermann, 2021), das eigene Planen und Gestalten von Unterrichtsszenarien und -materialien und die Zusammenarbeit mit Mitstudierenden (Tondeur et al., 2012) sowie der Aufbau auf fachdidaktischen Grundlagenwissen (Backfisch et al., 2021; Schmid et al., 2020; Zimmermann, 2021) hilfreich (4.2.2). Die bisherigen Ergebnisse beziehen sich jedoch nicht auf Lehramtsstudierende im Fach Physik.

Zusammenfassend wird demnach deutlich, dass hinsichtlich der Untersuchung zur Entwicklung von FDW zum Einsatz digitaler Medien bei Lehramtsstudierenden der Physik bereits relevante Forschungsergebnisse vorliegen, die einerseits die fachspezifische Untersuchung des FDW adressieren (jedoch ohne Fokus auf digitale Medien) oder andererseits TPACK in den Blick nehmen (jedoch nicht spezifisch für das Fach Physik). Wie sich das FDW zum Einsatz digitaler Medien bei Physiklehramtsstudierenden entwickelt und was für diese Entwicklung förderlich sein kann, ist demnach noch offen.

Im folgenden Kapitel werden die entwickelten und untersuchten Lerngelegenheiten im Verbundprojekt DiKoLeP beschrieben, welche den Untersuchungskontext dieser Arbeit darstellen. Kapitel 6 beinhaltet anschließend die Formulierung der Ziele und Forschungsfragen der vorliegenden Arbeit, welche sich durch abgeleitete Desiderate aus dem beschriebenen Forschungsstand ergeben.

5 Lerngelegenheiten im Verbundprojekt DiKoLeP

Dieses Kapitel beschreibt die untersuchten Lerngelegenheiten im Verbundprojekt DiKoLeP, die den Kontext der vorliegenden Studie darstellen. In 5.1 wird zunächst kurz das Verbundprojekt vorgestellt. Anschließend erfolgt in 5.2 die Erläuterung der fachdidaktischen Lehrveranstaltungen im Projekt, wobei das übergreifende Lehrkonzept aller Lehrveranstaltungen (5.2.1) sowie die Spezifika der einzelnen Standorte (5.2.2 und 5.2.3) beschrieben werden.

5.1 Das Verbundprojekt DiKoLeP im Überblick

Das Projekt DiKoLeP („Digitale Kompetenzen von Lehramtsstudierenden im Fach Physik“) ist ein Verbundprojekt der Universitäten Graz, Innsbruck, Paderborn und Tübingen und zielt auf die Entwicklung und Evaluation fachdidaktischer Lerngelegenheiten zur Förderung physikdidaktischer digitaler Kompetenzen wie dem FDW zum Einsatz digitaler Medien im Physikunterricht ab (Schubatzky et al., 2022). Zu Beginn des Projekts war die RWTH Aachen als Kooperationspartner beteiligt, sodass auch an diesem Standort eine Lerngelegenheit im Verbundprojekt untersucht wird.

Ein erstes Teilziel stellt zunächst die Entwicklung von Seminarkonzepten zum Einsatz digitaler Medien dar, die als gemeinsames Lehrkonzept abgestimmte Kerninhalte sowie standortspezifische Praxisanteile berücksichtigen. Ein zweites Teilziel bildet die standortübergreifende Evaluation des abgestimmten Lehrkonzepts. Als drittes Teilziel werden Zusammenhangsanalysen im Kontext digitaler Kompetenzen von Physiklehramtsstudierenden adressiert. Im Sinne dieser Teilziele verfolgt das Projekt verschiedene Forschungsinteressen, wie beispielsweise eine valide Messung des FDW zum Einsatz digitaler Medien (Große-Heilmann et al., 2022), die Untersuchung eines Kompetenzzuwachses über die beteiligten Seminare (Große-Heilmann et al., 2024), die Identifikation von Einstellungstypen hinsichtlich digitaler Medien (Weiler et al., 2024), die Identifikation von lernförderlichen und -hinderlichen Seminarelementen (Große-Heilmann et al., 2024) sowie die Untersuchung der Rolle des FDW zum Einsatz digitaler Medien im Hinblick auf die Motivation, digitale Medien im Unterricht zu nutzen (Schubatzky et al., 2023).

Die entsprechenden Untersuchungen im Verbundprojekt sollen schließlich zum übergeordneten Hauptziel des Projekts beitragen, welches in der Generierung von Hypothesen zu effektiven Lernangeboten in Physik-Lehramtsstudiengängen zum Einsatz digitaler Medien besteht (Große-Heilmann et al., 2024).

Dazu wurde im Projekt ein übergreifendes Lehrkonzept entwickelt, das gemeinsame Kerninhalte zum Einsatz digitaler Medien sowie standortspezifische Praxisanteile beinhaltet. Die Entwicklung des Lehrkonzepts erfolgte unter Berücksichtigung einer im Verbundprojekt durchgeführten Bedürfnisanalyse zu Vorerfahrungen, Interessen und Einstellungen zum Einsatz digitaler Medien bei Physiklehramtsstudierenden (Weiler et al., 2022a). Weiterhin erfolgten Interviews mit Fachdidaktiker:innen zur Einschätzung der ersten Gestaltung einzelner Seminareinheiten (zu verschiedenen digitalen Medien), um das Lehrkonzept und die besprochenen Inhalte weiterzuentwickeln (Weiler et al., 2023; Weiler et al., 2022b). Dieses Lehrkonzept wurde an den beteiligten Universitäten unter Berücksichtigung standortspezifischer Rahmenbedingungen in Form von fachdidaktischen Seminaren implementiert. Das Lehrkonzept sowie die Ausgestaltung der Praxisanteile an den verschiedenen Standorten werden in 5.2 erläutert.

In der Begleitforschung zur Evaluation des Lehrkonzepts im Hinblick auf den Kompetenzerwerb werden unterschiedliche Kompetenzaspekte fokussiert. Zum einen wird auf kognitiver Ebene der Erwerb von FDW zum Einsatz digitaler Medien über die Lehrveranstaltungen untersucht. Zum anderen wird die Entwicklung der motivationalen Orientierungen und Einstellungen zum Einsatz digitaler Medien der Lehramtsstudierenden im Verlauf der Lehrveranstaltungen erhoben. Weiterhin werden selbstberichtete Vorerfahrungen zu digitalen Medien vor Beginn sowie Einschätzungen der Teilnehmenden zur Seminarqualität am Ende der Lehrveranstaltungen abgefragt. Anknüpfende retrospektive Interviews mit den teilnehmenden Studierenden dienen der vertieften Untersuchung zur wahrgenommenen Seminarqualität sowie zu Ursachen für Veränderungen in den gemessenen Kompetenzaspekten im Zusammenhang mit Erfahrungen im Seminar.

In der vorliegenden Arbeit als Teil des Verbundprojekts DiKoLeP liegt der Fokus auf dem FDW zum Einsatz digitaler Medien. Dabei wird eine valide Messung dieses Wissens mithilfe eines Leistungstests angestrebt und die Entwicklung dieses gemessenen Wissens über die beteiligten Seminare im Verbund sowie der Zusammenhang von Wissensveränderungen mit Seminarinhalten und -aspekten untersucht. Im nächsten Abschnitt erfolgt die Beschreibung der untersuchten Lehrveranstaltungen im Verbundprojekt, welche den Studienkontext dieser Arbeit darstellen.

5.2 Fachdidaktische Lehrveranstaltungen an den beteiligten Standorten

Dieser Abschnitt beinhaltet die Vorstellung der im Verbundprojekt DiKoLeP entwickelten und untersuchten fachdidaktischen Lehrveranstaltungen zum Einsatz digitaler Medien im Physikunterricht. Dazu wird in 5.2.1 zunächst das Lehrkonzept zu den Veranstaltungen beschrieben, welches standortübergreifend abgestimmte Kerninhalte berücksichtigt. In den praktischen Anteilen unterscheiden sich die Lehrveranstaltungen der verschiedenen Standorte teilweise. Daher wird in 5.2.2 das Seminarkonzept an der RWTH Aachen vorgestellt, welches als Lehr-Lern-Seminar eine praktische Erprobung mit realen Schulklassen vorsieht. Die Seminarkonzepte der Standorte Graz, Paderborn und Tübingen werden in 5.2.3 beschrieben, in welchen die praktische Erprobung mit anderen Seminarteilnehmenden erfolgt. Das Seminar in Innsbruck ist in der untersuchten Stichprobe der vorliegenden Arbeit nicht vertreten, weshalb es an dieser Stelle weniger ausführlich erläutert wird. Im Konzept ähnelt es jedoch dem Lehr-Lern-Seminar in Aachen, da die Studierenden auch eine Erprobung mit Schülerinnen und Schülern durchführen können.

5.2.1 Übergreifendes Lehrkonzept

Das übergreifende Lehrkonzept zur Förderung fachdidaktischer digitaler Kompetenzen im Verbundprojekt DiKoLeP orientiert sich am Ansatz des SQD-Modells (Tondeur et al., 2012; 4.2.2). Das SQD-Modell beschreibt einen Ansatz, um angehende Lehrkräfte auf den Einsatz digitaler Medien im Unterricht vorzubereiten, und wurde aus empirischen Ergebnissen qualitativer

Untersuchungen in diesem Bereich gewonnen (4.2.2). Im Modell werden sieben Schlüsselaspekte auf Ebene der Lehrkonzeptgestaltung festgehalten (Tondeur et al., 2012, S. 138-140):

1. Verzahnung von Theorie und Praxis
2. Einnahme einer Vorbildfunktion von Seminarleiter:innen
3. Ermöglichung einer systematischen Reflexion über die Rolle digitaler Medien
4. Eigenes Entwerfen von Lernarrangements mit digitalen Medien
5. Arbeiten in Teams
6. Unterstützung authentischer Erfahrungen mit digitalen Medien
7. Kontinuierliches Feedback an Studierende

Diese Schlüsselaspekte wurden bei der Gestaltung des Lehrkonzepts sowie der einzelnen Seminarkonzepte im Verbundprojekt berücksichtigt, beispielsweise indem das Lehrkonzept eine Theorie-Praxis-Struktur aufweist. Im ersten Teil des Lehrkonzepts werden zunächst theoriebasiert fachdidaktische Grundlagen zum Einsatz digitaler Medien eingeführt, wobei es auch Gelegenheiten gibt, von der Seminarleitung vorgeführte Medieneinsätze selbst praktisch auszuprobieren und verschiedene digitale Medien kennenzulernen (Schlüsselaspekt 2 und 6). Auf diese Weise erfolgt eine erste Verzahnung von Theorie und Praxis (Schlüsselaspekt 1). Vertieft wird diese Verzahnung im zweiten Teil der Lehrkonzepts, in welchem die Studierenden die gelernten Inhalte praktisch anwenden und eigene Unterrichtssequenzen unter Einbindung digitaler Medien planen und durchführen (Schlüsselaspekte 4 und 6). In diesem praktischen Teil arbeiten die Studierenden in Gruppen zusammen (Schlüsselaspekt 5) und erhalten sowohl von der Seminarleitung als auch voneinander prozessbegleitend Feedback zu ihren erstellten Unterrichtssequenzen und -materialien (Schlüsselaspekt 7). Die Erprobung der Unterrichtssequenzen sowie vorgestellte Medieneinsätze werden im Seminar reflektiert (Schlüsselaspekt 3).

Die Orientierung an diesen Schlüsselaspekten des SQD-Modells bietet eine Grundlage für die strukturelle und methodische Gestaltung des Lehrkonzepts und der Seminare. Das Modell ist jedoch nicht spezifisch für das Fach Physik und konkretisiert keine fachspezifischen Inhalte oder Anforderungen

für die Gestaltung des Lehrkonzepts. An dieser Stelle ist erneut eine fachspezifische Sichtweise mit Orientierung an fachspezifischen und -typischen digitalen Medien und Arbeitsweisen (2.3.2) notwendig, indem zum einen theoretische Grundlagen zu diesen physiktypischen Medien thematisiert sowie die Implementation dieser Medien in den praktischen Seminarteilen ermöglicht werden.

Das Lehrkonzept ist in seiner Theorie-Praxis-Struktur in Abbildung 5-1 dargestellt. Im theoretischen Teil werden standortübergreifend abgestimmte Kerninhalte berücksichtigt, welche relevante Aspekte des FDW zum Einsatz digitaler Medien aufgreifen (2.3). Zu diesen Kerninhalten zählen die CTML sowie die CLT, empirische Befunde zum Medieneinsatz im naturwissenschaftlichen Unterricht (Hillmayr et al., 2020) sowie das SAMR-Modell (2.3.3). Weiterhin werden verschiedene für den Physikunterricht typische digitale Medien behandelt: Simulationen und Animationen, interaktive Bildschirmexperimente, Augmented Reality, digitale Messwerterfassung und mobile Messsensorik, Videoanalyse sowie Erklärvideos (2.3.2). Zu diesen Medien werden jeweils Gestaltungsmerkmale, didaktische Einsatzmöglichkeiten und empirische Befunde besprochen sowie Gelegenheiten zum Kennenlernen und Anwenden der Medien im Seminar ermöglicht.



Abbildung 5-1: Struktur des Lehrkonzepts im Verbundprojekt DiKoLeP aus gemeinsamen Kerninhalten (Theorie) und standortspezifischen praktischen Anteilen (Praxis)

Im praktischen Teil des Lehrkonzepts ist an allen Standorten die eigene Entwicklung von Unterrichtssequenzen unter Einbindung digitaler Medien sowie deren Erprobung und anschließende Reflexion vorgesehen. Aufgrund

verschiedener Rahmenbedingungen an den Standorten wird diese praktische Auseinandersetzung unterschiedlich ausgestaltet, wie es die folgenden Abschnitte darstellen.

5.2.2 Lehr-Lern-Seminar am Standort Aachen

Am Standort Aachen wurde das vorgestellte Lehrkonzept im Rahmen eines Lehr-Lern-Seminars implementiert, welches einen Teil des Studienjahrs zum Praxissemester darstellt. Im Studienjahr zum Praxissemester besuchen die Studierenden im Vorbereitungssemester fachdidaktische und bildungswissenschaftliche Lehrveranstaltungen zur Vorbereitung auf das im folgenden Semester anknüpfende ca. fünfmonatige Praktikum (Praxissemester) am Lernort Schule. Das Lehr-Lern-Seminar ist ein physikdidaktisches Seminar für Masterstudierende (nach Studienplan im ersten Mastersemester bzw. siebten Hochschulsesemester). Es stellt eines von zwei Vorbereitungsseminaren zum Praxissemester im Lehramt Physik dar.

Das Seminar war von 2015 bis 2023 Teil der Lehr-Lern-Gelegenheiten im Projekt *LeBiAC* („Gemeinsam verschieden sein – Lehrerbildung an der RWTH Aachen“) der Qualitätsoffensive Lehrerbildung (BMBF, 2016; 2023). In diesem LeBiAC-Teilprojekt wurden Lehr-Lern-Gelegenheiten mit Schülerinnen und Schülern zur Stärkung des Praxisbezugs sowie der Professionalisierung der Lehramtsstudierenden als integraler Bestandteil in die Lehrkräftebildung in Aachen implementiert. In der zweiten Förderphase *LeBiAC II* wurde unter dem Leitsatz „Gemeinsam verschieden sein in einer digitalen Welt“ weiterhin der Einsatz digitaler Medien in den Lehr-Lern-Gelegenheiten fokussiert⁸.

Das Lehr-Lern-Seminar im Fach Physik sieht eine enge Verknüpfung von Theorie- und Praxisphasen vor. Auf Basis von Theorieimpulsen zu fachdidaktischen Themen entwickeln die Studierenden Unterrichtseinheiten in Form von experimentellen Lernzirkeln, setzen diese in realen Schulklassen

⁸ Weitere Informationen zum Projekt LeBiAC (erste und zweite Phase) finden sich unter www.qualitaetsoffensive-lehrerbildung.de/lebiac oder www.lebiac.rwth-aachen.de (letzter Zugriff am 27.05.2024).

ein und reflektieren die Umsetzung der Unterrichtseinheit sowie ihr Auftreten als Lehrkraft auf Basis von begleitendem Feedback aus verschiedenen Perspektiven (Schülerinnen und Schüler, betreuende Lehrkraft, Seminarleitung und Mitstudierende). Diese Rahmenbedingungen zum Lehr-Lern-Seminar eignen sich durch die vorgesehene Entwicklung und Erprobung einer Unterrichtseinheit in Schulklassen insbesondere, um die Schlüsselaspekte 4 und 6 des SQD-Modells (eigenes Entwerfen von Lernarrangements und authentische Erfahrungen im Einsatz digitaler Medien; 5.2.1) zu adressieren. Damit stellt das Lehr-Lern-Seminar eine passende Grundlage zur Implementation des übergreifenden Lehrkonzepts aus dem Verbundprojekt DiKoLeP dar.

So sind die Theorieimpulse im ersten Seminarteil entsprechend den abgestimmten Kerninhalten des übergreifenden Lehrkonzepts (5.2.1) adaptiert und weiterentwickelt worden. Zusätzlich bilden Schülervorstellungen im Physikunterricht einen fachdidaktischen Inhaltsfokus des ersten Seminarteils im Lehr-Lern-Seminar (Schwerpunkt des Seminars in der ersten Förderphase). Im zweiten Seminarteil erfolgt die Entwicklung und Erprobung der Unterrichtseinheiten in Gruppen, wobei hier der Einsatz digitaler Medien fokussiert wird. Die Studierenden entwickeln dazu adressatengerechte Experimentierstationen sowie begleitende Arbeitsblätter für ihre Unterrichtseinheit als Lernzirkel. Die Lernzirkel sollen dabei den Einsatz digitaler Medien beinhalten, wobei nicht an jeder Station ein digitales Medium verwendet werden muss. Zusätzlich wählen die Gruppen ein inhaltlich passendes und fachdidaktisch geeignetes Erklärvideo für ihre Unterrichtseinheit aus. Die Entwicklung der Stationen und Arbeitsblätter wird durch Feedback der Seminarleitung sowie der Mitstudierenden begleitet. Als physikalischer Inhalt werden in den Lernzirkeln Energieformen und Energieumwandlungen fokussiert.

Die entwickelten Lernzirkel werden von den Studierenden in realen Schulklassen in einer Unterrichtsstunde (Doppelstunde) eigenständig eingesetzt und erprobt. Im Untersuchungszeitraum wurden dazu Schulklassen der achten Jahrgangsstufe besucht, welche gerade Formen mechanischer Energie im Unterricht behandelt hatten. Die Studierenden erhalten in der Erprobung

Feedback von den Lernenden, der betreuenden Lehrkraft sowie der Seminarleitung zur Umsetzung ihrer Unterrichtseinheit sowie ihrem Auftreten als Lehrkraft. Schließlich wird die Erprobung am Lernort Schule unter Berücksichtigung des erhaltenen Feedbacks im Seminar reflektiert.

Da die Studierenden nicht an jeder Station digitale Medien verwenden müssen, ist es möglich, dass einzelne Studierende an der von ihnen hauptverantwortlich entwickelten und in der Erprobung betreuten Station kein digitales Medium einsetzen. Nichtsdestotrotz erfolgt die Gestaltung des Gesamtkonzepts der Unterrichtseinheit inklusive Einbettung des Erklärvideos in Gruppen, sodass alle Studierenden sich mit der Implementation digitaler Medien in der gesamten Unterrichtseinheit auseinandersetzen, auch wenn an der eigenen Station kein digitales Medium zum Einsatz kommt. Die Lerngelegenheit, digitale Medien in einer Unterrichtssequenz zu implementieren, ist jedoch möglicherweise intensiver, wenn an der betreuten Station auch ein digitaler Medieneinsatz erfolgt. Diese fehlende Verpflichtung zum Einsatz eines digitalen Mediums an jeder Station ist eine bewusste Entscheidung für die Lernzirkelentwicklung. Den Studierenden wird das Unterrichtsthema vorgegeben, um für die Erprobung an der Schule eine Passung zum vorausgehenden Unterricht zu gewährleisten. Ausgehend von diesem vorgegebenen Unterrichtsthema entscheiden die Studierenden eigenständig, welche konkreten Inhalte sie auf welche Weise in ihren Lernzirkelstationen adressieren. Dabei steht es ihnen auch frei, an einzelnen Stationen kein digitales Medium zur Vermittlung der Inhalte zu verwenden – genauso wie es im späteren Berufsalltag als Lehrkraft der Fall wäre.

Im folgenden Abschnitt wird die Ausgestaltung des Lehrkonzepts an den Standorten Graz, Paderborn und Tübingen beschrieben, welche sich im praktischen Teil aufgrund anderer Gegebenheiten vom Lehr-Lern-Seminar in Aachen unterscheidet.

5.2.3 Seminare an den Standorten Graz, Paderborn und Tübingen

An den Standorten Graz, Paderborn und Tübingen wurde das Lehrkonzept in fachdidaktischen Seminaren im Lehramtsstudiengang Physik implementiert. Nach empfohlenem Studienplan ist das jeweilige Seminar in Graz im achten Semester (Bachelor), in Paderborn im sechsten Semester (Bachelor) und in

Tübingen im vierten Mastersemester verortet. Die einzelnen Seminarkonzepte weisen ausgehend vom übergreifenden Lehrkonzept eine Theorie-Praxis-Struktur auf, wobei im ersten Seminarteil die abgestimmten Kerninhalte thematisiert werden (5.2.1). Im zweiten Seminarteil erfolgen praktische Auseinandersetzungen, bei denen die Studierenden eine Unterrichtsskizze in Partner- oder Gruppenarbeit entwickeln und in Form von *Micro-Teachings* (z. B. Klinzing, 2002) mit ihren Mitstudierenden im Seminar erproben.

In Graz und Tübingen wählen die Studierendengruppen zu Beginn der praktischen Auseinandersetzung ein digitales Medium aus (je Gruppe verschiedenes Medium), wiederholen und vertiefen die gelernten Inhalte dazu in Eigenarbeit mithilfe vorbereiteter Materialien der Seminarleitung. Anschließend gestalten die Gruppen eine Unterrichtsskizze zu einem von ihnen gewählten fachlichen Inhalt unter Einbettung ihres gewählten Mediums sowie begleitende Arbeitsmaterialien (z. B. Arbeitsblatt, Tafelbild). Von der Seminarleitung erhalten sie dabei begleitendes Feedback zur entwickelten Unterrichtsskizze, um diese ggf. überarbeiten zu können. Am Ende des Seminars erproben die Studierenden eine Sequenz aus ihrer erstellten Unterrichtsskizze im Seminar (Micro-Teaching), wobei die Mitstudierenden die Rolle der Lernenden einnehmen. Die Studierendengruppen erhalten jeweils Feedback von der Seminarleitung und ihren Mitstudierenden, um ihre Erprobung reflektieren zu können. Die beschriebene praktische Auseinandersetzung wird in einem erneuten Durchlauf mit der Auswahl eines zweiten digitalen Mediums wiederholt, sodass alle Studierenden zu zwei digitalen Medien je eine Unterrichtsskizze erstellen und eine Sequenz im Seminar erproben.

In Paderborn erfolgt aufgrund zeitlicher Rahmenbedingen und der geringeren fachdidaktischen Vorkenntnisse der Bachelorstudierenden nur ein Durchlauf der praktischen Auseinandersetzung. Ebenso entwickeln die vergleichsweise weniger erfahrenen Studierenden hier nur eine Unterrichtssequenz in Form einer Erarbeitungsphase und keine vollständige Unterrichtsstunde. Die Studierendengruppen wählen im Seminar aus vorgegebenen Themen und Jahrgangsstufen den fachlichen Inhalt für ihre Unterrichtssequenz aus. Ausgehend von diesem gewählten Unterrichtsthema planen die Gruppen anschließend eine experimentelle Erarbeitungsphase unter Einbindung digitaler Medien und erstellen ein zugehöriges Arbeitsblatt. Die Themen

werden im Vorhinein von der Seminarleitung danach ausgewählt, dass sich verschiedene Einsätze digitaler Medien anbieten (z. B. Bewegungen mit Videoanalyse, mechanische Schwingungen mit Smartphones, ionisierende Strahlung mit interaktiven Bildschirmexperimenten). Die Entwicklung der Erarbeitungsphasen und Arbeitsblätter wird auch hier von der Seminarleitung durch mehrmaliges Feedback begleitet. Die entwickelten Unterrichtssequenzen werden im Seminar als Micro-Teaching erprobt, wobei auch hier die Mitstudierenden die Rolle der Schülerinnen und Schüler einnehmen. Diese Erprobung reflektieren die Teilnehmenden anschließend mithilfe des Feedbacks der Seminarleitung und Mitstudierenden.

Durch die Entwicklung und Erprobung einer eigenen Unterrichtseinheit unter Einbettung digitaler Medien werden auch hier die Schlüsselaspekte 4 und 6 des SQD-Modells (5.2.1) adressiert, wenngleich die Erprobung vor Mitstudierenden nicht dieselbe Authentizität aufweist wie vor einer realen Schulklasse. Eine Erprobung mit Schülerinnen und Schülern wie im Lehr-Lern-Seminar in Aachen ist an den Standorten Graz, Paderborn und Tübingen jedoch nicht umsetzbar, sodass die Micro-Teachings innerhalb des Seminars eine bestmögliche Alternative darstellen.

Nachdem nun mit dem Verbundprojekt DiKoLeP sowie den dort implementierten Lerngelegenheiten an den Standorten der Untersuchungskontext dieser Arbeit beschrieben wurde, werden im folgenden Kapitel die Ziele und Forschungsfragen der Arbeit erläutert.

6 Ziele der Arbeit

In diesem Kapitel werden die Konsequenzen aus dem in Kapitel 2, 3 und 4 beschriebenen Stand der relevanten Literatur und Forschung zusammengefasst (6.1) und die davon ausgehenden Ziele und Forschungsfragen für diese Arbeit formuliert (6.2).

6.1 Konsequenzen aus Literatur und bisheriger Forschung

Die vorliegende Arbeit beschäftigt sich mit der Messung und Entwicklung des FDW zum Einsatz digitaler Medien bei Physiklehramtsstudierenden. Dieses zentrale Konstrukt der Arbeit zählt zum Professionswissensbereich FDW, welcher wiederum einen wichtigen Teil der professionellen Handlungskompetenz von (angehenden) Lehrkräften darstellt (2.1) und für die Gestaltung und Qualität von Unterricht bedeutsam ist (z. B. Baumert et al., 2010; 4.1). Die Förderung des FDW ist daher ein wichtiger Bestandteil in der Professionalisierung angehender Lehrkräfte (Baumert & Kunter, 2006; Kaiser et al., 2020). Zudem erfordert die zunehmende Bedeutung digitaler Medien für Schule und Unterricht (Kapitel 1) auch eine entsprechende Förderung digitalisierungsbezogener Kompetenzen, um Lehrkräfte zu einem fachdidaktisch sinnvollen Einsatz digitaler Medien im Unterricht zu befähigen (KMK, 2016). Die Förderung solcher Kompetenzen in der Lehrkräftebildung ist dabei nicht nur Aufgabe der Bildungswissenschaften oder einer übergeordneten Mediendidaktik, sondern ebenso der einzelnen Fächer und Fachdidaktiken (KMK, 2016; Schmid et al., 2020; SWK, 2022). Dies begründet die Relevanz dafür, sich in der vorliegenden Arbeit dem FDW zum Einsatz digitaler Medien im Fach Physik zu widmen.

Der fokussierte Wissensbereich FDW zum Einsatz digitaler Medien ist für das Fach Physik bislang nicht ausführlich beschrieben worden, da bisherige Modelle zum FDW in Physik andere fachdidaktische Facetten fokussieren (2.2). Beschreibungen zu digitalisierungsbezogenen Kompetenzen oder Wissensaspekten wie das TPACK-Modell (Koehler et al., 2013; Mishra & Koehler, 2006) sind oft fachübergreifend und nicht spezifisch formuliert (2.3.1). Der zentrale Wissensbereich TPACK des gleichnamigen Modells ist zwar dem hier fokussierten FDW zum Einsatz digitaler Medien ähnlich, wird jedoch

auch nicht fachspezifisch konkretisiert. Dabei ist es aus fachdidaktischer Perspektive für den Medieneinsatz im Unterricht bedeutsam, sich an fachspezifischer und -typischer Mediennutzung sowie an Arbeitsweisen des Fachs zu orientieren (Ropohl et al., 2018). Für das FDW zum Einsatz digitaler Medien im Fach Physik spielt demnach vor allem das Wissen über Möglichkeiten und Anforderungen bei der Nutzung physikspezifischer und -typischer Medien eine wichtige Rolle. Zu physikspezifischen Medien gehören unter anderem digitale Messwerterfassung, Simulationen oder auch Erklärvideos (2.4). Die bislang noch fehlende konkrete fachspezifische Beschreibung des FDW zum Einsatz digitaler Medien bzw. des TPACK im Fach Physik stellt somit ein erstes für diese Arbeit bedeutsames Desiderat dar.

Desiderat 1: Es existiert keine konkrete fachspezifische Beschreibung des FDW zum Einsatz digitaler Medien (bzw. TPACK) im Fach Physik.

Ebenso existieren für das FDW zum Einsatz digitaler Medien im Fach Physik bisher keine fachspezifischen Messungen. Etablierte Leistungstests zum FDW in Physik aus dem deutschsprachigen Raum fokussieren nicht den Einsatz digitaler Medien und Testinstrumente zum ähnlichen Wissensbereich TPACK liegen für andere naturwissenschaftliche Fächer, nicht aber für die Physik vor (3.6). Zudem wird TPACK meist über Selbsteinschätzungen und auf generische Weise erfasst (Wang et al., 2018; Willermark, 2018), wobei solche Messverfahren zunehmend hinsichtlich fehlender Validität und Fachspezifität kritisiert werden (z. B. Akyuz, 2018; Kotzebue, 2022a; 3.5). Ein valides Messverfahren zur Erfassung des TPACK oder FDW zum Einsatz digitaler Medien ist demnach unabdingbar, um im Sinne eines argumentbasierten Validitätsverständnisses (3.3.2) angemessene Schlussfolgerungen aus erfolgten Messungen ziehen zu können. Dies gelingt jedoch nicht mit den bisher vielfach genutzten Selbsteinschätzungsinstrumenten in der TPACK-Forschung; es braucht dazu eine proximale Messung, z. B. über einen Leistungstest. Zur Entwicklung eines solchen Leistungstests sind Aspekte der Testökonomie zu beachten. Bisherige leistungsbasierte Messverfahren zum TPACK (in anderen Fächern) gehen mit hohem Aufwand für eine Sicherstellung der Objektivität einher, wenn z. B. die Bewertung von offenen Leistungstestaufgaben oder Produkten der Unterrichtsplanung notwendig ist (3.5.2).

Geschlossene Aufgabenformate sind in der Bewertung grundsätzlich eindeutig und damit weniger aufwändig; gleichzeitig sind diese Aufgaben jedoch schwieriger zu erstellen, da passende richtige und vor allem falsche Antwortalternativen entwickelt werden müssen (3.2.2). Zur Berücksichtigung der Fachspezifität in einem solchen Leistungstest ist zudem eine inhaltliche Orientierung an fachtypischen Medien und Arbeitsweisen (siehe oben) sinnvoll (3.6). Das zweite Desiderat bezieht sich demnach auf die fehlenden Messverfahren zum FDW zum Einsatz digitaler Medien:

Desiderat 2: Es existieren keine validen Messverfahren zur fachspezifischen Erfassung des FDW zum Einsatz digitaler Medien (bzw. TPACK) im Fach Physik.

Die Entwicklung eines validen Testinstruments ermöglicht schließlich eine Untersuchung des FDW zum Einsatz digitaler Medien in Physik bei angehenden Lehrkräften. Der beschriebene Forschungsstand zeigt hier, dass nationale Untersuchungen zum Professionswissen und FDW bei angehenden Lehrkräften in Physik sich nicht auf den Einsatz digitaler Medien beziehen (4.1). Im fachdidaktischen Forschungsfeld mit Bezug zur Nutzung digitaler Medien adressieren physikspezifische Studien bisher z. B. die Einstellungen zur Vermittlung digitaler Kompetenzen (Walpert & Wodzinski, 2023) oder den eigentlichen Einsatz digitaler Medien durch Physiklehramtsstudierende (Schaber & Friege, 2023). Die Entwicklung des medienbezogenen FDW oder TPACK bei Physiklehramtsstudierenden scheint jedoch bisher nicht untersucht (4.3). Ausstehend sind daher Untersuchungen zum FDW zum Einsatz digitaler Medien von Physiklehramtsstudierenden und dazu, wie sich dieses Wissen über bestimmte Lerngelegenheiten entwickelt.

Desiderat 3: Es existieren keine Untersuchungen zur Evaluation physikdidaktischer Lerngelegenheiten im Lehramtsstudium hinsichtlich der Entwicklung des FDW zum Einsatz digitaler Medien im Physikunterricht (bzw. TPACK).

In der Forschung zur Entwicklung des FDW von Lehramtsstudierenden wurde in der Vergangenheit zudem auch untersucht, welche Aspekte von Lerngelegenheiten die Entwicklung des FDW beeinflussen. Dabei hängen glo-

bale Faktoren wie das Fachsemester oder die absolvierten SWS in der Fachdidaktik positiv mit der Ausprägung des FDW zusammen (z. B. Riese et al., 2017; 4.1.2). Zudem scheint die Entwicklung des FDW auch abhängig vom inhaltlichen Fokus der Lerngelegenheit. So konnte Joswig-Käfer (2024) – passend zur inhaltlichen Ausrichtung der untersuchten Lehrveranstaltungen – einen Zuwachs im FDW in den Inhaltsbereichen Optik und Elektrizitätslehre, jedoch nicht im Bereich Mechanik feststellen (4.1.1). Über ihre vertiefende qualitative Interviewstudie ermittelte Joswig-Käfer (2024) zudem, inwiefern sich die gemessene Wissenszuwächse auf Aspekte der untersuchten Lerngelegenheiten oder auf andere Faktoren wie z. B. der Nutzung von Fachwissen, dem Zeitmanagement oder der Motivation bei der Testbearbeitung zurückführen lassen (4.1.2). Aus bisherigen Evaluationen zum TPACK in anderen Fächern deutet sich an, dass ein deutlicher Praxisbezug (z. B. Max et al., 2022; Tondeur et al., 2012; Zimmermann, 2021) und das Anknüpfen an fachdidaktisches Grundlagenwissen (Backfisch et al., 2021; Zimmermann, 2021) sowie das Sammeln authentischer Erfahrungen mit digitalen Medien und die Reflexion über deren Rolle (Tondeur et al., 2012) nützlich zur Förderung von TPACK sind (4.2.2). Für das FDW zum Einsatz digitaler Medien von Physiklehramtsstudierenden ist jedoch offen, wie sich dieses Wissen in entsprechenden Lerngelegenheiten entwickelt und welche Elemente dieser Lerngelegenheiten lernförderlich zur Entwicklung des FDW zum Einsatz digitaler Medien sind (4.3).

Desiderat 4: Es fehlen Erkenntnisse über Elemente von fachdidaktischen Lerngelegenheiten, die förderlich für den Erwerb von FDW zum Einsatz digitaler Medien (bzw. TPACK) sind.

Zusammenfassend zeigen sich aus dem beschriebenen Forschungsstand folgende Aspekte als Desiderate, die (nach Kenntnisstand der Autorin) bislang nicht vorliegen:

1. konkrete fachspezifische Beschreibung des FDW zum Einsatz digitaler Medien (bzw. TPACK) im Fach Physik
2. valide Messverfahren zur fachspezifischen Erfassung des FDW zum Einsatz digitaler Medien im Physikunterricht (bzw. TPACK)

3. Untersuchungen zur Evaluation physikdidaktischer Lerngelegenheiten im Lehramtsstudium hinsichtlich der Entwicklung des FDW zum Einsatz digitaler Medien im Physikunterricht (bzw. TPACK)
4. Erkenntnisse über Elemente von fachdidaktischen Lerngelegenheiten, die förderlich für den Erwerb von FDW zum Einsatz digitaler Medien (bzw. TPACK) sind.

Ausgehend von den dargestellten Desideraten und Konsequenzen werden im folgenden Abschnitt die Ziele und Forschungsfragen für diese Arbeit formuliert und erläutert.

6.2 Ziele und Forschungsfragen

Wie im vorausgehenden Abschnitt dargestellt, steht die Untersuchung des FDW zum Einsatz digitaler Medien bei Studierenden im Lehramt Physik im Fokus dieser Arbeit. Den Kontext der Arbeit stellen dabei die Seminare im Verbundprojekt DiKoLeP dar, welches sich der Entwicklung und Evaluation physikdidaktischer Lehrveranstaltungen zur Förderung fachdidaktischer digitaler Kompetenzen wie dem FDW zum Einsatz digitaler Medien widmet (Kapitel 5). Um zur Evaluation dieser Lehrveranstaltungen beizutragen sowie die in 6.1 beschriebenen Desiderate zu adressieren, werden in der vorliegenden Arbeit mehrere Ziele in Bezug auf das FDW zum Einsatz digitaler Medien in Physik verfolgt.

Dabei ist im Hinblick auf das erste Desiderat (6.1) zunächst eine fachspezifische Beschreibung des FDW zum Einsatz digitaler Medien in Physik vorgesehen. Diese Modellierung des adressierten Wissensbereichs stellt damit das **erste Ziel** der vorliegenden Arbeit dar und dient als Grundlage für die anschließende Testentwicklung.

Anknüpfend wird die valide Messung des fokussierten Wissensbereichs in den Blick genommen, um das zweite Desiderat zu adressieren, dass bislang kein derartiges Instrument für die Physik vorliegt (3.6; 6.1). Das **zweite Ziel** der Arbeit stellt somit die Entwicklung und Validierung eines Leistungstests zur Erfassung des FDW zum Einsatz digitaler Medien im Fach Physik in den Mittelpunkt. Anknüpfend an die Modellierung im Rahmen des ersten Ziels

zählt zu diesem zweiten Ziel die Testentwicklung sowie anschließende empirische Untersuchungen zur Validität. Für die zu entwickelnden Testaufgaben wird ein geschlossenes Antwortformat im Sinne von Mehrfachwahlaufgaben angestrebt, um eine ökonomische Bearbeitung und Auswertung bei gleichzeitiger Objektivität zu ermöglichen. Mit diesem Format des Leistungstests aus Mehrfachwahlaufgaben wird ein Kompromiss geschaffen: So erfolgt die Messung des FDW zum Einsatz digitaler Medien über einen proximalen Leistungstest, um Validitätsprobleme von in der TPACK-Messung häufig verwendeten Selbsteinschätzungen zu überwinden. Gleichzeitig wird ein geschlossenes Aufgabenformat anstelle der bei Leistungstests üblicherweise freien bzw. offenen Aufgabenformaten genutzt, da diese mit viel Aufwand in der Bearbeitung und insbesondere der Gewährleistung einer objektiven Bewertung einhergehen (6.1). In einer Pilotierung werden die entwickelten Aufgaben vor dem eigentlichen Testeinsatz erprobt und anschließend überarbeitet. Die Untersuchung der Validität erfolgt schließlich im Hinblick auf mehrere Aspekte der Konstruktvalidität (3.3.2), um Argumente zu gewinnen, inwieweit mit dem entwickelten Testinstrument eine angemessene Interpretation hinsichtlich der Messung des FDW zum Einsatz digitaler Medien bei Physiklehramtsstudierenden erfolgen kann. Zu diesem Ziel wird die folgende erste Forschungsfrage untersucht:

FF1: Inwiefern lässt sich das FDW zum Einsatz digitaler Medien bei Physiklehramtsstudierenden mithilfe des entwickelten Testinstruments valide messen?

Das **dritte Ziel** der Arbeit greift das dritte Desiderat auf (6.1) auf und adressiert die Evaluation der im Verbundprojekt DiKoLeP entwickelten und implementierten physikdidaktischen Seminare zum Einsatz digitaler Medien im Physikunterricht (Kapitel 5). In diesem Zusammenhang wird die Entwicklung des FDW zum Einsatz digitaler Medien bei den teilnehmenden Studierenden über den Verlauf der Seminare ermittelt. Wie in 5.2 bereits ausführlicher beschrieben, verfolgen die untersuchten Seminare dabei ein gemeinsames Lehrkonzept, das Theorie- und Praxisteile verknüpft und insbesondere im ersten (eher theoretischen) Seminarteil gemeinsam abgestimmte Kerninhalte zu fachdidaktischen Grundlagen zum Einsatz digitaler Medien und zu physiktypischen Medien berücksichtigt (5.2.1). Im zweiten Seminarteil wenden die

Studierenden die gelernten Inhalte praktisch an und entwickeln eigene Unterrichtseinheiten, die sie mit Schulklassen oder im Seminar erproben (5.2.2; 5.2.3). Das Lehrkonzept berücksichtigt damit Schlüsselaspekte des SQD-Modells, welche als förderlich für die Vorbereitung angehender Lehrkräfte auf den Einsatz digitaler Medien gelten (Tondeur et al., 2012; 5.2.1). Hinsichtlich dieses Ziels zur Evaluation der Seminare ergibt sich somit die folgende zweite Forschungsfrage:

FF2: Inwieweit verändert sich das gemessene FDW zum Einsatz digitaler Medien über die physikdidaktischen Seminare der kooperierenden Standorte im Verbundprojekt DiKoLeP⁹?

Für FF2 wird Hypothese H1 formuliert, dass das FDW zum Einsatz digitaler Medien im Verlauf der untersuchten Seminare zunimmt:

H1: Das FDW zum Einsatz digitaler Medien der teilnehmenden Studierenden nimmt über die untersuchten Seminare zu.

Diese Wirkungshypothese beruht zunächst auf der inhaltlichen Passung der abgestimmten Kerninhalte in den Seminaren zum untersuchten Konstrukt des FDW zum Einsatz digitaler Medien sowie der Orientierung an den Schlüsselaspekten des SQD-Modells (5.2.1). Weiterhin deutet der beschriebene Forschungsstand an, dass jede SWS in der Fachdidaktik (d. h. jede fachdidaktische Lerngelegenheit) zum Ausbildungserfolg im Hinblick auf das FDW beiträgt (Riese et al., 2017; 4.1.2) und sich das FDW abhängig vom inhaltlichen Fokus der Lerngelegenheit entwickelt (Joswig-Käfer, 2024; 4.1.1). Aus diesen Gründen wird davon ausgegangen, dass die entwickelten physikdidaktischen Seminare zu digitalen Medien geeignet sind, um das FDW zum Einsatz digitaler Medien der teilnehmenden Physiklehramtsstudierenden zu steigern.

⁹ Da im Untersuchungszeitraum für die vorliegende Arbeit nur eine Person in Innsbruck an dem Seminar und der Studie teilgenommen hat, wurden diese Daten nicht in die Stichprobe zur Arbeit einbezogen. Das bedeutet, dass im Rahmen von FF2 nur die Seminare an den Standorten Aachen, Graz, Paderborn und Tübingen hinsichtlich der Entwicklung des untersuchten FDW evaluiert wurden.

Anknüpfend an die Evaluation hinsichtlich des FDW zum Einsatz digitaler Medien wird das vierte Desiderat (6.1) im Rahmen des **vierten Ziels** der Arbeit aufgegriffen. Dabei werden Gründe für Veränderungen im gemessenen FDW erkundet und mit den untersuchten Lerngelegenheiten in Verbindung gebracht, um lernförderliche oder lernhinderliche Seminarelemente zu identifizieren. Zu diesem Ziel wird die folgende dritte Forschungsfrage formuliert:

FF3: Welche Elemente des gemeinsamen Kerns bzw. der standortspezifischen Teile des Lehrkonzepts sind besonders lernförderlich oder eher lernhinderlich und stehen mit Verbesserung oder Verschlechterung im gemessenen FDW zum Einsatz digitaler Medien in Verbindung?

FF3 wird explorativ und hypothesengenerierend untersucht, um eine Offenheit für verschiedene Erkenntnisse zu ermöglichen sowie einen Beitrag zum noch lückenhaften Forschungsstand zu Zusammenhängen von Lerngelegenheiten und der Entwicklung von TPACK bzw. FDW zum Einsatz digitaler Medien in Physik leisten zu können (4.2.2).

Schließlich stellt ein **fünftes Ziel** dieser Arbeit die Generierung von Hypothesen zur Gestaltung von Lerngelegenheiten zum Einsatz digitaler Medien dar. Dieses Ziel bildet wiederum ein zentrales Anliegen im Verbundprojekt DiKoLeP (5.1). Insbesondere sollen die Erkenntnisse der qualitativen Untersuchung zu FF3 Beiträge zu dieser Hypothesenbildung liefern und damit als Gestaltungswissen für ähnliche Lerngelegenheiten nutzbar gemacht werden.

Die Ziele der vorliegenden Arbeit sowie die damit einhergehenden Forschungsfragen können wie folgt zusammengefasst werden:

Ziel 1: Modellierung des FDW zum Einsatz digitaler Medien im Physikunterricht.

Ziel 2: Entwicklung und Validierung eines Testinstruments zur Erfassung des FDW zum Einsatz digitaler Medien von Physiklehramtsstudierenden.

FF1: Inwiefern lässt sich das physikdidaktische Wissen zum Einsatz digitaler Medien mithilfe des entwickelten Testinstruments valide messen?

Ziel 3: Untersuchung des Wissenserwerbs im FDW zum Einsatz digitaler Medien über die im Verbundprojekt DiKoLeP beteiligten Seminare.

FF2: Inwieweit verändert sich das gemessene physikdidaktische Wissen zum Einsatz digitaler Medien über die Seminare der kooperierenden Standorte im Verbundprojekt DiKoLeP?

H1: Das FDW zum Einsatz digitaler Medien der teilnehmenden Studierenden nimmt über die untersuchten Seminare zu.

Ziel 4: Erkundung von Gründen für Veränderungen im Testverhalten zum FDW zum Einsatz digitaler Medien und Identifikation lernförderlicher und lernhinderlicher Seminarelemente.

FF3: Welche Elemente des gemeinsamen Kerns bzw. der standortspezifischen Teile des Lehrkonzepts sind besonders lernförderlich oder eher lernhinderlich und stehen mit Verbesserung oder Verschlechterung im gemessenen FDW zum Einsatz digitaler Medien in Verbindung?

Ziel 5: Generierung von Hypothesen für wirksame Lerngelegenheiten zum Einsatz digitaler Medien im Physikunterricht.

Im weiteren Verlauf dieser Arbeit wird das Studiendesign der Arbeit vorgestellt sowie das methodische Vorgehen zur Beantwortung der jeweiligen Forschungsfragen beschrieben.

7 Studiendesign und methodisches Vorgehen

In diesem Kapitel werden das Design und methodische Vorgehen zur Untersuchung der in 6.2 genannten Ziele und Forschungsfragen beschrieben. Die nachfolgende Abbildung 7-1 stellt das übergeordnete Studiendesign der Arbeit dar. Zu Beginn werden die Testentwicklung und -pilotierung sowie verschiedene Validierungsstudien zur Untersuchung von FF1 adressiert. Anknüpfend erfolgt in mehreren Durchgängen die Evaluation der fachdidaktischen Seminare im Verbundprojekt DiKoLeP. Dabei wird die Entwicklung des FDW zum Einsatz digitaler Medien (FF2) in einer Prä-Post-Erhebung untersucht. Zur Erkundung von Gründen für veränderte Testantworten und Identifikation von lernförderlichen oder lernhinderlichen Seminarelementen (FF3) erfolgen nach dem Posttest mit einem Teil der teilnehmenden Studierenden retrospektive Interviews (Abbildung 7-1).

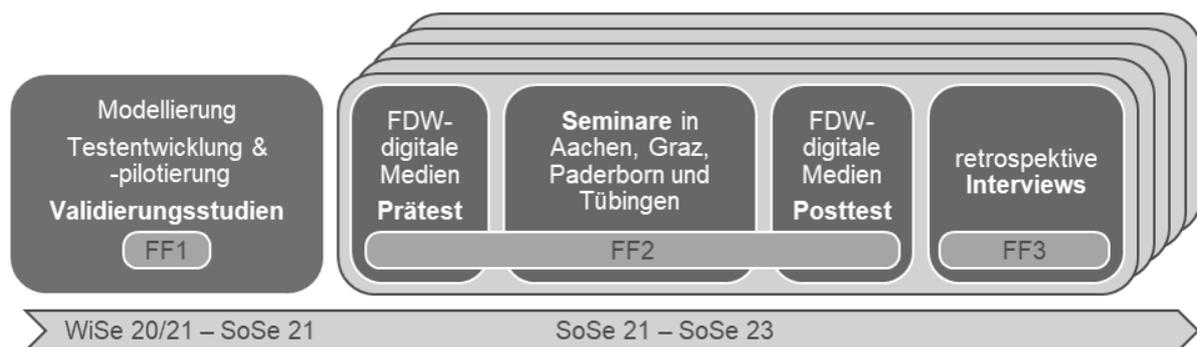


Abbildung 7-1: Studiendesign zur Untersuchung der Forschungsfragen

Zur Untersuchung der verschiedenen Forschungsfragen werden in der vorliegenden Arbeit gemäß eines Mixed-Methods-Ansatzes (3.1) sowohl quantitative als auch qualitative Methoden und Verfahren verwendet. So bilden die Untersuchung der Validität und des fachdidaktischen Wissenserwerbs im Rahmen von FF1 und FF2 den quantitativen Teil der Arbeit. In diesem Teil wird die Untersuchung der Wirksamkeit des Lehrkonzepts adressiert, wozu in der fachdidaktischen Forschung meist quantitative Verfahren genutzt werden (3.1). Im qualitativen Teil erfolgt die Untersuchung von FF3 im Rahmen der retrospektiven Interviewstudie. Durch diese qualitative Studie sollen Gründe für Studierendenantworten im Test explorativ identifiziert und darauf aufbauend Hypothesen für die Gestaltung von Lerngelegenheiten (6.2)

entwickelt werden. Für diese explorative Untersuchung und die Entwicklung von Hypothesen ist die Wahl qualitativer Verfahren naheliegend (3.1).

Im Folgenden werden das Vorgehen und die verwendeten Methoden zu den beiden Studienteilen sowie den einzelnen Forschungsfragen und Zielen ausführlicher erläutert. Dieses ist in Grundzügen auch teilweise in den Beiträgen von Große-Heilmann et al. (2021), Große-Heilmann et al. (2022) und Große-Heilmann et al. (2024) beschrieben.

7.1 Vorgehen zum quantitativen Teil

Im quantitativen Teil dieser Arbeit wird im ersten Schritt die Modellierung des FDW zum Einsatz digitaler Medien im Fach Physik sowie die darauf aufbauende Entwicklung des Testinstruments zur Messung dieses Wissens adressiert. Anknüpfend erfolgt die Untersuchung der Validität des entwickelten Testinstruments im Rahmen von FF1. Schließlich wird mithilfe des Tests die Entwicklung des FDW zum Einsatz digitaler Medien der an den Seminaren teilnehmenden Physiklehramtsstudierenden in mehreren Durchgängen untersucht, um FF2 zu beantworten (Abbildung 7-1).

Modellierung, Testentwicklung und Pilotierung

Zur Entwicklung des Testinstruments wird zunächst der fokussierte Wissensbereich des FDW zum Einsatz digitaler Medien inhaltlich modelliert. Diese Modellierung erfolgt auf Basis der Beschreibung zur Facette (Digitale Medien des FDW-Modells nach (Gramzow, 2015; Gramzow et al., 2013; Abbildung 2-4). Diese Beschreibung wird unter Heranziehen verschiedener Konzeptualisierungen zu digitalisierungsbezogenen Kompetenzen (Becker et al., 2020; Mishra & Koehler, 2006; Redecker, 2017; 2.3.1) sowie fachübergreifender (Mayer, 2009; Sweller, 1994) und physikdidaktischer Literatur (Girwidz, 2020c) zum Medieneinsatz ausdifferenziert. Das Ziel stellt hierbei die Entwicklung einer gemeinsamen Modellierung im Projektverbund DiKoleP dar, welche verschiedene (inhaltliche) Kategorien für eine systematische Aufgabenentwicklung unterscheidet. Die Modellierung des FDW zum Einsatz digitaler Medien wird in 8.1 vorgestellt.

Darauf aufbauend erfolgt die Entwicklung von Testaufgaben, die die Modellierung des FDW zum Einsatz digitaler Medien inhaltlich abdecken sollen. Dazu wird ein dreidimensionales Itementwicklungsmodell erstellt, welches eine Adaption des Itementwicklungsmodells im Projekt ProfiLe-P (Gramzow, 2015; Riese et al., 2017; Abbildung 3-2) im Hinblick auf das hier fokussierte FDW zum Einsatz digitaler Medien darstellt. Das Modell unterscheidet neben inhaltlichen Kategorien zum fokussierten FDW demnach in Anlehnung an das Itementwicklungsmodell nach Gramzow (2015) die drei kognitiven Anforderungen *Reproduzieren*, *Anwenden* und *Analysieren* (Anderson & Krathwohl, 2001) sowie den physikalischen Inhaltsbereich. Die Aufgabenentwicklung erfolgt systematisch unter Variation der inhaltlichen Kategorien des FDW zum Einsatz digitaler Medien und der kognitiven Anforderungen. Für den Test werden als Aufgabenformat Mehrfachwahlaufgaben angestrebt, welche jeweils mehrere richtige und falsche Antwortalternativen erfordern. Die Entwicklung dieser Antwortalternativen erfolgt dabei literaturbasiert sowie in engem Austausch mit den beteiligten Physikdidaktikern¹⁰ im Verbundprojekt DiKoLeP. Das Vorgehen zur Erstellung der Mehrfachwahlaufgaben anhand des Itementwicklungsmodells in dieser Arbeit wird in 8.2 beschrieben.

Die entwickelten Mehrfachwahlaufgaben werden schließlich über die Pilotierung in einer der Zielgruppe ähnlichen Stichprobe erprobt. Dazu erfolgen Befragungen bei Lehramtsstudierenden an sieben Hochschulen in Deutschland und Österreich mithilfe des entwickelten Testinstruments. Die Stichprobe zur Pilotierung wird in 8.3 beschrieben. Durch diese Pilotierung können erste statistische Analysen zur Qualität der Testaufgaben und einzelner Antwortalternativen erfolgen. Zudem werden im Rahmen der Pilotierung zu Validierungszwecken auch weitere Testinstrumente zu angrenzenden Wissensbereichen eingesetzt sowie Studierende anderer Lehrämter befragt. Auf Basis der gewonnenen Erkenntnisse in der Pilotierung wird das Testinstrument überarbeitet, indem einzelne Aufgaben und Antwortalternativen verändert, entfernt oder neu entwickelt werden. Die Pilotierung sowie darauf erfolgte Überarbeitungen zum Testinstrument werden in 8.3 beschrieben. Auf

¹⁰ Da neben der Autorin nur männliche Personen im Verbundprojekt beteiligt waren, muss an dieser Stelle nicht gegendert werden.

Basis der probabilistischen Testtheorie bzw. IRT (3.3.1) als methodische Grundlage erfolgen anknüpfende Untersuchungen zur Testgüte des überarbeiteten Testinstruments (Modellpassung, EAP-Reliabilität, Aufgabenschwierigkeiten; 8.4).

Untersuchung der Validität des Leistungstests zum FDW zum Einsatz digitaler Medien (FF1)

Zur Validierung hinsichtlich unterschiedlicher Aspekte der Konstruktvalidität (3.3.2) werden verschiedene Teilstudien (z. T. im Rahmen der Pilotierung und z. T. mit dem überarbeiteten Testinstrument) durchgeführt. Durch diese Validierungsstudien sollen Hinweise für eine angemessene Interpretation der mit dem Test erfassten Daten gewonnen werden. Nach Kane (2001) werden verschiedene Bereiche für die Interpretationen von Testergebnissen unterschieden: das Bewerten, das Verallgemeinern, das Extrapolieren oder das Erklären von Ergebnissen sowie das Fällen von weiterführenden Entscheidungen als Konsequenz von Ergebnissen (vgl. auch Hartig et al., 2020). So soll der vorliegende Test eine Testwertinterpretation erlauben, welche auf eine Erklärung und eine anknüpfende Bewertung von Leistungen abzielt. Dabei sollen für die Erklärung von Leistungen Annahmen über das theoretische Konstrukt des FDW zum Einsatz digitaler Medien ermöglicht werden, inwiefern dieses Wissen (1) universitär erwerbbar ist, (2) fachspezifisch ist und sich von anderen Wissensbereichen abgrenzt und (3) eine angenommene Strukturierung aufweist. Diese zu untersuchenden Annahmen und Validitätsaspekte sind dabei in dieser Arbeit zum Teil orientiert an den Validierungsstudien zum FDW-Test im Projekt ProfiLe-P (Gramzow, 2015). Das Vorgehen zu den Validierungsstudien dieser Arbeit sowie die Auswahl der untersuchten Validitätsaspekte wird hier kurz aufgezeigt. Kapitel 9 beinhaltet eine ausführlichere Beschreibung diese Studien mitsamt ihren Ergebnissen.

So wird die **Abgrenzung des untersuchten FDW zum Einsatz digitaler Medien zu angrenzenden Konstrukten** wie dem FDW (ohne Bezug zu digitalen Medien) und dem pädagogischen Wissen untersucht. Dafür werden im Rahmen der Pilotierung bei einem Teil der Befragten zusätzlich entsprechende Testinstrumente eingesetzt und anschließend Korrelationsanalysen der jeweiligen Testscores vorgenommen (9.1). Diese Teilstudie zur Abgrenzung

dient der Überprüfung, dass mit dem entwickelten Testinstrument Aussagen zum FDW als *fachspezifischen* Wissensbereich zum Einsatz digitaler Medien getroffen werden können und nicht „nur“ allgemein-mediendidaktisches Wissen erfasst wird. Diese Abgrenzung ist für die vorliegende Arbeit bedeutsam, da der entwickelte physikspezifische FDW-Test zum Einsatz digitaler Medien eine angemessene Fachspezifität aufweisen soll, an welcher es z. B. bisherigen TPACK-Instrumenten häufig mangelt (3.5).

Weiterhin wird die **diskriminante Validität** (3.3.2) im Rahmen der Pilotierung über den Einsatz des entwickelten Testinstruments bei Lehramtsstudierenden anderer Fächer adressiert. Deren Testleistungen werden mit denen der befragten Physiklehramtsstudierenden mithilfe eines t-test unabhängiger Stichproben (z. B. Bortz & Schuster, 2010) verglichen (9.2). Auch hier steht die Überprüfung einer angemessenen Fachspezifität des entwickelten Instruments im Fokus, indem untersucht wird, ob Lehramtsstudierende der Physik erwartungsgemäß im physikspezifischen FDW-Test bessere Leistungen erzielen als die Lehramtsstudierenden anderer Fächer.

Die Verständlichkeit der Aufgaben sowie die **kognitive Validität** (3.3.2) werden über eine Erhebung zum Lauten Denken bei vier Physiklehramtsstudierenden adressiert (9.3). Damit wird untersucht, inwiefern tatsächlich das angestrebte FDW zum Einsatz digitaler Medien von den Studierenden zur Beantwortung der Aufgaben herangezogen wird und nicht etwa Beweggründe durch Raten, sprachliche Hinweise oder andere Wissensquellen oder Wissensarten bei der Beantwortung überwiegen. Mit dieser Befragung zum Lauten Denken wird zur Validierung auch eine qualitative Datenerhebungsmethode verwendet, welche im Sinne einer mehrperspektivischen Betrachtung als Triangulation (3.1) zur Bewertung der Konstruktvalidität betrachtet werden kann.

Die Bewertung des **inhaltlich curricularen Aspekts der Validität** erfolgt über eine Befragung von Expert:innen zur curricularen Passung der Testaufgaben. Dazu werden Physikdidaktiker:innen an zehn Hochschulstandorten in Deutschland und Österreich zu den einzelnen Testaufgaben um eine Einschätzung gebeten, inwiefern das notwendige Wissen dazu in der physikdidaktischen Ausbildung an ihrer Hochschule erworben werden kann (9.4).

Diese Teilstudie dient der Überprüfung, inwiefern das mit dem Testinstrument erfasste FDW auch universitär erwerbbares Wissen darstellt, da die universitäre Lehrkräftebildung den Kontext für die zu evaluierenden Seminare und die Zielgruppe darstellt.

Im Sinne der **strukturellen Validität** (3.3.2) wird die in der Modellierung zum untersuchten FDW angenommene Modellstruktur empirisch überprüft, indem die Modellpassungen eines eindimensionalen und eines mehrdimensionalen IRT-Modells miteinander verglichen werden (9.5).

Anschließend beinhaltet Kapitel 10 die Beschreibung des optimierten und hinsichtlich verschiedenen Validitätsaspekten untersuchten Testinstruments in seiner finalen Form. Dabei werden zunächst übergeordnete Testmerkmale beschrieben (10.1) und anschließend exemplarisch mehrere Beispielaufgaben erläutert (10.2).

Untersuchung des Wissenserwerbs im FDW zum Einsatz digitaler Medien (FF2)

Dieses optimierte Testinstrument wird anknüpfend eingesetzt, um die Entwicklung des FDW zum Einsatz digitaler Medien über die beteiligten Seminare im Verbundprojekt zu untersuchen. Diese Evaluation basiert auf der Befragung der an den Seminaren teilnehmenden Physiklehramtsstudierenden im Prä-Post-Design (Abbildung 7-1). Im Rahmen des Verbundprojekts werden dabei neben dem entwickelten Test zum FDW zum Einsatz digitaler Medien weitere Instrumente (z. B. zu Einstellungen und motivationalen Orientierungen) eingesetzt (5.1). Die Stichprobe für diese Evaluation besteht aus $N_1 = 70$ Physiklehramtsstudierenden und wird in 11.1 näher beschrieben.

Zur Beantwortung der FF2 werden die mittleren Personenfähigkeiten (IRT-Analyse) im gemessenen FDW vor und nach den Lehrveranstaltungen mithilfe eines t-Tests für abhängige Stichproben (z. B. Bortz & Schuster, 2010) miteinander verglichen. Die Ergebnisse dieser Evaluation hinsichtlich des FDW zum Einsatz digitaler Medien werden in 11.2 beschrieben.

7.2 Vorgehen zum qualitativen Teil

Im qualitativen Teil dieser Arbeit wird eine Interviewstudie mit einer Gelegenheitsstichprobe von $N_2 = 19$ der befragten Studierenden nach Besuch der Lehrveranstaltungen und dem Posttest durchgeführt (Abbildung 7-1). Diese retrospektiven Interviews finden vom Wintersemester 2021/22 bis Wintersemester 2022/23 in drei Durchgängen an den Standorten Aachen, Graz und Tübingen statt. Weitere Merkmale der Stichprobe zur Interviewstudie werden in 12.1 beschrieben.

Gestaltung des Interviewleitfadens und der Interviews

Die retrospektiven Interviews werden als leitfadengestützte Interviews (z. B. Helfferich, 2011) durchgeführt und dienen der Untersuchung von (1) empfundener Lernförderlichkeit bestimmter Seminarelemente, (2) Ursachen für individuelle Veränderungen im Antwortverhalten von Prä- zu Posttest und (3) Zusammenhang von individuellen Veränderungen im Antwortverhalten mit den Seminarerfahrungen. Der dazu entwickelte Interviewleitfaden beinhaltet daher verschiedene Fragen, um diese Untersuchungsziele zu adressieren.

Die Gestaltung des Leitfadens erfolgt in Anlehnung an das SPSS-Prinzip (Helfferich, 2011), welches die vier Schritte *Sammeln*, *Prüfen*, *Sortieren* und *Subsummieren* beinhaltet. Zunächst werden sämtliche Fragen gesammelt, die für das Untersuchungsziel von Interesse sein könnten. Anknüpfend folgt das Prüfen der Fragen, z. B. dahingehend, welche als Faktenfragen in eine kurze schriftliche Abfrage ausgelagert werden können oder welche bereits bekanntes Vorwissen adressieren. Die danach übrigen Fragen werden schließlich in Fragebündel sortiert und strukturiert – beispielsweise inhaltlich oder chronologisch. Im letzten Schritt wird erfolgt die Festlegung von Oberfragen zu Fragebündeln, unter die die passenden Fragen und Stichworte subsummiert werden. Diese untergeordneten Fragen und Stichworte können auch als Nachfragen oder Impulse für das Interview dienen, wenn eine Erzählung abgeschlossen ist oder ins Stocken gerät (Helfferich, 2011).

Der entwickelte Interviewleitfaden, der Ablauf der Interviews sowie die im Interview verwendeten Zusatzmaterialien werden in 12.2 beschrieben.

Analyse der Interviews mithilfe qualitativer Inhaltsanalyse

Zur Auswertung werden die Interviews per Audio aufgenommen und anschließend transkribiert. Die Transkription sowie die weitere Auswertung und Analyse erfolgt mithilfe der QDA-Software (*Qualitative Data Analysis*) *MAXQDA* (z. B. Rädiker & Kuckartz, 2019). Die Interviewtranskripte werden zur Untersuchung von FF3 mithilfe der dafür in der Bildungsforschung bewährten und verbreiteten Methode der qualitativen Inhaltsanalyse (Gläser-Zikuda, 2013) ausgewertet. Die qualitative Inhaltsanalyse stellt eine regelgeleitete Analyse von Texten (z. B. Interviewtranskripte) mit dem Ziel der systematischen Beschreibung der Textbedeutung dar (z. B. Schreier, 2014). Im Zentrum für diese Systematisierung steht dabei das Kategoriensystem zur Zuordnung von Kategorien zum Textmaterial (Stamann et al., 2016). In dieser Arbeit wird als Form die inhaltlich-strukturierende qualitative Inhaltsanalyse (Kuckartz, 2018) gewählt, da thematische Kategorien für die Strukturierung und inhaltliche Bündelung der genannten Gründe für veränderte Testantworten zur Beantwortung der Forschungsfragen am besten geeignet sind. Die Zielstellungen anderer Formen wie der typenbildenden oder evaluierenden qualitativen Inhaltsanalyse (Kuckartz, 2018) sind für diese Arbeit weniger bedeutsam, da weder das Bilden von Typen anhand des Interviewmaterials adressiert wird, noch (hauptsächlich) Kategorien zur Bewertung oder Klassifizierung im Sinne einer ordinalen Ausprägung erwartet werden.

Im Ablauf der inhaltlich-strukturierenden qualitativen Inhaltsanalyse erfolgt die Kategorienbildung und Kodierung in mehreren Schritten. Zunächst wird das Material anhand thematischer Hauptkategorien kodiert, welche überwiegend deduktiv ausgehend vom Interviewleitfaden entwickelt werden. Anknüpfend werden mit einer Hauptkategorie kodierte Textstellen zusammengestellt und anhand dessen induktive Subkategorien am Material entwickelt. Schließlich erfolgt erneut eine Kodierung des gesamten Datenmaterials mithilfe dieser (Sub-)Kategorien (Kuckartz, 2018). Im Zuge des regelgeleiteten Verfahrens der qualitativen Inhaltsanalyse betont Kuckartz (2018) unabhängig von der Form der Analyse und der Art der Kategorienbildung die Bedeutung der Kategoriendefinition, welche während der Konstruktion des Kategoriensystems für jede Kategorie formuliert werden sollte. Als allgemeines

Schema schlägt Kuckartz folgende Darstellung für Kategoriendefinitionen vor (Tabelle 7-1).

Tabelle 7-1: Allgemeines Schema für Kategoriendefinitionen (Kuckartz, 2018, S. 40)

| Name der Kategorie | | Möglichst prägnante Bezeichnung |
|----------------------------------|--------------|--|
| Inhaltliche Beschreibung | Beschreibung | Beschreibung der Kategorie, unter Umständen mit theoretischer Anbindung |
| Anwendung der Kategorie | | „Kategorie xy“ wird kodiert, wenn folgende Aspekte genannt werden... |
| Beispiele für Anwendungen | | Zitate mit Quellenangaben (Dokument, Absatz) |
| Weitere Anwendungen (optional) | | Die Kategorie wird auch kodiert, wenn... Zitate mit Quellenangaben (Dokument, Absatz) |
| Abgrenzung zu anderen Kategorien | | Die Kategorie wird nicht kodiert, wenn...: ...sondern in diesem Fall wird „Kategorie z“ verwendet Zitate mit Quellenangaben (Dokument, Absatz) |

An den oben genannten Schritten zur inhaltlich-strukturierenden Inhaltsanalyse orientiert, werden in der vorliegenden Arbeit somit inhaltliche Kategorien sowohl deduktiv als auch induktiv gebildet. Die deduktiv gebildeten Hauptkategorien ergeben sich dabei insbesondere aus der Struktur des Interviewleitfadens und weiteren Vorüberlegungen zum Interview. Das in dieser Arbeit entwickelte Kategoriensystem zur qualitativen Inhaltsanalyse wird in 12.3 vorgestellt und anhand von Beispielen erläutert.

Im Vorgehen zur qualitativen Inhaltsanalyse als bewährte Auswertungsmethode der qualitativen Sozialforschung ist die Einhaltung und Überprüfung von Gütekriterien ebenfalls bedeutsam, um die Qualität des Forschungsprozesses abzusichern (Gläser-Zikuda, 2013). Da „klassische“ Gütekriterien der quantitativen Sozialforschung (Objektivität, Reliabilität und Validität; 3.3.1) nicht im gleichen Sinne für qualitative Zielsetzungen zu interpretieren sind, wurden spezifische qualitative Gütekriterien entwickelt, welche bei der qualitativen Inhaltsanalyse zu berücksichtigen sind (Gläser-Zikuda, 2013, S. 115; Mayring, 2015, S. 125):

- Verfahrensdokumentation zur Nachvollziehbarkeit
- Regelgeleitetheit
- Argumentative, konsensuelle und kommunikative Validierung
- Triangulation (Kombination qualitativer und quantitativer Analyse-schritte)
- Intercoder-Reliabilität oder Interrater-Reliabilität

Einige Kriterien betreffen dabei nicht nur die inhaltsanalytische Auswertungstechnik, sondern gelten für den gesamten qualitativen Forschungsprozess (Kuckartz, 2018). So ermöglicht ein genaues Festhalten des Vorgehens im Forschungsprozess (sowie in der eigentlichen qualitativen Inhaltsanalyse), dass Außenstehende diese Prozesse nachvollziehen können. Ein schrittweises, systematisches und durch Regelgeleitetheit gezeichnetes Vorgehen erleichtert ebenso die Nachvollziehbarkeit, wobei die prinzipielle Offenheit der qualitativen Forschung nicht durch zu enge Vorgaben eingeschränkt werden sollte (Gläser-Zikuda, 2013). Für die Methode der qualitativen Inhaltsanalyse stellt die Regelgeleitetheit bereits ein zentrales Merkmal dar (z. B. Mayring, 2000; Schreier, 2014). Um einen Beitrag zur Absicherung der Ergebnisse und deren Interpretation zu erhalten, können diese mit anderen Forschenden (argumentative Validierung), mit den Mit-Auswertenden (konsensuelle Validierung) sowie (erneut) mit den Befragten (kommunikative Validierung) diskutiert werden (Gläser-Zikuda, 2013). Die Interrater-Reliabilität als Maß für die Prüfung der Übereinstimmung zwischen verschiedenen Auswertenden stellt ein spezifisch inhaltsanalytisches Gütekriterium dar (Kuckartz, 2018; Mayring, 2015). Dabei wird neben der Angabe der prozentualen Übereinstimmung von z. B. zwei Kodierenden häufig auf Reliabilitätskoeffizienten zurückgegriffen, die eine Korrektur um zufällig übereinstimmende Kodierungen vorsieht (z. B. Cohens κ ; 3.3.1). Die Übereinstimmung kann zudem über konsensuelles Kodieren sichergestellt werden, indem gleiche Interviews von zwei Auswertenden unabhängig kodiert und die Kodierungen anschließend gemeinsam überprüft werden (Kuckartz, 2018).

Ausgehend von der Kategorienbildung anhand der inhaltlich-strukturierenden qualitativen Inhaltsanalyse untersucht diese Arbeit, inwiefern die Erfahrungen in den Seminaren zu einer Veränderung im FDW führen und welche weiteren Gründe es für Veränderungen in der Beantwortung der Aufgaben

gibt (13.1). Im Rahmen von FF3 wird dabei der Zusammenhang zwischen genannten Gründen für veränderte Testantworten und Inhalten oder Erfahrungen aus dem Seminar untersucht (13.2). Die Erkenntnisse aus dieser Analyse werden schließlich als Gestaltungswissen genutzt, um Hypothesen für Lerngelegenheiten zur Förderung FDW zum Einsatz digitaler Medien zu generieren (13.3).

Im weiteren Verlauf der Arbeit erfolgt eine Aufteilung in den quantitativen Teil und qualitativen Teil, um gewählte Operationalisierungen und Ergebnisse beider Teile vorzustellen. Die nachfolgenden Kapitel 8 bis Kapitel 11 zählen dabei zum quantitativen Teil und adressieren die Entwicklung des Testinstruments zum FDW zum Einsatz digitaler Medien (Kapitel 8), die Gestaltung und Ergebnisse der verschiedenen Studien zur Untersuchung der Validität (Kapitel 9), die Vorstellung des finalen Testinstruments (Kapitel 10) sowie die quantitative Untersuchung des fachdidaktischen Wissenserwerbs über die beteiligten Seminare (Kapitel 11). Kapitel 12 und Kapitel 13 zählen zum qualitativen Teil der Arbeit. Sie adressieren die Beschreibung der qualitativen Interviewstudie (Kapitel 12) und der Ergebnisse dieser zur Identifikation lernförderlicher und lernhinderlicher Seminarelemente sowie zur Entwicklung von Hypothesen zur Seminargestaltung (Kapitel 13).

Quantitativer Teil

8 Testentwicklung und Pilotierung

In diesem Kapitel werden die Entwicklung des Testinstruments sowie die anknüpfende Pilotierung beschrieben. Dafür wird in 8.1 zunächst die Modellierung des FDW zum Einsatz digitaler Medien vorgestellt, welche die inhaltliche Grundlage zur Aufgabenentwicklung bildet. Anknüpfend wird in 8.2 die systematische Aufgabenentwicklung mithilfe des Itementwicklungsmodells dieser Arbeit beschrieben. In 8.3 erfolgt die Beschreibung der Pilotierung des entwickelten Testinstruments und daraus resultierenden Schritten zur Überarbeitung. Schließlich werden in 8.4 Ergebnisse zur Testgüte des überarbeiteten Testinstruments vorgestellt. Das finale Testinstrument wird in Kapitel 10 anhand exemplarischer Aufgabenbeispiele beschrieben.

8.1 Modellierung des FDW zum Einsatz digitaler Medien

Ausgehend von der Definition zur Facette (Digitale) Medien im Projekt ProfiLe-P (Gramzow et al., 2013; Riese et al., 2017) wurde für diese Arbeit eine differenzierte Modellierung des FDW zum Einsatz digitaler Medien erstellt, welche verschiedene Inhalte als Grundlage für eine systematische Aufgabenentwicklung berücksichtigen sollte (7.1). Die zugrundeliegende Definition der Facette lautete wie folgt:

„Die Facette (Digitale) Medien beinhaltet Wissen von Möglichkeiten und Anforderungen bei der Mediennutzung und die inhaltsspezifische angemessene Nutzung von Medien (vgl. Korneck et al., 2010). Des Weiteren sollte Wissen über Einsatzszenarien für e-Learning und über ein mögliches Angebot und die Bewertung eben dieser im Unterricht vorhanden sein (vgl. Korneck et al., 2010). Auch das Wissen über den sinnvollen Einsatz von Computern zur Messwerterfassung und -auswertung, von Simulationen, Modellbildungsprogrammen, Lernprogrammen und ähnlichem fällt in diesen Bereich (vgl. Korneck et al., 2010). Außerdem wird das Wissen über den Einsatz von Medien zur Differenzierung dieser Facette zugeordnet.“ (Gramzow, 2015, S. 101)

Zur Erweiterung und Ausgestaltung der bestehenden Definition wurde fachdidaktische Literatur zum Medieneinsatz (z. B. Girwidz, 2020b, 2020c; Hoyer & Girwidz, 2019; Lampe et al., 2015; Laumann et al., 2019; Rutten et al., 2012;

Wieman et al., 2010; Wiesner et al., 2011; Wolf & Kulgemeyer, 2016) recherchiert und allgemein-didaktische Literatur zum Medieneinsatz adaptiert (z. B. Mayer, 2009; Sweller, 1994; Weidenmann, 2002). Weiterhin wurden zur Entwicklung der Modellierung auch die Kooperationspartner im Verbundprojekt DiKoLeP einbezogen und je Standort um eine Ausgestaltung der Facette (Digitale) Medien auf Basis der bestehenden Definition gebeten. Zudem wurden bestehende Strukturierungen und Rahmenmodelle hinsichtlich digitaler Kompetenzen oder Anforderungen an Lehrkräfte diesbezüglich (TPACK: Mishra & Koehler, 2006; DigCompEdu: Redecker, 2017; Anforderungen der KMK: KMK, 2016; 2019; Orientierungsrahmen DiKoLAN: Becker et al., 2020; 2.3.1) herangezogen.

In einer tabellarischen Übersicht wurden darauf basierend mögliche Inhalte für die kategorienbasierte Beschreibung des fokussierten Wissensbereichs gesammelt. Grundlage für diese Tabelle bildete die eigene Ausgestaltung mit vorläufigen Kategorien auf Basis der Ausgangsdefinition zur Facette (Digitale) Medien. Weiterhin wurden in der Tabelle die Ausgestaltungen der Projektpartner im Verbund sowie ausgewählte Inhalte des naturwissenschaftsspezifischen Orientierungsrahmens DiKoLAN und der weiteren (fachunspezifischen) Rahmenmodelle bzw. Strukturierungen integriert. Auf diese Weise wurden die eigenen recherchierten Aspekte erweitert, indem wesentliche Kernelemente identifiziert, Dopplungen zusammengefasst oder weitere Kategorien hinzugefügt wurden. Anknüpfend erfolgten eine Ausschärfung und Fokussierung auf zentrale Aspekte. Einerseits bedeutete dies die inhaltliche Fokussierung aus fachdidaktischer Perspektive auf physikspezifische und -typische Medien und Situationen zum Medieneinsatz (2.3.2); andererseits die Abgrenzung zu weiteren Facetten im zugrundeliegenden Modell zum FDW in Physik (Abbildung 2-4).

So erfolgte eine Kürzung der Übersicht dahingehend, wenn Aspekte zu allgemein oder nicht spezifisch für den Physikunterricht waren oder wenn Aspekte anderen Facetten des FDW (z. B. Experimente, Aufgaben, Instruktionsstrategien) zuzuordnen waren. Da sich physiktypische Medien nicht-digitaler Art (z. B. Experimente ohne digitale Anteile oder klassische Arbeitsblätter) bereits in anderen Facetten des zugrundeliegenden Rahmenmodells zum FDW in Physik wiederfinden, ging die Ausschärfung zur Beschreibung der

Facette (Digitale) Medien im Zuge der Modellierung mit einer Fokussierung auf ausschließlich digitale Medien einher. Demnach zählen solche Aspekte und Inhalte zum fokussierten Wissensbereich, die speziell den Physikunterricht betreffen und den Mehrwert des Digitalen betrachten (z. B. im Sinne der oberen Ebenen im SAMR-Modell; 2.3.3). Zum Teil wurden auch grundsätzlich allgemein-mediendidaktische oder lernpsychologische Aspekte in der Modellierung belassen, jedoch wurde markiert, dass diese einen konkreten Bezug zum Physikunterricht bzw. zum Fach Physik benötigen, um für die Aufgabenentwicklung zum FDW zum Einsatz digitaler Medien nutzbar zu sein. Diese tabellarische Übersicht ist in Tabelle 8-1 skizziert, um einen Eindruck zur Struktur der Tabelle und dem Vorgehen zur Kürzung zu vermitteln; die Inhalte werden zur Übersichtlichkeit nur exemplarisch dargestellt.

Tabelle 8-1: exemplarische Darstellung der tabellarischen Übersicht zur Modellierung der Facette (Digitale) Medien des FDW bzw. des FDW zum Einsatz digitaler Medien im Fach Physik

| | mögliche Kategorien | eigene Ausgestaltung | Standort Graz | Standort Tübingen | DiKoLAN | KMK, TPACK, DigCompEdu |
|--------------------|---------------------------------------|---|--|--|--|---|
| aus Gramzow (2015) | Möglichkeiten bei der Mediennutzung | <ul style="list-style-type: none"> • Erweiterung des Erfahrungskegels (Wiesner et al., 2011) • Aspekte von Multimedia: Multimodalität, Multicodierung & Interaktivität (Girwidz, 2020c, Wiesner et al., 2011) | <ul style="list-style-type: none"> •... •... | <ul style="list-style-type: none"> • Erweiterung von Erfahrungen • ... | <ul style="list-style-type: none"> •... •... | <ul style="list-style-type: none"> • kann sich positiv auf individualisierte, selbstgesteuerte sowie kollaborative Lernprozesse und -ergebnisse auswirken (KMK, 2016) • ... |
| | Anforderungen bei der Mediennutzung | <ul style="list-style-type: none"> •... •... | <ul style="list-style-type: none"> •... •... | <ul style="list-style-type: none"> •... •... | <ul style="list-style-type: none"> •... •... | <ul style="list-style-type: none"> •... •... |
| | Inhaltsspezifische Nutzung von Medien | <ul style="list-style-type: none"> • Arbeitsblätter bei Schülerexperimenten...(Girwidz, 2020b) • Physikalische Versuche am Arbeitsprojektor (Girwidz, 2015a) •... | <ul style="list-style-type: none"> •... •... | <ul style="list-style-type: none"> •... •... | <ul style="list-style-type: none"> •... •... | <ul style="list-style-type: none"> • Medieneinsatz zum Anknüpfen an vorhandenes Wissen, Beseitigen von Lernproblemen/-schwierigkeiten (Mishra & Koehler, 2006) •... |
| | ... | <ul style="list-style-type: none"> •... | <ul style="list-style-type: none"> •... | <ul style="list-style-type: none"> •... | <ul style="list-style-type: none"> •... | <ul style="list-style-type: none"> •... |

durchgestrichen: zusammengefasste Dopplungen; **blaue Schrift:** grundsätzlich nicht fachspezifisch, aber mit Potenzial dazu, d.h. zur Aufgabenentwicklung muss ein Physikbezug erfolgen; **türkis markiert:** eher zur Facette *Aufgaben* gehörig; **grün markiert:** eher zur Facette *Experimente* gehörig; **gelb markiert:** zu allgemein bzw. nicht fachtypisch/-spezifisch genug, um eine Rolle im Testinstrument zu spielen.

Die tabellarische Übersicht wurde mit den markierten Kürzungsvorschlägen der Autorin erneut im Verbundprojekt zur Diskussion gestellt. Entsprechend der Diskussionsergebnisse wurde die tabellarische Übersicht schließlich auf eine (lineare) stichwortartige Beschreibung reduziert.

Die finale Modellierung des FDW zum Einsatz digitaler Medien im Fach Physik berücksichtigt dabei vier verschiedene Kategorien als Grundlage zur Entwicklung der Testaufgaben: *fachbezogene Grundlagen*, *digitale Messwerterfassung*, *Simulationen* und *Erklärvideos* (Abbildung 8-1).

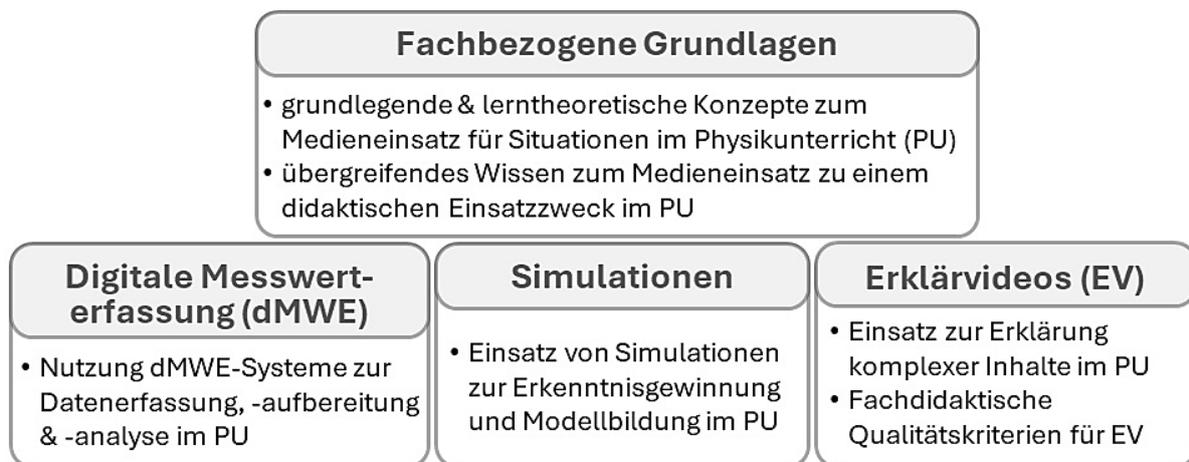


Abbildung 8-1: Modellierung zum FDW zum Einsatz digitaler Medien im Fach Physik mit den vier Kategorien *fachbezogene Grundlagen*, *digitale Messwerterfassung (dMWE)*, *Simulationen* und *Erklärvideos (EV)*

Die übergreifende Kategorie **fachbezogene Grundlagen** ergibt sich aus der Bedeutung grundlegender Aspekte zum Medieneinsatz in spezifischen Situationen des Physikunterrichts. Sie beinhaltet übergreifendes Wissen über Möglichkeiten und Anforderungen bei der Mediennutzung speziell im Physikunterricht sowie relevantes Wissen zu lernpsychologischen Grundlagen wie der CTML und CLT (Mayer, 2009; Sweller, 1994; 2.3.3) für Situationen im Physikunterricht. Auch übergreifendes Wissen über den Einsatz digitaler Medien zu einem bestimmten Einsatzzweck (z. B. Förderung des konzeptionellen Verständnisses oder Differenzierung) wird zu dieser Kategorie gezählt.

Die Kategorie **digitale Messwerterfassung** beinhaltet das Wissen über die sinnvolle Nutzung von Messwerterfassungssystemen zur (automatisierten) Datenerfassung, -aufbereitung und -analyse im Physikunterricht beispielsweise zur Erweiterung experimenteller Möglichkeiten oder Auslagerung von

Routinearbeiten zur Fokussierung auf die Dateninterpretation (Girwidz, 2020c; Lampe et al., 2015). Dabei werden sowohl vollständige Systeme von Lehrmittelfirmen als auch die Messsensorik von mobilen Endgeräten oder selbstgebaute Systeme mit Mikrocontrollern und externen Sensoren eingeschlossen. Ebenso zählt das Wissen über den sinnvollen Einsatz der Videoanalyse, bei der Messwerte aus Videoaufzeichnungen von Bewegungen gewonnen werden, als gängiges Verfahren zur Untersuchung von Bewegungen im Physikunterricht (Wilhelm, 2020) zu dieser Kategorie. Die Kategorie findet sich auch als fachspezifischer Bereich im Orientierungsrahmen DiKoLAN wieder (2.3.1).

Die Kategorie **Simulationen** enthält das Wissen über Simulationen als Modellierungen oder (vereinfachte) Rekonstruktionen von Phänomenen oder Systemen, die durch Veränderung von Parametern interaktiv gesteuert werden können (Blake & Scanlon, 2007; Wiesner et al., 2011). Die Kategorie umfasst das Wissen über den sinnvollen Einsatz von Simulationen als fachtypische Arbeitsweise zur Erkenntnisgewinnung und Modellbildung im Physikunterricht (Becker et al., 2020; Wieman et al., 2010), beispielsweise durch die Visualisierung nicht sichtbarer Sachverhalte oder die Vermittlung von physikalischen Zusammenhängen durch einfache systematische Variablenkontrolle (Girwidz, 2020c; Rutten et al., 2012). Auch diese Kategorie stellt einen Bereich im Orientierungsrahmen DiKoLAN dar (2.3.1).

Zu der Kategorie **Erklärvideos** zählt das Wissen über die fachdidaktische Qualität und den Einsatz von Erklärvideos im Physikunterricht, z. B. zum Erklären komplexer physikalischer Inhalte (Kulgemeyer, 2018a, 2018b). Aufgrund der zunehmenden Bedeutung von Erklärvideos für den Physikunterricht sowie in der physikdidaktischen Forschung (2.3.2) werden diese als eigene Kategorie im FDW zum Einsatz digitaler Medien berücksichtigt.

Alle vier Kategorien stellen zudem thematische Kerninhalte der untersuchten physikdidaktischen Lerngelegenheiten zum Einsatz digitaler Medien dar (5.2.1; Abbildung 5-1). Die lineare (stichwortartige) Version dieser Modellierung des FDW zum Einsatz digitaler Medien bzw. der Facette Digitale Medien des FDW mit den entsprechenden Verweisen zur verwendeten Literatur ist im Anhang der Arbeit zu finden (Anhang A).

Diese Modellierung des FDW zum Einsatz digitaler Medien im Fach Physik bildet mit ihren vier Kategorien die inhaltliche Grundlage zur Erstellung der Testaufgaben für die vorliegende Arbeit. Im nachfolgenden Abschnitt wird die Aufgabenentwicklung anhand des Itementwicklungsmodells beschrieben.

8.2 Aufgabenentwicklung anhand eines Itementwicklungsmodells

In diesem Abschnitt wird das in dieser Arbeit verwendete Itementwicklungsmodell vorgestellt und die systematische Aufgabenentwicklung damit erläutert. Dabei wird auch beschrieben, wie besondere Anforderungen für die Erstellung von Mehrfachwahlaufgaben sowie die Gütekriterien bei der Aufgabenentwicklung berücksichtigt wurden.

Vorstellung des Itementwicklungsmodells zur Aufgabenerstellung

Für die Entwicklung der Testaufgaben zum FDW zum Einsatz digitaler Medien im Fach Physik wurden bestimmte Rahmenbedingungen berücksichtigt. So sollte das Testinstrument die in 8.1 beschriebene Modellierung im Sinne einer inhaltlichen Validität repräsentieren. Zudem wurde eine inhaltlich-strukturelle Passung zum bestehenden FDW-Test nach Gramzow (2015) aus dem Projekt ProfiLe-P angestrebt. Daher wurde für die Aufgabenentwicklung ein Itementwicklungsmodell erstellt, welches am Modell des Projekts ProfiLe-P (Abbildung 3-2) angelehnt ist.

Das Itementwicklungsmodell dieser Arbeit unterscheidet drei Dimensionen und ist in Abbildung 8-2 dargestellt. In der ersten Dimension finden sich die vier Kategorien zur Modellierung der Facette digitale Medien bzw. des fokussierten FDW zum Einsatz digitaler Medien. Die zweite Dimension unterscheidet drei kognitive Anforderungen *Reproduzieren*, *Anwenden* und *Analysieren* (Anderson & Krathwohl, 2001). Die dritte Dimension betrachtet den physikalischen Inhaltsbereich.

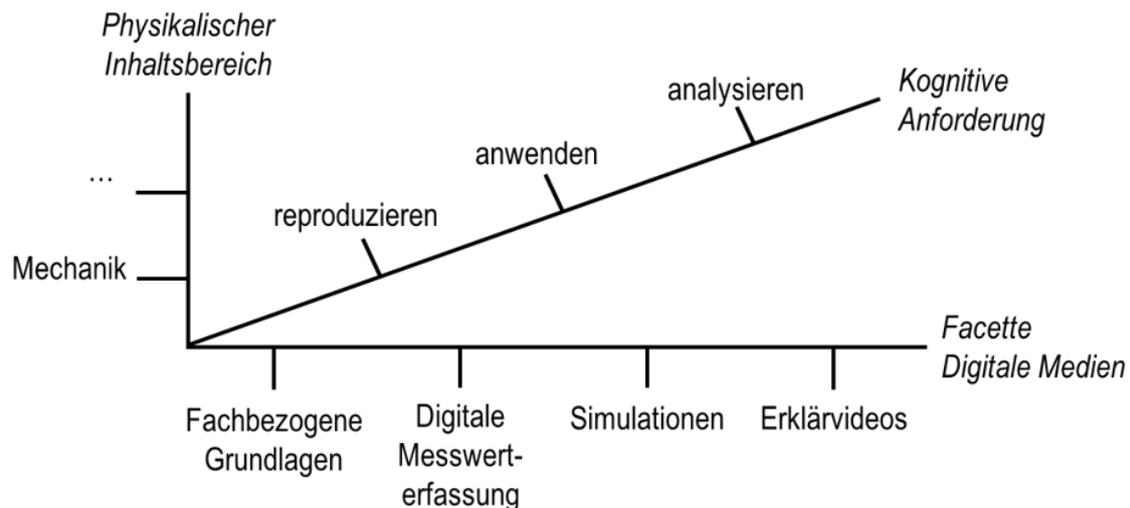


Abbildung 8-2: Itementwicklungsmodell zur Entwicklung der Testaufgaben zum FDW zum Einsatz digitaler Medien im Fach Physik (in Anlehnung an Gramzow, 2015)

Zur Aufgabenentwicklung erfolgte eine gezielte Variation der ersten beiden Dimensionen, um die verschiedenen Aspekte der modellierten Facette digitale Medien des FDW sowie verschiedene Anforderungsbereiche abzudecken. Für die Variation hinsichtlich der Kategorien wurde auf deren inhaltliche Beschreibung aus der Modellierung zur Facette digitale Medien in 8.1 zurückgegriffen. Die Variation der kognitiven Anforderungen basierte auf den Beschreibungen der drei Anforderungen *Reproduzieren*, *Anwenden* und *Analysieren* in Gramzow (2015), welche im Laufe der Testentwicklung im Hinblick auf den Kontext ‚Einsatz digitaler Medien‘ sowie das Aufgabenformat geschlossener Mehrfachwahlaufgaben adaptiert wurden. Hier ist anzumerken, dass mit dem Anforderungsbereich *Analysieren* in dieser Arbeit auch eine höhere kognitive Anforderung berücksichtigt wird als z. B. in den Kompetenzniveaus im DiKoLAN-Orientierungsrahmen, in welchem die Anwendung das höchste Niveau darstellt (2.3.1).

Als physikalischer Inhaltsbereich wurde in Passung an den FDW-Test im Projekt ProfiLe-P grundsätzlich der Bereich Mechanik fokussiert. Gleichzeitig sind auch Aufgaben möglich, die inhaltsübergreifend oder inhaltsunabhängig sind, sowie Aufgaben, die in einen anderen Inhaltsbereich eingebettet sind, wenn das betrachtete Medium dies anbietet. Die Beschreibungen zur Anwendung und Abgrenzung der einzelnen Kategorien und kognitiven Anforderungen sind in Tabelle 8-2 und Tabelle 8-3 dargestellt.

Tabelle 8-2: Beschreibung und Anwendung der Kategorien des Itementwicklungsmodells

| Kategorie | Inhaltliche Beschreibung (8.1) | Dazu zählen... |
|----------------------------|---|---|
| Fachbezogene Grundlagen | Wissen über grundlegende und lerntheoretische Konzepte zum Einsatz digitaler Medien für Situationen des Physikunterrichts | Aufgaben, in denen mehrere Beispiele zu verschiedenen Medien aufgegriffen werden. Aufgaben, in denen übergreifende Aspekte zu einem (spezifischen) Medieneinsatz abgefragt werden, die nicht für diese Medienart charakteristisch sind. |
| Digitale Messwerterfassung | Wissen über den Einsatz digitaler Messwertfassungssysteme (Lehrmittelsysteme, mobile Endgeräte, Mikrocontroller und Videoanalyse) | Aufgaben, die sich ausschließlich mit der Nutzung digitaler Messwertfassung auseinandersetzen. Aufgaben, die sich mit der Nutzung digitaler Messwertfassung im Vergleich zu traditioneller Messmethodik auseinandersetzen. Aufgaben, die eine digital gestützte Auswertung und Analyse von Messwerten mit solchen Systemen adressieren. |
| Simulationen | Wissen über den Einsatz von Simulationen im Physikunterricht z. B. zur Erkenntnisgewinnung oder Modellbildung | Aufgaben, die sich ausschließlich mit der Nutzung von Simulationen im Physikunterricht auseinandersetzen. |
| Erklärvideos | Wissen über die fachdidaktische Qualität und den Einsatz von Erklärvideos im Physikunterricht | Aufgaben, die sich ausschließlich mit Erklärvideos zur Physik oder deren Einsatzmöglichkeiten im Physikunterricht auseinandersetzen. |

Tabelle 8-3: Beschreibung und Anwendung der kognitiven Anforderungen des Itementwicklungsmodells

| Kognitive Anforderung | Inhaltliche Beschreibung | Dazu zählen... |
|------------------------------|---|---|
| Reproduzieren | Wiedergabe von Wissen zum Einsatz digitaler Medien im Physikunterricht | Aufgaben, in denen allgemeine Vorteile, Prinzipien oder Kriterien zum Einsatz digitaler Medien im Physikunterricht unabhängig von einer konkreten Unterrichtssituation erkannt werden sollen. |
| Anwenden | Anwendung des Wissens zum Einsatz digitaler Medien im Physikunterricht | <p>Aufgaben, in denen ein geeignetes Medium oder ein geeigneter Medieneinsatz zu einem bestimmten didaktischen Einsatzzweck ausgewählt werden soll.</p> <p>Aufgaben, in denen Begründungen oder Argumente für einen gegebenen Medieneinsatz in einer Situation oder zu einem Zweck ausgewählt werden sollen.</p> <p>Aufgaben, in denen Aspekte für einen bestimmten Medieneinsatz berücksichtigt werden sollen.</p> |
| Analysieren | Analyse von Unterrichtssituationen mit Medieneinsätzen oder -beispielen im Physikunterricht unter einem bestimmten fachdidaktischen Gesichtspunkt (ggfs. bewertend) | <p>Aufgaben, in denen zu einer beschriebenen Unterrichtssituation geprüft werden soll, ob allgemeine Kriterien für einen Medieneinsatz erfüllt sind oder nicht.</p> <p>Aufgaben, in denen aus einer gegebenen Unterrichtssituation Zielstellungen rekonstruiert werden sollen.</p> <p>Aufgaben, in denen mehrere verschiedene Situationen im Hinblick auf einen Gesichtspunkt bewertet werden sollen.</p> |

Ziel für die Aufgabenentwicklung war es, mehrere Aufgaben je Kategorie und kognitiver Anforderung zu erstellen. Dazu wurden zunächst Aufgabenideen gesammelt, die systematisch verschiedene mögliche Hauptziele im Sinne der verschiedenen Kategorien und kognitiven Anforderungen adressierten; beispielsweise Funktionen oder Vorteile beim Einsatz eines bestimmten Mediums nennen (*Reproduzieren*), Beispiele für einen Medieneinsatz angeben (*Anwenden*) oder einen Medieneinsatz in einer konkreten Situation unter fachdidaktischen Gesichtspunkten bewerten (*Analysieren*). Zudem wurde angestrebt, wenn möglich, konkrete Beispiele für das Fach Physik einzubinden, indem beispielsweise Unterrichtssituationen mit einem Medieneinsatz oder Beispiele zu einem Medium in den Aufgabenstamm integriert wurden (z. B. Screenshot einer Animation oder Simulation zu einem physikalischen Inhalt, Darstellung eines physikalischen Experimentes mit digitaler Messwerterfassung oder ein erstelltes Erklärvideo zu einem physikalischen Inhalt). Dementsprechend beinhalteten diese Aufgabenideen folgende Aspekte:

- angestrebtes Hauptziel
- Zuordnung zu kognitiven Anforderungen und Kategorien
- Unterrichtssituation im Aufgabenstamm (ohne Einbettung, mit kurzer oder ausführlicher Unterrichtssituation)
- Medienbeispiel im Aufgabenstamm (mit oder ohne konkretes Medienbeispiel)
- ggfs. Einbettung in einen physikalischen Inhalt

Insbesondere für die (weniger physiktypischen) Kategorien *fachbezogene Grundlagen* und *Erklärvideos* war es bedeutsam, die Fachspezifität z. B. durch eine passende Einbettung im Aufgabenstamm zu berücksichtigen, um eine Abgrenzung zu allgemein-mediendidaktischem Wissen zu gewährleisten. Dies führte dazu, dass für diese beiden Kategorien keine Items zur kognitiven Anforderung *Reproduzieren* entwickelt wurden. Die Gewährleistung einer Fachspezifität für diese Kategorien erfordert eine Einbettung in konkrete Situationen bzw. eine Darstellung konkreter Medienbeispiele für den Physikunterricht, wofür die situationsunabhängige Anforderung *Reproduzieren* nicht passend ist (Tabelle 8-3).

Berücksichtigung besonderer Anforderungen für die Entwicklung von Mehrfachwahlaufgaben

Zur ökonomischen und gleichzeitig objektiven Auswertung wurde ein geschlossenes Aufgabenformat (Mehrfachwahlaufgaben) für den FDW-Test angestrebt (6.2). So wurden die geschlossenen Mehrfachwahlaufgaben ausgehend von den zuvor beschriebenen Aufgabenideen erstellt. Die erstellten Aufgaben folgen dabei einem einheitlichen Aufbau. Im Aufgabenstamm wird (für die Anforderungen *Anwenden* und *Analysieren*) eine Unterrichtssituation beschrieben, die ggfs. ein Medienbeispiel enthält. Im Anschluss ist die Fragestellung so formuliert, dass die darauffolgend gelisteten Antwortalternativen jeweils mit *ja* oder *nein* bewertet werden können. Zum Beispiel kann eine Fragestellung wie folgt lauten: „*Stellen die folgenden Aussagen sinnvolle Begründungen für diesen Medieneinsatz im Physikunterricht dar?*“ (mit Bezug zu einer vorausgehend beschriebenen Unterrichtssituation). Für die Entwicklung der einzelnen Antwortoptionen war zunächst das Ziel, möglichst viele richtige und falsche Optionen (Attraktoren und Distraktoren) zu erstellen. Dabei wurden fachdidaktische Lehrwerke (z. B. Girwidz, 2020c; Wiesner et al., 2011), Beiträge aus unterrichtspraktischen Zeitschriften (z. B. Girwidz, 2013; Lampe et al., 2015; Laumann et al., 2019) sowie empirische Befunde (z. B. Becker et al., 2019; Kulgemeyer, 2018a, 2018b; Rutten et al., 2012) zur Mediennutzung im Physikunterricht als Grundlage für die Formulierung der Aufgaben und Antwortoptionen genutzt.

Insbesondere stellte die Formulierung der Distraktoren eine Schwierigkeit dar. Diese falschen Antwortoptionen müssen so konstruiert sein, dass sie zwar eindeutig als falsch zu bewerten sind, gleichzeitig jedoch ausreichend attraktiv erscheinen. Für unwissende Testteilnehmende sollen die Antwortoptionen (d. h. Attraktoren sowie Distraktoren) gleichermaßen plausibel und attraktiv erscheinen (Döring & Bortz, 2016a). Zudem ist es bei der Entwicklung von Mehrfachwahlaufgaben wichtig, dass nicht durch Struktur oder sprachliche Formulierung der Aufgaben und Antwortalternativen bereits Hinweise auf die richtigen und falschen Antworten gegeben werden (Gronlund, 1974; Krebs, 2004, 2019). So ist es beispielsweise ungünstig, wenn sprachliche oder formale Hinweise die Aufmerksamkeit auf wenige

Antwortoptionen einschränken, da somit letztendlich zwischen weniger Alternativen ausgewählt werden muss. Richtige Antwortoptionen sollten also nicht aus sprachlichen oder formalen Gründen herausstechen, z. B. wenn sie als einzige Antworten grammatikalisch zur Frage passen, gegenüber falschen Optionen deutlich länger sind oder ein höheres oder formelleres Sprachniveau aufweisen (Gronlund, 1974; Krebs, 2004, 2019; Moosbrugger & Brandt, 2020). Ebenso sollten in den Antwortoptionen Extremformulierungen wie z. B. *niemals* oder *alle* sowie das Aufgreifen von Schlüsselwörtern aus der Fragestellung vermieden werden (Gronlund, 1974; Krebs, 2004, 2019). Die richtigen Antworten sollten zudem nicht immer an den gleichen Positionen stehen, sondern über das Instrument ausgeglichen verteilt sein, um den Einfluss von Antworttendenzen (z. B. Tendenz zur Mitte) zu verringern (Krebs, 2004, 2019).

Aus den erstellten Antwortoptionen konnten durch einen intensiven Austausch im Projektverbund die geeignetsten Optionen (z. B. hinsichtlich Eindeutigkeit, Plausibilität und Unabhängigkeit untereinander) für die Testaufgaben ausgewählt werden, sodass jede Mehrfachwahlaufgabe schließlich vier bis acht Antwortoptionen enthielt. Dabei wurden bewusst viele Optionen je Aufgabe für die Pilotierung eingebunden, um diese zu erproben und die Aufgaben anschließend hinsichtlich möglicherweise nicht funktionierender Antwortoptionen zu überarbeiten (8.3). Vor dem Einsatz der Aufgaben wurde zudem geprüft, dass die Häufigkeiten der Positionen richtiger Antwortalternativen über das gesamte Instrument gleichmäßig verteilt ist.

Zudem war bei der Erstellung der geschlossenen Mehrfachwahlaufgaben zu bedenken, dass dieses Antwortformat im Allgemeinen Wiedererkennungslösungen (z. B. Erkennen von gelerntem Faktenwissen) einfordert, da die Teilnehmenden keine eigenen Antworten konstruieren müssen (Döring & Bortz, 2016a; Moosbrugger & Brandt, 2020). Für die Aufgabenentwicklung in dieser Arbeit wird dem durch die Einbindung spezifischer Situationen in Form kurzer Unterrichtsvignetten (3.2.2) entgegengewirkt. So wird in den meisten Aufgaben eine situationsbezogene Analyse oder Anwendung auf eine Situation notwendig, sodass für eine richtige Beantwortung das Wiedererkennen von Lehrbuchwissen nicht genügt.

Weiterhin wird die Beantwortung geschlossener Mehrfachwahlaufgaben anders als bei offenen Formaten durch mögliches Raten beeinflusst. Diesem Einfluss kann durch die Art der Bepunktung der Aufgaben begegnet werden (z. B. Döring & Bortz, 2016a). So werden die Aufgaben in der vorliegenden Arbeit durch eine Schwellenbepunktung bewertet (Krebs, 2004, 2019), welche die Antwortoptionen einer Aufgabe inhaltlich bündelt sowie den Einfluss des Ratens auf die zu erreichenden Punkte je Aufgabe etwas reduziert. Die gewählte Schwellenbepunktung ermöglicht je Mehrfachwahlaufgabe das Erreichen von 0, 1 oder 2 Punkten je nach Anzahl richtig beantworteter Antwortalternativen. In 10.1 wird diese Bepunktungsregel genauer beschrieben.

Berücksichtigung von Gütekriterien bei der Aufgabenentwicklung

Das beschriebene Vorgehen zur Aufgabenentwicklung hatte zum Ziel, bereits gewisse Aspekte der Testgüte (3.3.1) im Verlauf der Testentwicklung zu berücksichtigen. So wird die Objektivität im Sinne der Durchführungsobjektivität dadurch gewährleistet, dass sich das entwickelte Testinstrument in einer standardisierten Online-Befragung unabhängig von einer testleitenden Person einsetzen lässt. Die Auswertungsobjektivität ist aufgrund der geschlossenen Mehrfachwahlaufgaben trivialerweise erfüllt. Die Berücksichtigung der oben genannten Hinweise und Anforderungen für die Erstellung von Mehrfachwahlaufgaben und deren Antwortoptionen zielt auf die Entwicklung möglichst reliabler und valider Testaufgaben ab.

Zudem lässt sich im Hinblick auf die Inhaltsvalidität bewerten, inwiefern die Testaufgaben das zugrundeliegende Itementwicklungsmodell abbilden. Dazu wurde die Passung der Aufgaben zum Itementwicklungsmodell untersucht, indem die Zuordnung der Aufgaben zu den Kategorien der Facette und den kognitiven Anforderungen durch die Projektpartner im Verbundprojekt DiKoLeP unabhängig vorgenommen und mit der Zuordnung durch die Autorin verglichen wurde. Die daraus resultierenden Beurteiler:innenübereinstimmungen (Cohens κ) lagen zwischen $\kappa = 0,35$ bis $\kappa = 0,72$ (kognitive Anforderungen) und $\kappa = 0,68$ bis $\kappa = 0,92$ (Kategorien). Anknüpfend wurden die Definitionen der Kategorien und kognitiven Anforderungen ausgeschärft (siehe Beschreibungen oben) und eine erneute Zuordnung von zwei Physikdidaktiker:innen außerhalb des Verbundprojekts vorgenommen. Dadurch konnten

verbesserte Übereinstimmungen von $\kappa = 0,54$ bis $\kappa = 0,65$ (kognitive Anforderungen) bzw. $\kappa = 0,84$ (Kategorien) erreicht werden, welche im mittlerem bis guten bzw. sehr gutem Bereich liegen (Döring & Bortz, 2016a; 3.3.1). Demnach ist die Passung der Aufgaben in das Itementwicklungsmodell als angemessen zu bewerten, was als ein Argument für die Inhaltsvalidität aufgefasst werden kann.

Insgesamt konnten durch das beschriebene Verfahren 17 Testaufgaben für die Pilotierung entwickelt werden.

8.3 Pilotierung des Testinstruments

Wie es in der Entwicklung von Leistungstests in der fachdidaktischen Forschung üblich ist (3.1), wurde auch das in dieser Arbeit entwickelte Testinstrument vor dem Einsatz in der Hauptstudie im Rahmen einer Pilotierung erprobt. In diesem Abschnitt werden die Stichprobe und Erhebung zur Pilotierung sowie daraus folgende Überarbeitungen zur Optimierung des Testinstruments vorgestellt.

Stichprobe zur Pilotierung

Die erste Testversion bestehend aus 17 Aufgaben wurde für die Pilotierung als Online-Fragebogen (über *Sosci Survey*) implementiert. Die Befragungen zur Pilotierung erfolgten zu Beginn des Sommersemesters 2021 an insgesamt sieben Hochschulen in Deutschland und Österreich¹¹. Dabei wurden 116 Lehramtsstudierende befragt; teilweise zu mehreren Messzeitpunkten, so dass für die Pilotierung $N_P = 142$ Datensätze ausgewertet werden konnten. Wie in 7.1 bereits erläutert, wurden in manchen Teilstichproben zusätzliche Instrumente zur Abgrenzung zu anderen Wissensbereichen im Sinne einer Validierung (9.1) eingesetzt. Zudem wurden demografische Daten abgefragt (Geschlecht, Alter, Hochschule/Universität, Ausbildungsphase, Studienfächer, Fachsemester, Schulabschluss, absolvierte Schulpraktika, Medieneinsatz in Schulpraktika). Die Stichprobe der 116 befragten Lehramtsstudierenden in der Pilotierung ist in Tabelle 8-4 beschrieben.

¹¹ Beteiligt waren neben den Projektstandorten Aachen, Graz und Tübingen zudem die TU Dortmund, die Universität Göttingen, die PH Karlsruhe und die Universität Paderborn.

Tabelle 8-4: Übersicht der Stichprobe zur Pilotierung des entwickelten Testinstruments

| Stichprobe | Geschlecht | Alter | Fachsemester | Studienphase | Physik als Fach |
|----------------------|------------|----------|--------------|-----------------------|------------------|
| N ₀ = 116 | 45 % w | M = 24 | M = 7 | 63 % BA | 79 % Physik |
| | 50 % m | (SD = 3) | (SD = 3) | 34 % MA 3 % Sonst. | 21 % kein Physik |

w: weiblich; m: männlich; M: Mittelwert; SD: Standardabweichung; BA: Bachelorstudium; MA: Masterstudium; Sonst.: sonstige Studiengänge. Erläuterung zu fehlenden Werten: an einem Erhebungsstandort (sechs Befragungspersonen) wurden Geschlecht und Alter nicht miterhoben.

Von den 116 Befragten konnten nur von 97 Personen Abfragen zum Abitur erfolgen¹². Zwei Drittel dieser Personen schlossen das Abitur nach Klasse 12 (G8) und circa ein Drittel nach Klasse 13 (G9) mit einer mittleren Abiturnote von 2,1 (SD = 0,6) ab. Von den 116 Befragten gaben 24 Personen (21 %) an, an noch keinem Schulpraktikum im Rahmen des Lehramtsstudiums teilgenommen zu haben; die anderen 79 % absolvierten demnach mindestens ein Praktikum. Etwa 38 % der Studierenden hatten bereits zwei oder mehr Praktika abgeschlossen.

Beschreibung der Testversion

Von den 17 Aufgaben der ersten Testversion zur Pilotierung waren 16 bereits Mehrfachwahlaufgaben mit je vier bis acht Antwortoptionen. Eine Aufgabe stellte eine Auswahlaufgabe mit anknüpfender Erläuterung im offenem Antwortformat dar – mit der Absicht, diese nach der Pilotierung mithilfe der erhaltenen Antworten zu schließen. Im Mittel benötigten die Teilnehmenden etwa 33 Minuten zur Beantwortung der zu pilotierenden Aufgaben.

Die Verteilung der Aufgaben zur Pilotierung auf die kognitiven Anforderungen und Kategorien zeigt Tabelle 8-5. Es ist zu sehen, dass für alle Kategorien

¹² Aufgrund fehlender Abfragen an einem deutschen Erhebungsstandort sowie am österreichischen Standort Graz beruhen die Angaben zum Schulabschluss bzw. dem Abitur nur auf 97 gültigen Fällen.

nahezu gleich viele Aufgaben entwickelt werden konnten. Da zur Anforderung *Reproduzieren* nur Aufgaben in den physikspezifischeren Kategorien *digitale Messwerterfassung* und *Simulationen* entwickelt wurden (8.2), ergibt sich hier eine geringere Anzahl an Aufgaben als in den kognitiven Anforderungen *Anwenden* und *Analysieren*.

Tabelle 8-5: Übersicht zur Verteilung der 17 Testaufgaben für die Pilotierung in Kategorien und kognitive Anforderungen

| | Reproduzieren | Anwenden | Analysieren | Summe |
|----------------------------|---------------|----------|-------------|-------|
| Fachbezogene Grundlagen | - | 2 | 2 | 4 |
| Digitale Messwerterfassung | 2 | 2 | 1 | 5 |
| Simulationen | 1 | 1 | 2 | 4 |
| Erklärvideos | - | 2 | 2 | 4 |
| <i>Summe</i> | 3 | 7 | 7 | 17 |

Die Einträge in der Tabelle geben die Anzahl der Aufgaben an, die dieser Kategorie oder kognitiven Anforderung zugehörig sind.

In Abbildung 8-3 ist eine Beispielaufgabe der Testversion zur Pilotierung dargestellt. Diese Mehrfachwahlaufgabe zählt zur Kategorie *Simulationen* und zur kognitiven Anforderung *Anwenden*. In der Aufgabe wird beschrieben, dass eine Lehrkraft eine Simulation zu einfachen Stromkreisen im Unterricht einsetzt. Die Befragten sollen angeben, ob die einzelnen Antwortoptionen sinnvolle Begründungen für diesen Medieneinsatz darstellen (Abbildung 8-3). Bei der Erstellung der Aufgabe wurden bewusst viele Antwortoptionen konstruiert, um nach der Pilotierung eventuelle nicht funktionierende Optionen entfernen zu können.

1. Fachdidaktisches Wissen zum Einsatz digitaler Medien – Teil 15 von 17

Eine Lehrkraft lässt Schülerinnen und Schüler im Anfangsunterricht zur Elektrizitätslehre eine Simulation zu einfachen Stromkreisen verwenden. Stellen die folgenden Aussagen sinnvolle Begründungen für diesen Medieneinsatz im Physikunterricht dar?

| | Nein | Ja |
|--|-----------------------|-----------------------|
| In einer geeigneten Simulation können nicht sichtbare physikalische Größen wie der elektrische Strom oder das elektrische Potenzial visualisiert werden. | <input type="radio"/> | <input type="radio"/> |
| Durch die Verwendung der Simulation als zentrales Medium zur Erarbeitung einfacher Stromkreise kann die Lehrkraft Zeit und Arbeitsaufwand einsparen, um schneller neue Unterrichtsinhalte durchnehmen zu können. | <input type="radio"/> | <input type="radio"/> |
| Das Überprüfen von vorab aufgestellten Vermutungen kann durch eine einfache und gezielte Variablenkontrolle in der Simulation ermöglicht werden. | <input type="radio"/> | <input type="radio"/> |
| Eine geeignete Simulation ermöglicht das Fokussieren auf zentrale Aspekte, indem sichergestellt wird, dass diese ohne überlagernde Nebeneffekte erkennbar werden. | <input type="radio"/> | <input type="radio"/> |
| Den Lernenden wird durch die Verwendung der Simulation anstelle des Realexperiments eine Interaktivität geboten, da die Simulation auf Veränderungen der Parameter reagiert. | <input type="radio"/> | <input type="radio"/> |
| Die Simulation kann als Vorbereitung auf ein folgendes analoges Realexperiment genutzt werden, um eine erste Vorstellung der physikalischen Zusammenhänge zu bekommen. | <input type="radio"/> | <input type="radio"/> |
| Die vereinfachte Darstellungsform und die Visualisierung nicht sichtbarer physikalischer Größen können zum konzeptionellen Verständnis der Lernenden von Stromkreisen beitragen. | <input type="radio"/> | <input type="radio"/> |

Abbildung 8-3: Beispielaufgabe (Mehrfachwahlaufgabe) aus der Testversion zur Pilotierung

(Vorversion der jetzigen Aufgabe 4 im finalen Test; Kategorie: *Simulationen*; kognitive Anforderung: *Anwenden*; Musterlösung: ja, nein, ja, ja, nein, nein, ja)

In Abbildung 8-4 ist eine weitere Aufgabe aus der Testversion zur Pilotierung abgebildet. Diese stellt die einzige Aufgabe mit offenem Antwortformat dar. In der Aufgabe sollen zunächst zu gegebenen didaktischen Einsatzzwecken mögliche Medien aus einer Liste ausgewählt werden. Die vorgegebenen Einsatzzwecke sind:

1. Interesse für das Thema wecken
2. Aufbau und Funktion des Plattenkondensators erlernen
3. Zusammenhang zwischen der Kapazität eines Plattenkondensators und ihren Einflussgrößen untersuchen
4. Lade- und Entladekurve eines Plattenkondensators für verschiedene Gesamtwiderstände untersuchen
5. Vorwissen zum Thema erkunden.

Jeweils zur Auswahl aufgelistet sind folgende digitale Medien:

- Übungsprogramm mit adaptiver Zuweisung von Aufgaben,
- Digitales Lernquiz,

- Video aus der Tagespresse,
- Digitale Messwerterfassung,
- Datenaufnahme mit einer Wärmebildkamera,
- Digitale Videoanalyse,
- Simulation und
- Erklärvideo.

Zu jeder Auswahl soll anknüpfend der Einsatz des Mediums zum gegebenen Einsatzzweck durch einen kurzen Freitext beschrieben werden (Abbildung 8-4). Diese Aufgabe gehört zur Kategorie *fachbezogene Grundlagen* und der kognitiven Anforderung *Anwenden*.

1. Fachdidaktisches Wissen zum Einsatz digitaler Medien – Teil 3 von 17

Im Folgenden sind verschiedene didaktische Einsatzzwecke für eine Unterrichtsstunde zum Plattenkondensator in der Sekundarstufe II aufgeführt. Wählen Sie für jeden didaktischen Einsatzzweck ein passendes Medium aus der Liste aus, dessen Einsatz sich für den jeweiligen Zweck am besten eignet. Beschreiben Sie danach kurz, wie ein entsprechender Medieneinsatz mit diesem Medium für den jeweiligen Zweck aussehen könnte.

| Didaktischer Einsatzzweck | Medium | Kurze Beschreibung des ausgewählten Medieneinsatzes |
|--|---|---|
| Interesse für das Thema wecken | [Bitte auswählen] v | |
| Aufbau und Funktion des Plattenkondensators erlernen | [Bitte auswählen] Übungsprogramm mit adaptiver Zuweisung von Aufgaben Digitales Lernquiz Video aus der Tagespresse Digitale Messwerterfassung | |
| Zusammenhang zwischen der Kapazität eines Plattenkondensators und ihren Einflussgrößen untersuchen | Datenaufnahme mit einer Wärmebildkamera Digitale Videoanalyse Simulation Erklärvideo | |
| Lade- und Entladekurve eines Plattenkondensators für verschiedene Gesamtwiderstände untersuchen | | |
| Vorwissen zum Thema erkunden | [Bitte auswählen] v | |

Abbildung 8-4: Beispielaufgabe (offene Aufgabe) aus der Testversion zur Pilotierung. (Screenshot aus dem Online-Fragebogen; Vorversion der jetzigen Aufgabe 3 im finalen Test; Kategorie: *fachbezogene Grundlagen*; kognitive Anforderung: *Anwenden*; Musterlösung zur Auswahl „Medium“: Video aus der Tagespresse, Erklärvideo, Simulation, Digitale Messwerterfassung, Digitales Lernquiz; Erwartungshorizont zur offenen Antwort in Anhang C)

Aufgabenanalysen und Überarbeitung

Die pilotierten Testaufgaben wurden in einer deskriptiven Analyse im Hinblick auf verschiedene Aspekte untersucht. Diese Analyse erfolgte klassisch (erst nach Überarbeitungen zur Pilotierung und weiteren Erhebungen mit dem optimierten Testinstrument wurde die IRT als Grundlage für Analysen mit dem vergrößerten Datensatz gewählt). So wurden in der Analyse zur Pilotierung beispielsweise zunächst Histogramme für die Verteilung der erreichten Punkte in den einzelnen Aufgaben betrachtet, um die Antworten auf Logik und mögliche fehlerhafte Werte zu prüfen. Zudem wurde damit untersucht, ob Boden- oder Deckeneffekte vorliegen (wenn die Mehrzahl der Befragten keine oder alle möglichen Punkte in einer Aufgabe erreicht hat). Weiterhin wurden die Aufgabenschwierigkeiten (mittlere relative Lösungshäufigkeiten) der Aufgaben sowie der einzelnen Antwortalternativen betrachtet, d. h. der Anteil der Befragten, der die Aufgabe oder das Item richtig gelöst hat. Die Lösungshäufigkeit einer Aufgabe sollte zwischen 20 % bis 80 % liegen (z. B. Döring & Bortz, 2016a). Für die einzelnen Antwortalternativen wurde dieser Bereich weniger streng eingehalten; hier wurden Alternativen, die von mehr als 90% der Teilnehmenden korrekt beantwortet wurden, als problematisch betrachtet und für eine mögliche Entfernung oder Umformulierung markiert. Als Analysen auf Skalenebene wurden zudem in Bezug auf die Gesamtskala Reliabilitätsanalysen (Cronbachs α) vorgenommen sowie die Trennschärfe als korrigierte Item-Skala-Korrelation überprüft (d. h. wie sehr die mit der Aufgabe erreichte Differenzierung mit der Differenzierung durch den Gesamttest einhergeht; z. B. Kelava & Moosbrugger, 2020a). Dabei wurde für die Aufgaben jeweils auch geprüft, wie sich das Entfernen der Aufgabe auf die Reliabilität auswirkt. Für die Teilstichproben, in denen auch ein Test zum pädagogischen Wissen (PW) zum Einsatz kam, konnte zudem die Korrelation der erreichten Punkte in den einzelnen Aufgaben mit der Testleistung im PW-Test untersucht werden – als Indiz dafür, dass möglicherweise nicht ausreichend fachspezifisches Wissen zum Einsatz digitaler Medien in diesen Aufgaben adressiert wird.

Mithilfe dieser verschiedenen Kriterien konnten Hinweise gesammelt werden, welche Aufgaben als „kritisch“ zu betrachten sind und ggfs. entfernt

oder überarbeitet werden sollten. Dabei wurden nur gesamte Aufgaben entfernt, wenn sie in mehreren Aspekten Probleme aufwiesen; ansonsten erfolgte eine Überarbeitung der einzelnen Antwortalternativen oder der Formulierung des Aufgabenstamms. Zudem wurde im Hinblick auf das Entfernen ganzer Aufgaben berücksichtigt, inwiefern danach noch eine angemessene Abdeckung der Kategorien und kognitiven Anforderungen des Itementwicklungsmodells gegeben ist.

Aufgaben, die einen Boden- oder Deckeneffekt aufzeigten, konnten durch das Entfernen der am seltensten oder am häufigsten richtig gelösten Antwortoptionen meist so überarbeitet werden, dass diese Effekte verschwanden. Gleichzeitig wurden stets nur so viele Optionen entfernt, dass für jede Aufgabe noch mindestens vier Antwortalternativen übrigblieben. So zeigte beispielsweise die Aufgabe aus Abbildung 8-3 einen leichten Deckeneffekt. Aus der Aufgabe wurden daher die erste und siebte Antwortoption (beides Attraktoren) entfernt, da sie sehr hohe Lösungshäufigkeiten von 94 % und 93 % aufwiesen. In manchen Aufgaben wurde zudem eine Konkretisierung oder Umformulierung der Aufgabenstellung vorgenommen. Die Aufgabe aus Abbildung 8-4 wurde nach der Pilotierung zu einer Mehrfachwahlaufgabe geschlossen. Dafür wurde nur der erste didaktische Einsatzzweck ausgewählt (Interesse am Thema wecken), um die Bearbeitungszeit des Instruments in einem angemessenen Rahmen zu halten. Die Entwicklung der Antwortoptionen für diesen Einsatzzweck erfolgte mithilfe der gegebenen Antworten der offenen Abfrage zur Beschreibung des Medieneinsatzes. Zudem dienten die Antworten zu den anderen didaktischen Zwecken als Anregung für die Formulierung der Distraktoren. Die geschlossene Aufgabe wird in 10.2.1 dargestellt.

Die Überarbeitung der Aufgaben erfolgte erneut im Austausch mit den Kollegen im Verbundprojekt DiKoLeP. Zwei Aufgaben wurden nach der Pilotierung ganz entfernt, bei einigen Aufgaben wurden einzelne Antwortalternativen verändert oder entfernt sowie teilweise die Aufgabenstellung ausgearbeitet. In Tabelle 8-6 findet sich eine Auflistung der kritischen Aufgaben in der Pilotierung sowie davon ausgehend vorgenommenen Überarbeitungen.

Tabelle 8-6: Kritische Aufgaben der Pilotierung und vorgenommene Überarbeitungen

| Aufgabe | Einordnung | Probleme | Veränderungen |
|------------------|---|--|--|
| A. 3 | Fachbezogene Grundlagen, Anwenden | Deckeneffekt, Schließung für alle Einsatzzwecke schwierig & zu viel | nur für ersten Einsatzzweck (Interesse wecken) schließen |
| A. 4 (entfernt) | Simulationen, Analysieren | Deckeneffekt, sign. Korrelation mit PW-Testscore | zwei Optionen entfernen |
| A. 5 (entfernt) | Digitale Messwerterfassung, Anwenden | Deckeneffekt, Option mit Lösungshäufigkeit > 90 %, geringe Trennschärfe | ganz entfernen |
| A. 7 (8) | Simulationen, Reproduzieren | leichter Deckeneffekt, Option mit Lösungshäufigkeit > 90 % | eine Option entfernen |
| A. 8 (14) | Digitale Messwerterfassung, Reproduzieren | Option mit Lösungshäufigkeit > 90 % | Aufgabenstellung anpassen, eine Option austauschen |
| A. 10 (7) | Fachbezogene Grundlagen, Analysieren | Bodeneffekt, geringe Trennschärfe | zwei Optionen entfernen, Aufgabenstellung anpassen |
| A. 13 (entfernt) | Erklärvideos, Anwenden | geringe Trennschärfe, Reliabilität ohne die Aufgabe höher | ganz entfernen |
| A. 15 (4) | Simulationen, Anwenden | sign. Korrelation mit PW-Testscore, leichter Deckeneffekt, Optionen mit Lösungshäufigk. > 90 % | zwei Optionen entfernen |
| A. 17 (13) | Erklärvideos, Analysieren | leichter Bodeneffekt | eine Option austauschen, Aufgabenstellung überarbeiten |

In der ersten Spalte ist in Klammern zusätzlich jeweils die Nummer der Aufgabe im finalen Testinstrument angegeben bzw. „entfernt“, wenn diese nicht mehr im finalen Test enthalten ist (Hinweis: die gelistete Aufgabe 4 wurde nicht nach der Pilotierung entfernt, sondern erst nach einer Studie zur Validierung).

Aus Tabelle 8-6 wird jedoch ersichtlich, dass dort nicht alle Testaufgaben gelistet sind, da es auch mehrere Aufgaben gab, die in der Pilotierung bereits gut funktionierten und nicht überarbeitet werden mussten. Insgesamt beinhaltete das nach der Pilotierung überarbeitete Testinstrument 15 geschlossene Mehrfachwahlaufgaben mit je vier bis sieben Antwortoptionen. Auf Basis der verschiedenen Teilstudien zur Validität ergaben sich weitere Überarbeitungen sowie die Entfernung einer Aufgabe, sodass das finale Testinstrument aus 14 Aufgaben besteht. Diese Überarbeitungen werden bei der Beschreibung der Untersuchungen zur Validität in Kapitel 9 erläutert.

Im nächsten Abschnitt werden verschiedene Merkmale zur Testgüte des optimierten Testinstruments beschrieben, welche auf weiteren Erhebungen zur Erprobung des Tests zum FDW zum Einsatz digitaler Medien basieren.

8.4 Untersuchungen zur Testgüte des optimierten Testinstruments

Das optimierte Testinstrument wurde neben der Evaluationsstudie in dieser Arbeit zusätzlich in mehreren weiteren Erhebungen bei Lehramtsstudierenden im Fach Physik eingesetzt. Insgesamt wurden dabei 399 Datensätze von 367 Studierenden (142 weiblich, 225 männlich) zum FDW zum Einsatz digitaler Medien erhoben (teilweise mehrere Messzeitpunkte). Diese Studierenden waren im Mittel im 7. Fachsemester ($SD = 3$) und 25 Jahre alt ($SD = 5$).

Für die Analyse der Testdaten in diesem vergrößerten Datensatz wurde die probabilistische Testtheorie bzw. IRT als methodische Grundlage verwendet (3.3.1). Die statistischen Analysen erfolgten mithilfe der Statistik-Software *R* und des Pakets *TAM* (*test analysis modules*: Robitzsch et al., 2024). Im Folgenden werden Untersuchungen zur Testgüte und Aufgabenqualität wie z. B. die Modellpassung der IRT-Modellierung, die EAP-Reliabilität oder die Analyse der Aufgabenschwierigkeiten erläutert.

So stellt die Schwierigkeit des Testinstruments bzw. der Testaufgaben einen wichtigen Aspekt hinsichtlich der Testqualität dar. Als erstes (klassisches) Maß für die Schwierigkeit des entwickelten Tests ergibt sich zunächst für alle Aufgaben eine angemessene mittlere erreichte Punktzahl zwischen 0,59 und 1,37 von möglichen 2 Punkten (8.2) für die betrachtete Stichprobe von 399 Datensätzen. In der probabilistischen Testtheorie oder IRT wird in diesem

Zusammenhang hingegen die Wahrscheinlichkeit betrachtet, dass eine Person eine Aufgabe richtig löst (3.3.1). Diese Lösungswahrscheinlichkeit hängt zum einen von der Schwierigkeit der Aufgabe¹³ und zum anderen von der Fähigkeit der Person ab.

In der IRT werden dabei Personenparameter (Personenfähigkeit η_i einer Person i) und Aufgabenparameter (Aufgabenschwierigkeit β_j einer Aufgabe j) auf einer gemeinsamen Skala quantifiziert („Logit-Skala“ als Differenz ($\eta_i - \beta_j$); Kelava & Moosbrugger, 2020b). Wenn die Personenfähigkeit und Aufgabenschwierigkeit gleich sind ($\eta_i = \beta_j$), liegt die Lösungswahrscheinlichkeit genau bei 50 %. Ist die Person fähiger als die Aufgabe schwierig ist ($\eta_i > \beta_j$), ist die Lösungswahrscheinlichkeit größer. Ist die Person hingegen weniger fähig als die Aufgabe schwierig ist ($\eta_i < \beta_j$), ist die Lösungswahrscheinlichkeit kleiner (z. B. Kelava & Moosbrugger, 2020b; Strobl, 2015).

Da die Bewertung der Testaufgaben für den vorliegenden FDW-Test jeweils das Erreichen von 0, 1 oder 2 Punkten ermöglicht (8.2; die Erläuterung der Bepunktungsregel erfolgt in 10.1), erfordert die Analyse eine entsprechende IRT-Modellierung, die diese Mehrstufigkeit berücksichtigt. So wurde für die Modellierung ein eindimensionales *Partial-Credit-Modell* genutzt (Masters, 1982), wobei *Thurstone Thresholds* für die Schätzung der Itemschwierigkeiten der mehrstufigen Testaufgaben verwendet wurden (Linacre, 1998). Anstelle der Schätzung eines Aufgabenparameters β_j für eine Aufgabe j , wird hier für jede Schwelle k der Aufgabe j ein Threshold-Wert ($t_{j,k}$) geschätzt – für die vorliegenden Testaufgaben also je ein Wert zum Erreichen von einem Punkt ($t_{j,1}$) und von zwei Punkten ($t_{j,2}$). Diese Thurstone Thresholds haben den Vorteil, dass sie analog interpretiert werden können wie Aufgabenschwierigkeiten bei binären Aufgaben, in denen eine Person eine Aufgabe entweder löst (1 Punkt) oder nicht (0 Punkte). Das bedeutet, dass eine Person einer Fähigkeit entsprechend eines Threshold-Wertes $t_{j,k}$, diese Punkteschwelle k mit einer Wahrscheinlichkeit von 50 % erreicht (Linacre, 1998).

¹³ Die Aufgabenschwierigkeit stellt in der IRT einen Schwierigkeitsparameter dar, bei dem ein höherer Wert einer höheren Schwierigkeit entspricht. Die „Aufgabenschwierigkeit“ in der KTT im Sinne der mittleren relativen Lösungshäufigkeit stellt hingegen einen Leichtigkeitsparameter dar, da höhere Werte einer leichteren Aufgabe entsprechen (Kelava & Moosbrugger, 2020b).

Neben einer eindimensionalen Modellierung lassen sich auch mehrdimensionale IRT-Modellierungen betrachten, die eine bestimmte Strukturierung der Aufgaben berücksichtigen. Gemäß dem Itementwicklungsmodell in dieser Arbeit (Abbildung 8-2) ist hier eine Strukturierung nach den modellierten Kategorien (*fachbezogene Grundlagen, digitale Messwerterfassung, Simulationen* und *Erklärvideos*) in einem vierdimensionalen Modell sowie eine Strukturierung nach kognitiven Anforderungen (*Reproduzieren, Anwenden, Analysieren*) in einem dreidimensionalen Modell sinnvoll. Im Folgenden werden nun Kennwerte zur Güte dieser drei IRT-Modellierungen erläutert, welche die Modellpassung, die EAP-Reliabilität sowie die Aufgabenschwierigkeiten (bzw. Schwierigkeiten der Aufgabenschwellen) adressieren. Diese statistischen Kennwerte sind in der nachfolgenden Tabelle 8-7 zusammenfassend aufgeführt.

Für die drei IRT-Modellierungen lässt sich jeweils die Modellpassung anhand verschiedener Kennwerte bewerten. So zeigt die eindimensionale IRT-Modellierung insgesamt eine annehmbare Modellpassung. Der Itemfit (gewichteter Mean Square bzw. *weighted MNSQ*) liegt für alle Items in einem guten Bereich ($0,93 < \text{MNSQ} < 1,06$)¹⁴. Ebenso zeigt sich mit einem mittleren RMSD von 0,029 (SD = 0,010) und einem SRMR von 0,058 ein angemessener Modellfit (Gäde et al., 2020)¹⁵. Für das vierdimensionale Modell nach Kategorien sowie für das dreidimensionale Modell nach kognitiven Anforderungen ergeben sich ähnliche Kennwerte (Tabelle 8-7).

¹⁴ Der MNSQ-Itemfit (Infit) ist ein Maß für die Rasch-Homogenität der Items und sollte möglichst bei 1 liegen. Nach Wright und Linacre (1994) werden Werte im Bereich von 0,8 bis 1,2 als angemessen betrachtet.

¹⁵ RMSD/RMSEA (*Root Mean Square Error of Approximation*) und SRMR (*Standardized Root Mean Square Residual*) sind Maße für die Passung von Modell und Daten und sollten möglichst nah bei 0 liegen. Nach Gäde et al. (2020) werden Werte unter 0,05 für beide Kriterien als guter Modellfit bezeichnet; als akzeptabler Modellfit gilt SRMR < 0,10 oder RMSEA < 0,08.

Tabelle 8-7: statistische Kennwerte zu Modellpassung, EAP-Reliabilitäten und mittleren Schwierigkeiten der Aufgabenschwellen für die eindimensionale sowie die beiden mehrdimensionalen Modellierungen nach Kategorien bzw. kognitiven Anforderungen

| IRT-Modellierung | Modellpassung | EAP-Reliabilitäten | mittlere Aufgabenschwierigkeiten |
|------------------|---|---|--|
| 1D | Gesamtskala (FDW zum Einsatz digitaler Medien) | 0,93 < MNSQ < 1,06 RMSD: 0,029 (SD = 0,010) SMSR 0,058 | 0,60 insgesamt: -0,116 1P-Schwelle: -0,848 2P-Schwelle: 0,617 |
| | Fachbezogene Grundlagen | 0,43 | insgesamt: 0,444 1P-Schwelle: -0,288 2P-Schwelle: 1,117 |
| 4D | Digitale Messwerterfassung | 0,93 < MNSQ < 1,09 RMSD: 0,040 (SD = 0,007) SMSR = 0,052 | 0,59 insgesamt: -0,451 1P-Schwelle: -1,093 2P-Schwelle: 0,191 |
| | Simulationen | 0,52 | insgesamt: -0,612 1P-Schwelle: -1,474 2P-Schwelle: 0,251 |
| | Erklärvideos | 0,54 | insgesamt: -0,004 1P-Schwelle: -0,979 2P-Schwelle: 0,971 |
| | Reproduzieren | 0,58 | insgesamt: -0,704 1P-Schwelle: -1,286 2P-Schwelle: -0,121 |
| 3D | Anwenden | 0,95 < MNSQ < 1,09 RMSD: 0,030 (SD = 0,010) SMSR = 0,054 | 0,50 insgesamt: 0,014 1P-Schwelle: -0,742 2P-Schwelle: 0,770 |
| | Analysieren | 0,57 | insgesamt: 0,028 1P-Schwelle: -0,788 2P-Schwelle: 0,844 |

1D: eindimensionale Modellierung; 4D: vierdimensionale Modellierung nach Kategorien; 3D: dreidimensionale Modellierung nach kognitiven Anforderungen; MNSQ: *weighted Mean Square*; RMSD: *Root Mean Square Error of Approximation*; SMSR: *Standardized Root Mean Square Residual*; EAP: *Expected-a-posteriori*; 1P: ein Punkt, 2P: zwei Punkte.

Als Maß für die Reliabilität der Gesamtskala in der eindimensionalen Modellierung sowie der möglichen Teilskalen in den mehrdimensionalen Modellierungen werden die EAP-Reliabilitäten (3.3.1) betrachtet. Beim Gesamttest liegt diese (für die betrachtete Stichprobe von 399 Datensätzen) bei 0,60. In den mehrdimensionalen Modellen fallen diese mit Werten zwischen 0,43 und 0,59 (Kategorien) und zwischen 0,50 und 0,58 (kognitive Anforderungen) geringer aus (Tabelle 8-7).

Aus den verschiedenen IRT-Modellierungen ergeben sich auch jeweils verschiedene Schätzungen für Personenfähigkeiten und Aufgabenschwierigkeiten (hier: Schwierigkeiten der Aufgabenschwellen bzw. Threshold-Werte, siehe oben) auf der Logit-Skala. Um einen Eindruck zur Schwierigkeit der Testaufgaben bzw. Aufgabenschwellen gemäß den verschiedenen Strukturierungen zu erhalten, sind in Tabelle 8-7 die mittleren Schwierigkeiten (beide Schwellen zusammengenommen sowie einzeln betrachtet) für den Gesamttest sowie für die jeweiligen Aufgaben der einzelnen Kategorien und kognitiven Anforderungen aufgeführt. Niedrige Werte sprechen dabei für leichte Aufgaben bzw. leichter zu erreichende Aufgabenschwellen und höhere Werte umgekehrt für schwieriger zu erreichende Aufgabenschwellen.

So wird im Vergleich der mittleren Schwierigkeiten für die Teilskalen deutlich, dass das Erreichen der Punkteschwellen in den Aufgaben zu den kognitiven Anforderungen *Anwenden* und *Analysieren* vergleichsweise schwieriger ist als zur Anforderung *Reproduzieren*. Deren mittlere Schwierigkeit ist sowohl für die Ein-Punkt-Schwelle (1P-Schwelle), als auch für die Zwei-Punkte-Schwelle (2P-Schwelle) niedriger als die entsprechenden Werte für die beiden anderen kognitiven Anforderungen (Tabelle 8-7). Hinsichtlich der Kategorien ist das Erreichen der Punkteschwellen im Mittel für die Aufgaben zu den Kategorien *digitale Messwerterfassung* und zu *Simulationen* vergleichsweise leicht. Für die Aufgaben zur Kategorie *Erklärvideos* liegen die mittleren Schwierigkeiten der Punkteschwellen im Mittelfeld, während die zu den Aufgaben der Kategorie *fachbezogene Grundlagen* höher ausfallen (Tabelle 8-7).

Neben den Aufgabenschwierigkeiten werden in einer IRT-Modellierung zudem die Personenfähigkeiten geschätzt. Eine Wright Map (*person-item map*;

Boone et al., 2014) veranschaulicht die Verteilung der Aufgabenschwierigkeiten gegenüber der Personenfähigkeiten (Abbildung 8-5). Auf der rechten Seite der Grafik zeigt die vertikale Achse die Skala für Aufgabenschwierigkeiten und Personenfähigkeiten (logits). Auf der linken Seite wird die Häufigkeitsverteilung der Personenfähigkeiten auf dieser Skala verdeutlicht. Auf der horizontalen Achse sind die Bezeichnungen der Aufgabenschwellen dargestellt, zu denen die Datenpunkte zählen (z. B. A121 für den Threshold-Wert $t_{12,1}$ für das Erreichen von einem Punkt in Aufgabe 12). Personen in einem bestimmten Fähigkeitsbereich (d. h. in einem Häufigkeitsbalken) lösen eine Aufgabe (bzw. erreichen die 1P- oder 2P-Schwelle einer Aufgabe), deren Datenpunkt auf Höhe des Häufigkeitsbalkens liegt, mit einer Wahrscheinlichkeit von 50 %. Aufgabenschwellen, die darüber liegen, fallen diesen Personen schwerer und erreichen sie weniger wahrscheinlich; darunterliegende Aufgabenschwellen erreichen sie hingegen wahrscheinlicher (Boone et al., 2014).

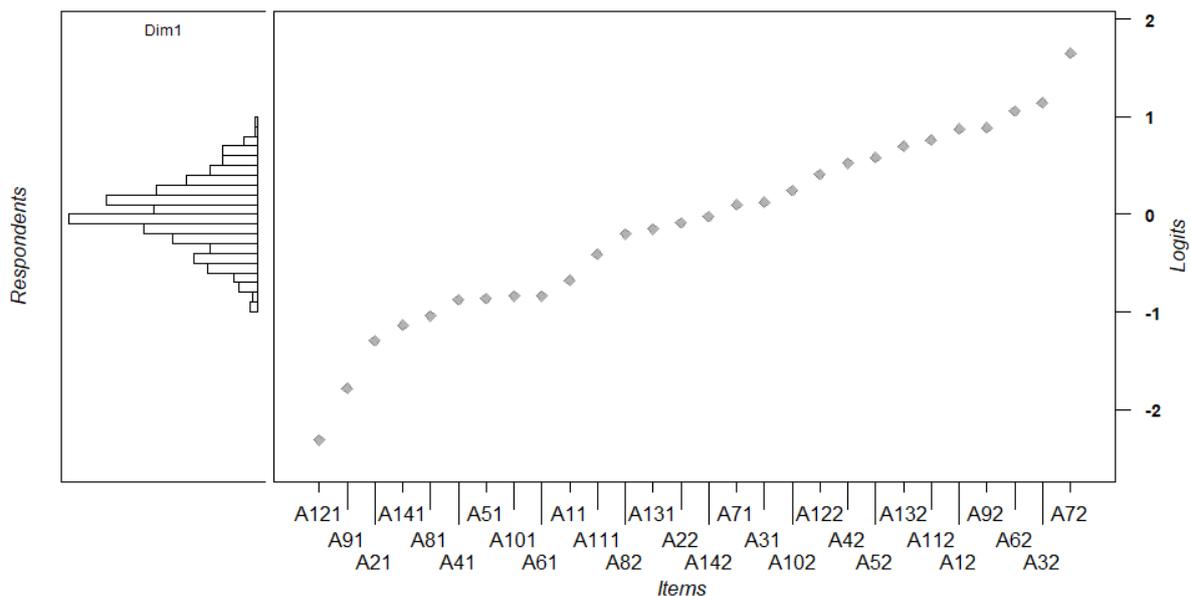


Abbildung 8-5: Wright Map (für die eindimensionale IRT-Modellierung).

Beschriftung der Aufgabenschwellen: A11 entspricht dem Threshold-Wert $t_{1,1}$ für das Erreichen von einem Punkt in Aufgabe 1 und A12 dem Wert $t_{1,2}$ für das Erreichen von zwei Punkten in Aufgabe 1 usw. Höhere Werte auf der Logit-Skala entsprechen schwieriger zu erreichenden Aufgabenschwellen sowie höheren Personenfähigkeiten.

Die Wright Map zur eindimensionalen IRT-Modellierung in Abbildung 8-5 zeigt für die betrachtete Stichprobe von 399 Lehramtsstudierenden der Phy-

sik, dass sich die Aufgabenschwierigkeiten für die einzelnen Punkteschwellen der Aufgaben angemessen über den Bereich der Personenfähigkeiten verteilen. Der mittlere Schwierigkeitsbereich ist gut besetzt und es liegen nur wenige Werte der Aufgabenschwellen über oder unter dem Bereich der Personenfähigkeiten. Dies spricht dafür, dass die Schwierigkeit des Gesamttests im statistischen Sinne annehmbar war für die untersuchte Stichprobe. Dies unterstützt auch die geringe Abweichung von 0,12 logits zwischen der mittleren Schwierigkeit aller Aufgabenschwellen und mittlerer Personenfähigkeit (Boone et al., 2014).

Mithilfe der WrightMap lässt sich zudem verdeutlichen, dass aufgrund der Mehrstufigkeit der vorliegenden Aufgaben keine eindeutige Angabe möglich ist, wie schwierig eine Aufgabe im Vergleich zu einer anderen ist. Zwar lassen sich die Mittelwerte der beiden Schwellenwerte $t_{j,k}$ für eine Aufgabe berechnen, um nur einen Schwierigkeitswert je Aufgabe zu erhalten, dieser kann jedoch nicht eindeutig interpretiert werden (Strobl, 2015). So ergibt sich beispielsweise für die beiden Schwellenwerte zu Aufgabe 12 ein Mittelwert von -0,95 und für Aufgabe 8 ein Mittelwert von -0,62. Betrachtet man jedoch die Wright Map in Abbildung 8-5, so zeigt sich, dass das Erreichen der 1P-Schwelle in Aufgabe 12 zwar leichter ist als bei Aufgabe 8, umkehrt jedoch das Erreichen der vollen Punktzahl (2P-Schwelle) in Aufgabe 8 leichter ist als in Aufgabe 12. Es lässt sich demnach nicht sagen, welche der beiden Aufgaben im Gesamten leichter oder schwieriger ist. Bezogen auf den Gesamttest ist Aufgabe 12 die Testaufgabe, in der am leichtesten die 1P-Schwelle erreicht wird (Datenpunkt A121), Aufgabe 8 hingegen diejenige, in welcher am leichtesten die volle Punktzahl erreicht wird (Datenpunkt A82). In Aufgabe 3 ist es am schwierigsten, die 1P-Schwelle zu erreichen (Datenpunkt A31), während es in Aufgabe 7 am schwierigsten ist, zwei Punkte zu erhalten (Datenpunkt A72; Abbildung 8-5). Die Aufgaben 3, 7 und 8 werden exemplarisch als Aufgabenbeispiele des finalen Testinstruments in Kapitel 10 vorgestellt.

Mithilfe einer Farbkodierung in der Wright Map lässt sich zudem darstellen, wie die Aufgabenschwierigkeiten bzw. Schwellenwerte zu den Aufgaben sich je nach Kategorie oder kognitiver Anforderung verteilen. Abbildung 8-6 zeigt die Wright Map erneut, jedoch sind die Datenpunkte einmal nach Kategorie und einmal nach kognitiver Anforderung farbkodiert.

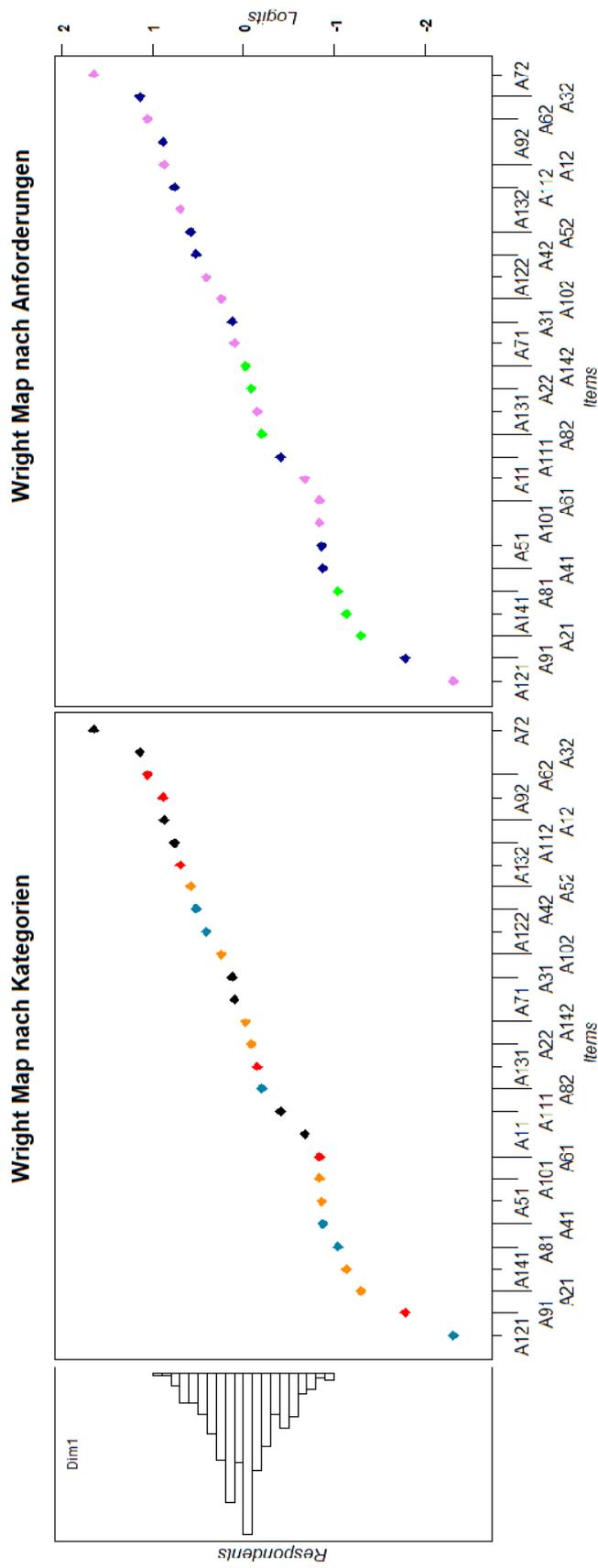


Abbildung 8-6: Wright Maps mit Farbkodierung nach Kategorien und kognitiven Anforderungen (links: Farbkodierung nach Kategorien, schwarz: fachbezogene Grundlagen, orange: digitale Messwerterfassung, hellblau: Simulationen, rot: Erklärvideos; rechts: Farbkodierung nach kognitiven Anforderungen, grün: Reproduzieren, dunkelblau: Anwenden, rosa: Analysieren)

Passend zu den dargestellten mittleren Schwierigkeiten in Tabelle 8-7 wird hier im Hinblick auf die kognitiven Anforderungen (Abbildung 8-6; rechte Wright Map) eine Tendenz sichtbar, dass die Anforderung *Reproduzieren* (grün) vermehrt im leichten Schwierigkeitsbereich auftritt, während die Anforderungen *Analysieren* (rosa) und *Anwenden* (dunkelblau) sich in der Breite verteilen und damit den schwierigeren Bereich ausmachen.

Die Kategorien (Abbildung 8-6; linke Wright Map) sind insgesamt über das Schwierigkeitsspektrum verteilt. Auch hier deckt sich das Bild mit den mittleren Schwierigkeiten je Kategorie (Tabelle 8-7): in den Kategorien *Simulationen* (hellblau) und *digitale Messwerterfassung* (orange) gibt es häufiger leichter zu erreichende Schwellen, während in der Kategorie *fachbezogene Grundlagen* (schwarz) überwiegend schwieriger zu erreichende Schwellen auftreten (Abbildung 8-6). Die Bezeichnung „Grundlagen“ ist für den Test demnach nicht so zu verstehen, dass in dieser Kategorie leichtere Aufgaben enthalten sind. Stattdessen sind die Aufgaben vergleichsweise schwierig, da die Kategorie kein reproduziertes Wissen abfragt, sondern die Anwendung des FDW zum Einsatz digitaler Medien für eine bestimmte Situation im Physikunterricht oder die Analyse und ggfs. Bewertung einer solchen Situation erfordert (8.2).

Insgesamt zeigen sich für die Testaufgaben zufriedenstellende Kennwerte hinsichtlich der Schwierigkeiten für die einzelnen Aufgabenschwellen. Für den Gesamttest liegt eine angemessene Verteilung auf den Fähigkeitsbereich der befragten Personen vor. Auch im Hinblick auf die Kategorien und kognitiven Anforderungen liegt eine Verteilung der Schwierigkeiten in eine gewisse Breite vor, da die einzelnen Kategorien und kognitiven Anforderungen jeweils Aufgabenschwellen unterschiedlicher Schwierigkeiten beinhalten. Die auftretenden Tendenzen, dass die kognitive Anforderung *Reproduzieren* eher leichtere Aufgaben und umgekehrt die Kategorie *fachbezogene Grundlagen* eher schwierigere Aufgaben enthält, sind zudem eingängig.

Für alle drei vorgestellten IRT-Modellierungen liegt eine annehmbare Modellpassung vor (Tabelle 8-7). Im Rahmen einer Untersuchung zur strukturellen Validität werden in 9.5 die eindimensionale und die vierdimensionale Model-

lierung nach Kategorien aufgegriffen und miteinander verglichen, um die angenommene Struktur des modellierten FDW zum Einsatz digitaler Medien zu prüfen.

Hinsichtlich der EAP-Reliabilität zeigt sich für die Gesamtskala mit 0,60 ein eher geringer Wert im Vergleich zu anderen Instrumenten zum FDW in Physik; mit nochmals niedrigeren Werten in den möglichen Teilskalen (insbesondere für die Kategorie *fachbezogene Grundlagen* mit 0,43). So wurde für den entwickelten FDW-Test im Projekt KiL eine Reliabilität von $\alpha = 0,74$ (Cronbachs α) berichtet (3.4; Kröger, 2019). Schiering et al. (2019) nutzten in KiL für ihre Analyse ebenfalls eine IRT-Modellierung und berichteten eine WLE-Reliabilität von 0,73¹⁶. Im Projekt ProfiLe-P ergab sich für den dort entwickelten FDW-Test eine EAP-Reliabilität von 0,72 (Gramzow, 2015). Die von Kotzebue (2022b) entwickelten Leistungstestaufgaben zum TPACK im Fach Biologie wiesen eine ähnliche Reliabilität von $\alpha = 0,59$ auf (3.5.2) wie die Aufgaben im vorliegenden Test. Wie in 3.3.1 bereits angeführt, kann es bei Leistungstests in der Fachdidaktik vorkommen, dass diese geringere Reliabilitäten als 0,7 aufweisen, da z. T. inhaltlich komplexere und schwerer zu operationalisierende Konstrukte erfasst werden, als es typischerweise in der Psychologie der Fall ist (Schecker, 2014). Nach Schmitt (1996) kann auch ein Test mit geringer Reliabilität (von z. B. 0,49) seine Berechtigung haben, wenn damit beispielweise ein Gebiet inhaltlich aussagekräftig abgedeckt wird oder die Messung eine passende Eindimensionalität aufweist. In diesem Sinne wird auch an dieser Stelle argumentiert, dass das Testinstrument trotz der eher geringen EAP-Reliabilität nutzbar ist, um das komplexe Konstrukt des FDW zum Einsatz digitaler Medien zu erfassen. Für die Nutzung des Tests in der Haupterhebung dieser Arbeit (11.2) muss jedoch zusätzlich die EAP-Reliabilität für die dort befragte Stichprobe betrachtet werden.

Im nachfolgenden Kapitel werden nun die Ergebnisse der Teilstudien zur Untersuchung der Konstruktvalidität des entwickelten Testinstruments vorgestellt. Aus manchen dieser Teilstudien lassen sich zudem Indizien für eine angemessene inhaltliche Erfassung des Wissens ableiten, was wiederum die

¹⁶ Die WLE-Reliabilität ist ein weiteres Reliabilitätsmaß in der IRT, welche auf der WLE-Schätzung (*Weighted Likelihood Estimation*) nach Warm (1989) basiert.

Berechtigung der Testnutzung auch bei geringeren Reliabilitätswerten unterstützt.

9 Ergebnisse zur Untersuchung der Validität des Testinstruments

In diesem Kapitel werden die verschiedenen Studien zur Untersuchung der Validität des entwickelten Testinstruments sowie deren Ergebnisse erläutert. Diese Studien dienen der Beantwortung der ersten Forschungsfrage (FF1) in dieser Arbeit (6.2). Dabei wurden verschiedene Aspekte der Konstruktvalidität (3.3.2) in den Blick genommen, um Hinweise für eine angemessene Interpretation der mit dem Test erfassten Daten zu erhalten, welche auf eine Erklärung und Bewertung¹⁷ von Leistungen abzielt (7.1). Diese Hinweise beziehen sich darauf, inwiefern das mit dem Test untersuchte Konstrukt des FDW zum Einsatz digitaler Medien ein Wissen darstellt, das (1) universitär erwerbbar ist, (2) fachspezifisch ist und sich von anderen Wissensbereichen abgrenzt sowie (3) eine angenommene Strukturierung aufweist.

Die Auswahl und Begründung der betrachteten Validitätsaspekte wurde in 7.1 beschrieben. So widmen sich die nachfolgenden Abschnitte jeweils den verschiedenen Validierungsstudien im Hinblick auf die Abgrenzung zu verwandten Wissensbereichen (9.1) sowie den diskriminanten (9.2), kognitiven (9.3), inhaltlich-curricularen (9.4) und strukturellen (9.5) Aspekten der Konstruktvalidität. In 9.6 werden die Ergebnisse der Studien zusammengefasst.

9.1 Abgrenzung von fachunabhängigem pädagogischem Wissen durch Korrelationsanalysen

Um das fokussierte FDW zum Einsatz digitaler Medien von anderen Wissensbereichen abzugrenzen, wurden im Rahmen der Pilotierung in einem Teil der Stichprobe zusätzliche Testinstrumente zum FDW (ohne Bezug zu digitalen Medien) und zum Pädagogischen Wissen (PW) eingesetzt.

¹⁷ Hier sei darauf hingewiesen, dass sich die „Bewertung von Leistungen“ an dieser Stelle auf die Leistung im Testinstrument zu Untersuchungszwecken bezieht. Es ist keinesfalls so zu verstehen, dass die Studierenden im Sinne einer Studienleistung bewertet werden. Die Teilnahme an den Befragungen im Rahmen dieser Arbeit hatte für die Studierenden keinerlei Einfluss auf ihre Leistungen im Studium.

Konkret wurde eine geschlossene Testversion des FDW-Tests in Physik (Gramzow, 2015) aus dem Projekt ProfiLe-P Transfer (dritte Förderphase im Verbundprojekt ProfiLe-P) genutzt (Jordans et al., 2022). Diese umfasst 24 Mehrfachwahlaufgaben in den vier fachdidaktischen Facetten *Instruktionsstrategien*, *Schülervorstellungen*, *Experimente* und *Vermittlung eines angemessenen Wissenschaftsverständnisses* und *Fachdidaktische Konzepte*. Weiterhin wurde eine adaptierte Kurzsкала eines etablierten Testinstruments zum PW nach Seifert und Schaper (2012) bzw. Seifert et al. (2009) verwendet. Die Kurzsкала (11 Testaufgaben, offenes und geschlossenes Format) entwickelte Riese (2009) auf Basis dieses etablierten Tests zum PW. Sie wurde zudem auch in den Projekten ProfiLe-P und ProfiLe-P+ eingesetzt. Die beiden nachfolgenden Abbildungen zeigen jeweils eine Beispielaufgabe des zusätzlichen eingesetzten FDW-Tests (Abbildung 9-1) und der Kurzsкала zum PW-Test (Abbildung 9-2).

1. Fachdidaktisches Wissen – Teil 4 von 24

Handelt es sich im Folgenden um typische Ursachen für fachlich nicht korrekte Schülervorstellungen?

| | Nein | Ja |
|--|-----------------------|-----------------------|
| Doppeldeutigkeit von Begriffen in Fach- und Alltagssprache | <input type="radio"/> | <input type="radio"/> |
| Physikalisch vereinfachte oder falsche Darstellungen in den Massenmedien | <input type="radio"/> | <input type="radio"/> |
| Eigene spontane Erklärungen ohne tiefere fachliche Überlegungen | <input type="radio"/> | <input type="radio"/> |
| Vereinfachende Vorstellung aufgrund begrenzter Rechenfertigkeiten | <input type="radio"/> | <input type="radio"/> |
| Oberflächliche Kenntnisse aus vorausgehendem Unterricht | <input type="radio"/> | <input type="radio"/> |

Abbildung 9-1: Beispielaufgabe aus dem eingesetzten Test zum FDW in Physik (Jordans et al., 2022) (Facette: Fachdidaktische Konzepte; Musterlösung: ja, ja, nein, nein, ja)

1. Pädagogisches Wissen – Teil 5 von 11

Stellen Sie sich bitte Folgendes vor:

Sie führen als Lehrperson ein Elterngespräch mit einem Vater über das Verhalten seines 15-jährigen Sohnes im Unterricht. Im Laufe des Gesprächs sagt der Vater: „Mein Sohn ist ein kluges Kind!“

Bitte kreuzen Sie an, wie diese Aussage gemäß des Vier-Ohren-Modells nach Schulz von Thun zu verstehen ist!

- Auf der Sachebene bedeutet diese Aussage, dass der Vater mir mitteilen möchte, dass sein Sohn klug sei.
- Auf der Sachebene bedeutet diese Aussage, dass der Vater mich anhält, die Klugheit seines Sohnes zu erkennen.
- Auf der Komplementärebene bedeutet diese Aussage, dass ich als Lehrperson in dieser Gesprächssituation eine andere Rolle einnehme als der Vater.
- Auf der Beziehungsebene bedeutet diese Aussage, dass der Vater mir nicht zutraut, erkennen zu können, dass sein Sohn klug ist, weswegen er zu mir sagt: „Mein Sohn ist ein kluges Kind“.
- Auf der Selbstoffenbarungsebene bedeutet diese Aussage, dass der Vater seinen Sohn nicht als Jugendlichen, sondern als Kind wahrnimmt.

Abbildung 9-2: Beispielaufgabe aus dem eingesetzten Test zum PW (Seifert & Schaper, 2012) (richtige Antworten: erste, vierte und fünfte Option)

Die beiden Instrumente wurden jeweils bei einem Teil der Stichprobe in der Pilotierung des entwickelten Tests zum FDW zum Einsatz digitaler Medien eingesetzt. Von den $N_0 = 116$ in der Pilotierung befragten Lehramtsstudierenden (Tabelle 8-4) beantworteten 105 zusätzlich den PW-Test und wiederum davon 43 ebenfalls den FDW-Test.

Um zu prüfen, dass das entwickelte Testinstrument einen bestimmten Aspekt des FDW in Physik erfasst und nicht fachunabhängiges PW (zu digitalen Medien), wurden Korrelationsanalysen zwischen den erreichten Punkten (Summenscore) im entwickelten Testinstrument zum FDW zum Einsatz digitaler Medien mit den Scores im PW-Test sowie im FDW-Test vorgenommen. Gemäß dieser Zielstellung wurde erwartet, dass die Korrelation mit dem FDW-Test höher ausfällt als die Korrelation mit dem PW-Test.

Tabelle 9-1 stellt die entsprechenden Korrelationen dar. Alle bivariaten Korrelationen sind signifikant. Wie erwartet zeigt sich jedoch eine höhere Korrelation des Testscores im FDW zum Einsatz digitaler Medien mit dem Score im FDW als mit dem Score im PW.

Tabelle 9-1: Manifeste Korrelationen des FDW zum Einsatz digitaler Medien im Fach Physik mit den Leistungstests zum FDW im Fach Physik und PW (Pearson-Korrelation)

| | FDW-digitale Medien ($\alpha = 0,69$) | allgemeines FDW ohne digitale Medien |
|---|---|---|
| allgemeines FDW ohne digitale Medien ($\alpha = 0,67$) | 0,46 ** (N = 43) | – |
| PW ($\alpha = 0,74$) | 0,32 ** (N = 105) | 0,34 * (N = 41) |

Signifikanzniveaus * $p < 0,05$, ** $p < 0,01$. Cronbachs α als klassisches Reliabilitätsmaß bezieht sich jeweils auf die mit den Instrumenten befragte Stichprobe im Rahmen der Pilotierung.

Das erwartungskonforme Ergebnis dieser Korrelationsanalyse gibt somit einen ersten Hinweis darauf, dass das mit dem entwickelten Test gemessene Wissen weniger mit (fachunspezifischem) PW zusammenhängt und stattdessen stärker mit dem FDW im Fach Physik. Dieser Hinweis stützt somit die Validität des entwickelten Tests im Hinblick auf die Frage, inwiefern das damit gemessene Wissen fachspezifisch ist und sich von anderen Wissensbereichen wie dem PW abgrenzen lässt.

9.2 Diskriminante Validierung durch Befragung von Lehramtsstudierenden anderer Fächer

Ebenfalls im Rahmen der Pilotierung wurden neben Physiklehramtsstudierenden zusätzlich Lehramtsstudierende anderer Fächer mit dem entwickelten Testinstrument befragt, um zu ermitteln, wie diese im Vergleich zu den Physiklehramtsstudierenden im Test abschneiden.

So stellten 92 der 116 Befragten in der Pilotierung Studierende im Lehramt Physik dar (Tabelle 8-4). Die restlichen 24 waren Lehramtsstudierende in den Fächern Chemie oder Technik (ohne Kombinationsfach Physik). Im Sinne einer diskriminanten Validierung war die Vermutung für diese Untersuchung, dass die Studierenden im Lehramt Physik bessere Testergebnisse erzielen als die des Lehramts Chemie oder Technik, weil der Test physikdidaktisches Wissen adressiert. In Tabelle 9-2 sind die deskriptiven Testergebnisse für die beiden Gruppen dargestellt.

Tabelle 9-2: Deskriptive Statistik der befragten Lehramtsstudierenden (Physik oder andere Fächer bzw. kein Physik) zur Testleistung im FDW zum Einsatz digitaler Medien

| Lehramtsstudierende | M | SD | Minimum | Maximum |
|----------------------------|----------|-----------|----------------|----------------|
| Physik (N = 92) | 12,74 | 4,11 | 3 | 20 |
| andere Fächer (N = 24) | 10,04 | 3,20 | 5 | 16 |

Insgesamt konnten 24 Punkte erreicht werden (unter Berücksichtigung der Überschneidung gleicher Aufgaben in Pilotierung und finalem Testinstrument); M: Mittelwert des Summenscores; SD: Standardabweichung.

Die Tabelle zeigt erwartungskonform, dass die Physiklehramtsstudierenden im Mittel einen höheren Testscore erreichen als die befragten Studierenden der Fächer Chemie und Technik (ohne Physik). Ein Vergleich der mittleren Testscores beider Gruppen mithilfe eines t-Tests unabhängiger Stichproben (Bortz & Schuster, 2010) zeigt einen signifikanten Gruppenunterschied ($t(114) = 2,986$; $p = 0,001$; $d = 0,68$) mit mittlerer Effektstärke (Cohen, 1988). Dieses Ergebnis spricht zunächst für die diskriminante Validität des Tests im Sinne einer Messung fachspezifischen Wissens.

Die befragten Gruppen der Lehramtsstudierenden in Physik gegenüber denen in Chemie oder Technik (ohne Physik) waren jedoch im Hinblick auf ihre Personenmerkmale nicht ganz vergleichbar und unterschieden sich signifikant in ihrer Abiturnote mit großem Effekt ($t(92) = -4,423$; $p < 0,001$; $d = -1,08$). Da die Abiturnote als möglicher Indikator für kognitive Leistungsfähigkeit (Riese et al., 2017; Riese & Reinhold, 2012) wiederum mit der Testleistung im gemessenen FDW zusammenhing (Pearson-Korrelation von $r = -0,43$; $p < 0,001$), wurde betrachtet, inwiefern der Gruppenunterschied zwischen Physik- und nicht-Physik-Lehramtsstudierenden sich auch unter Kontrolle der Abiturnote zeigt. Eine lineare Regression mit den entsprechenden Einflussvariablen auf die abhängige Variable FDW zum Einsatz digitaler Medien zeigte, dass der Einfluss des Fachs (Dummy-Variable: Physik = 1, nicht Physik = 0) nicht mehr signifikant ist, wenn die Abiturnote in das Regressionsmodell aufgenommen wird (Tabelle 9-3).

Tabelle 9-3: Ergebnisse der Regressionsanalyse zur Testleistung (Summenscore) im FDW zum Einsatz digitaler Medien (N = 94; $R^2 = 0,193$; korr. $R^2 = 0,175$; $F(2,91) = 10,881$; $p < 0,001$)

| Koeffizienten | b | SE | β | T | p |
|-------------------------------------|--------|-------|---------|--------|--------|
| Konstante | 16,912 | 1,962 | | 8,619 | <0,001 |
| Abiturnote | -2,676 | 0,713 | -0,389 | -3,753 | <0,001 |
| Dummy Fach (Physik/ kein Physik) | 0,913 | 0,967 | 0,098 | 0,944 | 0,348 |

b: nicht standardisierter Regressionskoeffizient; SE: Standardfehler; β : β -Wert (standardisierter Koeffizient).

Der beobachtete Gruppenunterschied im FDW zum Einsatz digitaler Medien lässt sich demnach scheinbar nicht durch die Fachzugehörigkeit erklären, sondern durch die Unterschiede in der Abiturnote der Befragten (Tabelle 9-3). Hier sei einschränkend angemerkt, dass die Abiturnote nicht von allen 116 Befragten erhoben werden konnte, sondern dazu nur 94 gültige Fälle vorliegen. Nichtsdestotrotz ist der Einfluss der Abiturnote im Regressionsmodell deutlich.

Dieses Ergebnis schränkt die oben dargestellte Beobachtung des Gruppenunterschieds im gemessenen FDW (t-Test) als Hinweis für die diskriminante Validität ein, da der Effekt der Fachzugehörigkeit unter Einbezug der Abiturnote als Einflussvariable verschwindet. Durch den positiven β -Wert für die Variable zur Fachzugehörigkeit deutet sich jedoch der Effekt in erwarteter Richtung an. Bei einer Erhebung in einer größeren Stichprobe könnte sich möglicherweise auch ein signifikanter Effekt zeigen. Auf Basis der vorliegenden Daten lässt sich jedoch aus der Regression kein Hinweis für die Fachspezifität des gemessenen FDW zum Einsatz digitaler Medien ableiten. Um den möglichen Effekt der Fachzugehörigkeit aufklären zu können, wäre eine Erhebung in einer größeren Stichprobe nützlich.

Zudem könnte eine mögliche Erklärung für das Ergebnis sein, dass die betrachteten Fächer Chemie und Technik vergleichsweise nah am Fach Physik liegen und das Testinstrument eher MINT-spezifisches Wissen als physikspezifisches Wissen adressiert. Für eine Aufklärung dahingehend wären ebenso weitere Erhebungen notwendig, beispielsweise auch unter Einbezug

von Lehramtsstudierenden entfernterer Fächer wie z. B. aus dem Bereich der Sprachen oder Gesellschaftswissenschaften.

9.3 Kognitive Validierung durch Think-Aloud-Studien

Zur Untersuchung, inwiefern von den Teilnehmenden bei der Bearbeitung der Testaufgaben auch der intendierte Wissensbereich aktiviert wird, wurden Physiklehramtsstudierende mit der Methode des Lauten Denkens bzw. *Think-Aloud* (3.3.2) zum entwickelten Testinstrument befragt. Neben dieser kognitiven Validierung war ein weiteres Ziel dieser Think-Aloud-Studien, die Verständlichkeit und Angemessenheit des Tests für die Zielgruppe zu prüfen und Hinweise zu möglichen Schwierigkeiten oder missverständlichen Formulierungen in den Aufgaben zu gewinnen.

So wurden im Rahmen einer Masterarbeit an der RWTH Aachen vier Think-Aloud-Interviews mit Studierenden im Lehramt Physik geführt (Conrads, 2022). Die Teilnehmenden der Interviews erhielten zu Beginn Informationen und Hinweise zur Think-Aloud-Studie (Anhang D). Anknüpfend bearbeiteten sie die Testaufgaben, wobei sie ihre Antworten und insbesondere alle ihre Gedankengänge, Handlungen, Absichten und Probleme bei der Bearbeitung laut aussprechen sollten. Die Testleitung war während dieser Bearbeitungszeit durchgängig anwesend; jedoch nur zur Beobachtung und nicht als Ansprechperson für inhaltliche Fragen. Im Anschluss an die Testbearbeitung folgte eine kurze Interviewphase, in denen die Teilnehmenden zum Testinstrument befragt wurden. Dabei ging es um die eigene Beschreibung des Tests, die Bewertung der sprachlichen und inhaltlichen Schwierigkeit des Tests sowie des eigenen Gelingens. Zudem gab es die Möglichkeit, anderweitige Anmerkungen zum Test oder zur Studie zu machen. Diese Interviewfragen zur Think-Aloud-Erhebung sind in Anhang D hinterlegt. Die Think-Aloud-Interviews mitsamt Testbearbeitung und anschließender Befragung wurden per Audio aufgezeichnet und im Rahmen der Masterarbeit von Conrads (2022) ausgewertet.

Für diese Auswertung wurden die Ausführungen der Teilnehmenden bei der Testbearbeitung dahingehend kodiert, welche Wissensart für die Beantwortung der einzelnen Antwortalternativen herangezogen wurde. In Anlehnung an die Validierungsstudie von Gramzow (2015) zum FDW-Test im Projekt

Profile-P wurden dabei vier Kategorien zur Einteilung vorgesehen: (1) Physikdidaktisches Wissen, (2) Fachliches Wissen, (3) Erfahrungen aus der Schulzeit oder eigener Lehrtätigkeit und (4) Subjektive Theorien (Conrads, 2022). Tabelle 9-4 stellt die die Anwendung der Kategorien in der Masterarbeit dar.

Tabelle 9-4: Beschreibung und Anwendung zu verwendeten Wissensarten bei der Testbearbeitung im Rahmen einer Masterarbeit zur kognitiven Validierung (Conrads, 2022)

| Kategorie | Beschreibung | Beispiele zur Anwendung |
|---|--|--|
| Physikdidaktisches Wissen | Begründung durch explizite oder implizite Verwendung von Wissen oder Nicht-Wissen aus fachdidaktischen Lehrveranstaltungen | <ul style="list-style-type: none"> • konkreter Bezug zu fachdidaktischer Lehrveranstaltung • fachdidaktische Aspekte und Potentiale zum Einsatz digitaler Medien |
| Fachliches Wissen | Begründung durch explizite fachliche Einschätzung | <ul style="list-style-type: none"> • Bewertung eines Diagramms aus fachlicher Sicht |
| Erfahrungen (aus der Schulzeit; als Lehrperson) | Beantwortung aufgrund persönlicher Erfahrungen aus der eigenen Schulzeit oder aus Praktika oder anderen Lehrerfahrungen | <ul style="list-style-type: none"> • konkreter oder indirekter Bezug zu eigener Schulzeit • konkreter oder indirekter Bezug zu eigener Lehrtätigkeit |
| Subjektive Theorien | keine begründete Argumentation passend zu den vorausgehenden Kategorien | <ul style="list-style-type: none"> • unbegründete Aussagen wie „das würde ich nicht so sehen“ |

Bei Aussagen der Kategorie Subjektive Theorien ist nicht eindeutig, was die Ursache für die Begründung oder die geäußerte subjektive Theorie ist oder ob dort tatsächlich eine subjektive Meinung geäußert oder nicht vielmehr fachdidaktisches (oder fachliches) Wissen schlecht expliziert wird (Gramzow, 2015). Somit wird diese Kategorie auch verwendet, wenn Aussagen nicht klar zu den anderen Kategorien zuzuordnen sind.

Abbildung 9-3 zeigt die Verteilung der vergebenen Kategorien in der Masterarbeit von Conrads (2022) für die einzelnen Think-Aloud-Interviews. Dabei sind einerseits die relativen Häufigkeiten für die Kategorien für alle Testpersonen (TP) einzeln (ausgefüllte Balken) und andererseits die relative Häufigkeit insgesamt (schraffierter Balken) dargestellt.

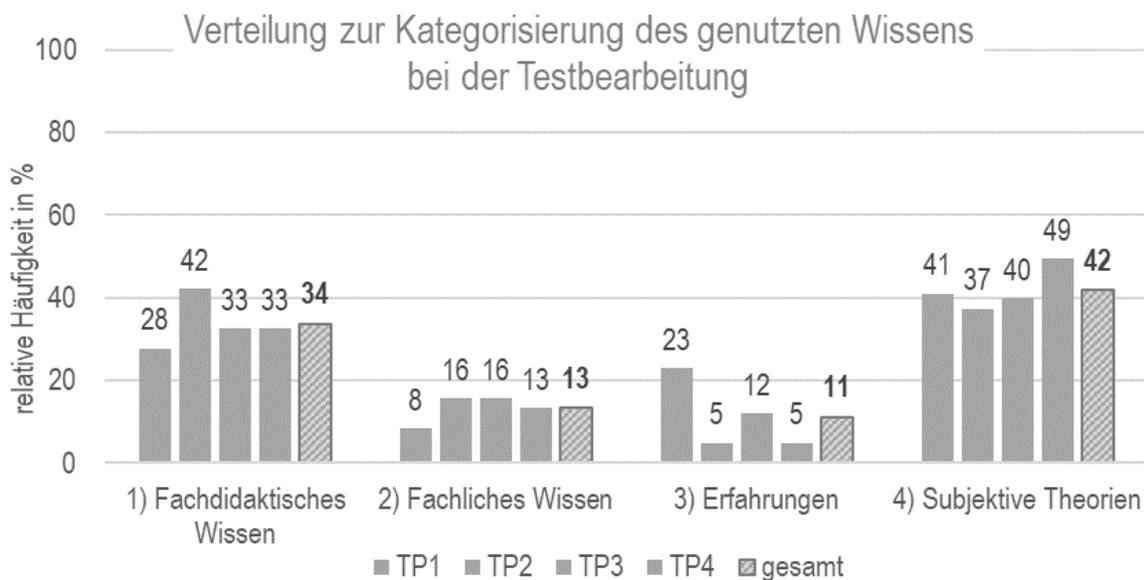


Abbildung 9-3: Verteilung zur Kategorisierung des bei der Testbearbeitung genutzten Wissens im Rahmen von Think-Aloud-Studien (Conrads, 2022)

Es zeigt sich, dass die Teilnehmenden neben subjektiven Theorien oder Äußerungen, die keine klare Zuordnung zuließen, häufig FDW bei der Beantwortung der Aufgaben heranzogen. Dieser Anteil zum FDW lag für die vier Testpersonen zwischen 28 % und 42 % und war damit stets größer als die Anteile zum Fachlichen Wissen oder den Erfahrungen. Für TP 2 stellte das FDW sogar den größten Anteil dar, für die anderen Testpersonen wurde häufiger die Kategorie Subjektive Theorien kodiert. TP 1 zeigte im Vergleich zu den anderen Teilnehmenden eine höhere Nutzung von Erfahrungen aus eigener Schulzeit oder Lehrtätigkeit, welche einen annähernd hohen Anteil wie den des FDW darstellt (Abbildung 9-3).

Insgesamt zeigt die Analyse, dass die Befragten sich bei der Bearbeitung der Testaufgaben vermehrt auf subjektive Einschätzungen berufen und dass trotz der Methode des Lauten Denkens nicht immer eindeutige Wissensres-

sources für die Beantwortung identifiziert werden können. Bei den Aussagen, für die eine Zuordnung möglich war, zeigt sich jedoch eine klare Dominanz des FDW als Wissensart oder Ressource, die bei der Beantwortung der Testaufgaben genutzt wurde. Dies lässt sich als Hinweis für die Validität des entwickelten Testinstruments im Hinblick auf die Abgrenzung des gemessenen FDW zu anderen Wissensarten wie dem Fachwissen oder rein auf Schulerfahrungen basierendem Wissen ableiten.

Weiterhin wurde zur Testbearbeitung der Teilnehmenden untersucht, welche Aufgaben oder Antwortalternativen Verständnisschwierigkeiten oder Probleme aufzeigten. Hier zeigten sich zwar manche Stellen, in denen einzelne Testpersonen aufgrund eines Ausdrucks, einer Formulierung oder einer Darstellung Schwierigkeiten äußerten, jedoch gab es keine systematischen Problemstellen, die bei allen oder fast allen Teilnehmenden deutlich wurden. Aus der Bepunktung der beantworteten Aufgaben ergaben sich für die vier Teilnehmenden Summenscores von 17, 22, 14 und 15 von 30 möglichen Punkten (57 %, 73 %, 47 % und 50 %). Beispielsweise erreichten sie die wenigsten Punkte in Aufgabe 7 (die finale Version der Aufgabe wird in Abbildung 10-2 dargestellt). Die Erläuterungen der Teilnehmenden bei der Beantwortung dieser Aufgabe gaben Hinweise auf eine mögliche Uneindeutigkeit der Antwortalternativen, welche ein Grund für die niedrigen Punktzahlen in der Aufgabe sein könnte.

In den an die Testbearbeitung anknüpfenden kurzen Interviews äußerten die Teilnehmenden ihre Einschätzungen zum Testinstrument gemäß den oben genannten Aspekten. Die sprachliche Schwierigkeit bewerteten sie trotz einiger Fachbegriffe und zum Teil anspruchsvoller Sätze als insgesamt angemessen für eine Bearbeitung durch Studierende mit Deutsch als Muttersprache oder einem ähnlich hohem Sprachniveau. Die Teilnehmenden schätzten die inhaltliche Schwierigkeit als mittelmäßig ein und gaben an, dass für sie fachdidaktische Aspekte ausschlaggebender waren als fachliche. Gleichzeitig wurde jedoch angemerkt, dass die Umsetzung mancher beschriebener Medieneinsätze nicht ganz klar wurde und dass die Teilnehmenden teilweise persönliche Meinungen über konkrete digitale Medien miteinbezogen. Diese

Einschätzung ist wiederum passend zu den vereinzelt geäußerten Schwierigkeiten bzgl. der Aufgaben sowie der hohen Anteile der Kategorie Subjektive Theorien.

Aufgrund dieser insgesamt positiven Hinweise zur Verständlichkeit und Angemessenheit des Testinstruments sowie zur kognitiven Validität wurden die Testaufgaben nach den Think-Aloud-Studien zunächst nicht überarbeitet. Zwar könnte eine präzisere Ausformulierung aller Medieneinsätze in dem Test die vereinzelt aufgetretenen Probleme womöglich beheben, jedoch sollte auch der Leseaufwand und die Testlänge in einem angemessenen Rahmen bleiben, weswegen auf eine dahingehende Überarbeitung verzichtet wurde. Die durch die Masterarbeit von Conrads (2022) identifizierten vereinzelt Problemstellen (wie z. B. in Aufgabe 7) wurden jedoch für den weiteren Validierungsprozess im Blick behalten. So wurde Aufgabe 7 letztendlich auch aufgrund von Anmerkungen im Rahmen der Expert:innenbefragung zur curricularen Validierung überarbeitet (Tabelle 9-5), wobei die Hinweise aus den Think-Aloud-Interviews einbezogen wurden.

9.4 Inhaltlich curriculare Validierung durch schriftliche Expert:innenbefragungen

Zur Untersuchung, ob die entwickelten Testaufgaben eine curriculare Passung aufweisen und somit mit dem Test auch universitär erwerbbares FDW zum Einsatz digitaler Medien erfasst wird, erfolgte eine schriftliche Expert:innenbefragung mit Physikdidaktiker:innen. So wurden Personen von insgesamt zehn Hochschulen in Deutschland und Österreich befragt, die in der fachdidaktischen Lehrkräftebildung tätig waren. Neben den Projektbeteiligten an den damaligen Projektstandorten Aachen, Graz und Tübingen nahmen zudem Personen von Universitäten bzw. Hochschulen der Standorte Dresden, Erlangen, Innsbruck, Münster, Paderborn, Schwäbisch Gmünd und Würzburg teil.

In der Befragung erhielten die Teilnehmenden die entwickelten Testaufgaben (Testversion nach der Pilotierung aus 15 Mehrfachwahlaufgaben) und sollten für jede Aufgabe eine Einschätzung geben, inwiefern das notwendige Wissen zum Lösen der Aufgabe am eigenen Standort in der fachdidaktischen Ausbil-

dung erworben werden kann. Diese Einschätzung wurde auf einer sechsstufigen Skala eingefordert. Zudem hatten die Befragten optional die Möglichkeit, bei jeder Aufgabe in einem Freitextfeld ihre Einschätzung zu begründen oder eine anderweitige offene Bemerkung zur Aufgabe abzugeben. Die Befragung erfolgte mithilfe eines interaktiven PDF-Dokuments (zum Ankreuzen und Ausfüllen), welches per E-Mail an die teilnehmenden Physikdidaktiker:innen geschickt wurde. Abbildung 9-4 zeigt beispielhaft eine Seite des Dokuments, in der die Einschätzung und optionale Bemerkung zu Aufgabe 7 abgefragt werden.

Aufgabe 7

Im Folgenden sind Situationen im Physikunterricht aufgeführt, in denen auf die Heterogenität einer Lerngruppe eingegangen werden soll. Zeigen digitale Medien in der jeweiligen Situation dafür ein besonderes Potenzial gegenüber nicht-digitalen Medien?

| | ja | nein |
|---|----|------|
| Umsetzung interessensspezifischer Experimentierprojekte | | |
| Vorbereitende Hausaufgabe zur individuellen Erarbeitung erster Grundlagen eines bevorstehenden Unterrichtsinhalts | | |
| Reduktion mathematischer Anforderungen bei der Auswertung von Experimenten | | |
| Bereitstellung passender Übungsangebote nach individuell diagnostiziertem Lernstand | | |

*** Frage zur Einschätzung***

Kann im Fachdidaktik-Studium Ihres Standorts das notwendige Wissen erworben werden, um diese Aufgabe korrekt zu lösen? (Bitte nur ein Kreuz setzen)

Ja, absolut

Nein, in keinem Fall

Anmerkung / Begründung (optional):

Abbildung 9-4: Beispiel aus der Befragung zur curricularen Passung der Testaufgaben¹⁸

¹⁸ Das abweichende Layout zu sonstigen Aufgabenbeispielen, die Screenshots des Online-Fragebogens aus SoSci Survey darstellen, ergibt sich, weil die Befragung zur curricularen Passung mithilfe eines bearbeitbaren PDF-Formulars mit einer anderen Formatierung erfolgte.

Zur Auswertung wurden die numerischen Einschätzungen von 1 (*Ja, absolut*) bis 6 (*Nein, in keinem Fall*) für alle Befragten in eine Datentabelle eingegeben. Für jede Aufgabe wurde der Mittelwert über alle zehn Einschätzungen gebildet, um einen Wert für die curriculare Passung der Aufgaben zu erhalten. Dieser Wert kann gemäß der Skala für jede Aufgabe zwischen 1 (sehr gute Passung) und 6 (gar keine Passung) liegen.

So ergeben sich für die curriculare Passung der fünfzehn untersuchten Testaufgaben Werte von 1,4 bis 2,7 und eine mittlere Passung von 1,9 (SD = 0,4). Abbildung 9-5 zeigt die Häufigkeitsverteilung für die berechneten Werte. Dabei wird deutlich, dass elf der fünfzehn Aufgaben eine sehr gute bis gute Passung ($\leq 2,0$) und die weiteren vier Aufgaben eine annehmbare Passung ($\leq 2,7$) aufweisen. Die geringste Passung mit einem Wert von 2,7 zeigt Aufgabe 7 (Beispielaufgabe in Abbildung 9-4). Insgesamt liegen jedoch alle Aufgaben in der „positiven“ Hälfte der Skala von 1 bis 6 und es gibt somit keine Aufgabe, die eine unbefriedigende Passung zeigt (Abbildung 9-5).

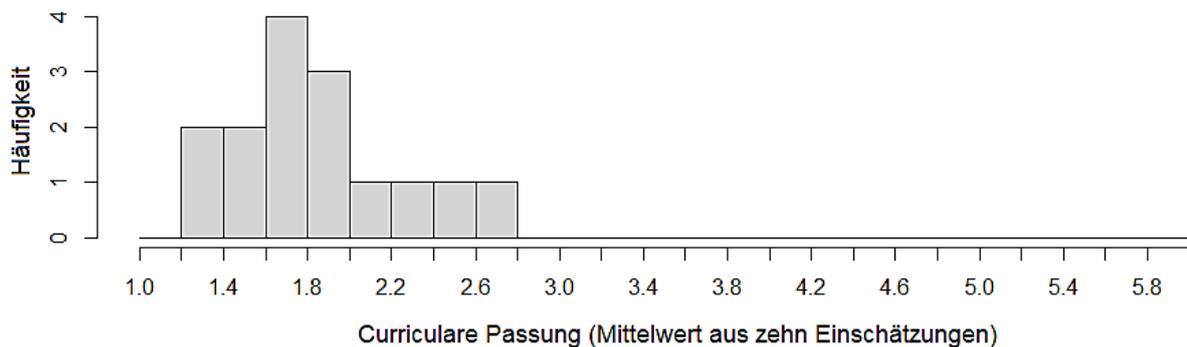


Abbildung 9-5: Häufigkeitsverteilung zur curricularen Passung der 15 Testaufgaben

Dieses Ergebnis der mindestens befriedigenden und überwiegend guten Passungen für alle Aufgaben stellt einen Hinweis für den inhaltlich-curricularen Aspekt der Validität im Hinblick darauf dar, dass mit dem entwickelten Test universitär erwerbbares Wissen adressiert wird.

Da alle Aufgaben mindestens eine annehmbare Passung aufzeigten, gab es aufgrund dieser Analyse der Einschätzungen zunächst keinen Anlass für das Entfernen von Aufgaben. Die zusätzliche optionale Abfrage offener Begründungen oder Bemerkungen ergab jedoch Hinweise auf mögliche Probleme

oder Verständnisschwierigkeiten in den Aufgaben oder einzelnen Antwortalternativen. So äußerten mehrere Expert:innen zu einer Aufgabe Bedenken hinsichtlich der Zugehörigkeit zum FDW zum Einsatz digitaler Medien. Aufgrund dieser Rückmeldungen und da die Aufgabe mit 2,4 eine der weniger guten curricularen Passungen besaß, wurde sie anknüpfend aus dem Testinstrument entfernt. Weiterhin ergaben sich aus den offenen Rückmeldungen Anlässe zu kleineren Überarbeitungen bei drei weiteren Aufgaben (z. B. Entfernen von Antwortalternativen oder Anpassen der Fragestellung).

In Tabelle 9-5 sind alle Überarbeitungen nach der Befragung zur curricularen Validierung dargestellt. Anschließend wurde das Testinstrument nicht mehr verändert.

Tabelle 9-5: Überarbeitungen der Testaufgaben nach der curricularen Validierung

| Aufgabe | Einordnung | Probleme | Veränderungen |
|-----------------|---|--|--|
| A. 1 | Fachbezogene Grundlagen, Analysieren | <ul style="list-style-type: none"> • Optionen könnten uneindeutig werden, wenn Einzelfälle mit untypischen Lernvoraussetzungen auftreten • zu viele leichte Attraktoren (Optionen 1, 2 und 4) • Option 3: Begriff „Tiefenstruktur“ unpassend • Option 5 spricht Lesekompetenz an | <ul style="list-style-type: none"> • im Aufgabenstamm „typischerweise“ ergänzen • Option 3: „sachlogische Struktur“ statt „Tiefenstruktur“ • Option 4 entfernen • Option 5 entfernen |
| A. 2 | Digitale Messwerterfassung, Reproduzieren | <ul style="list-style-type: none"> • Formulierung „allgemein als Vorteile“ | <ul style="list-style-type: none"> • „allgemein“ zu „typischerweise“ umformulieren |
| A. 4 (entfernt) | Simulationen, Analysieren | <ul style="list-style-type: none"> • Fragliche Zugehörigkeit zum Konstrukt | <ul style="list-style-type: none"> • ganz entfernen |

| | | | |
|-----------|--------------------------------------|--|--|
| A. 7 | Fachbezogene Grundlagen, Analysieren | <ul style="list-style-type: none"> • Option 1 nicht eindeutig | <ul style="list-style-type: none"> • Option 1 entfernt • neuer Distraktor: „Adressierung unterschiedlicher Interessenslagen“ |
| A. 15 (4) | Simulationen, Anwenden | | <ul style="list-style-type: none"> • verschieben/ umbenennen zu A. 4 |

In der ersten Spalte ist in Klammern zusätzlich jeweils die Nummer der Aufgabe im finalen Testinstrument angegeben, falls diese von der derzeitigen Nummerierung abweicht bzw. „entfernt“ angegeben, wenn diese nicht mehr im finalen Test enthalten ist.

9.5 Strukturelle Validierung durch Vergleich von IRT-Modellen

Um die angenommene Struktur des modellierten FDW zum Einsatz digitaler Medien zu prüfen, wurden das eindimensionale IRT-Modell und das vierdimensionale IRT-Modell (8.4) miteinander verglichen. Das vierdimensionale Modell berücksichtigt als Strukturierung die vier Kategorien *fachbezogene Grundlagen*, *digitale Messwerterfassung*, *Simulationen* und *Erklärvideos* der Facette digitale Medien des FDW (8.1).

Wie die Untersuchungen zur Testgüte auf Grundlage von 399 Datensätzen in 8.4 bereits verdeutlichen, zeigen sich annehmbare Modellfitwerte sowohl für die eindimensionale als auch für die vierdimensionale IRT-Modellierung nach Kategorien (Tabelle 8-7). Die EAP-Reliabilitäten für die vier möglichen Teilskalen liegen bei 0,43 (*fachbezogene Grundlagen*), 0,57 (*digitale Messwerterfassung*), 0,52 (*Simulationen*) und 0,54 (*Erklärvideos*). Sie fallen damit zum Teil deutlich geringer aus als die Reliabilität des Gesamttests in dieser Stichprobe (0,60; 8.4).

Der Vergleich beider Modellierungen ergibt eine statistisch signifikante Modellverbesserung im χ^2 -Differenzentest ($\chi^2(9, N = 399) = 22,77; p = 0,007$) zugunsten des vierdimensionalen Modells. Tabelle 9-6 stellt die Kennwerte des Modellvergleichs dar. So deutet der geringere Deviance-Wert für das vierdimensionale Modell zwar dessen Überlegenheit an, jedoch unterstützen die Informationskriterien AIC (*Akaike Information Criterion*) und BIC (*Bayesian Information Criterion*) diese nicht, da das AIC im vierdimensionalen Modell nahezu gleich und das BIC sogar höher ausfällt. Zudem handelt es sich beim

χ^2 -Differenzentest auch nicht um einen hoch signifikanten Unterschied. Insgesamt zeigt sich demnach keine deutliche Überlegenheit des vierdimensionalen Modells, was dafür spricht, dass die eindimensionale Modellierung mit ähnlich guten Modellfitwerten ebenso funktioniert.

Tabelle 9-6: Vergleich zwischen eindimensionalem IRT-Modell ohne Unterteilung (1D-Modell) und vierdimensionalem IRT-Modell mit der Unterteilung in vier Kategorien (4D-Modell)

| | 1D-Modell | 4D-Modell |
|-----------|-----------|-----------|
| Deviance | 10767,12 | 10744,36 |
| Parameter | 29 | 38 |
| AIC | 10825,12 | 10820,36 |
| BIC | 10940,80 | 10971,94 |

AIC: Akaike Information Criterion, BIC: Bayesian Information Criterion

Weiterhin lässt sich für das vierdimensionale Modell betrachten, wie die latenten Korrelationen zwischen den einzelnen Kategorien bzw. Dimensionen des IRT-Modells ausfallen. Hier deuten hohe Korrelationen darauf hin, dass statistisch betrachtet nicht von unabhängigen Dimensionen ausgegangen werden kann (Gramzow, 2015; Rindermann, 2006). Die latenten Korrelationen für die hier betrachteten vier Kategorien weisen Werte zwischen 0,46 und 0,76 auf (Abbildung 9-6). Die Korrelationen fallen nicht so hoch aus, dass statistisch trennbare Dimensionen auszuschließen sind. Jedoch stellt dies umgekehrt auch keinen Beleg dafür dar, dass es sich deshalb bei den Kategorien um inhaltliche Dimensionen des FDW zum Einsatz digitaler Medien handelt.

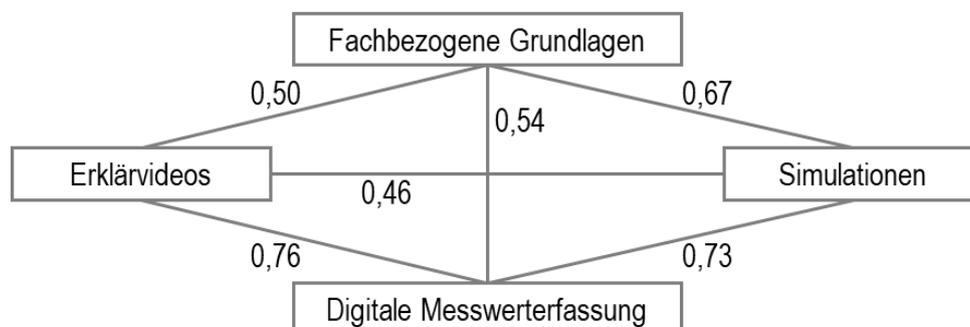


Abbildung 9-6: Latente Korrelationen der vier Dimensionen (Kategorien) im vierdimensionalen IRT-Modell

Der Vergleich der IRT-Modelle zur Prüfung der angenommenen Struktur des FDW zum Einsatz digitaler Medien aus vier Kategorien zeigt keine deutliche Überlegenheit des vierdimensionalen Modells. Stattdessen scheint die eindimensionale Modellierung ebenso geeignet zur Beschreibung der vorliegenden Daten und weist zudem eine höhere EAP-Reliabilität auf als die der vier möglichen Teilskalen. Hinsichtlich der strukturellen Validität lässt sich aus dieser Studie demnach kein klarer Hinweis ableiten, um die modellierte Struktur aus vier Kategorien empirisch zu stützen. Dies spricht dafür, das Testinstrument primär als eine Gesamtskala zu verwenden.

9.6 Zusammenfassung der Ergebnisse zur Untersuchung der Validität des entwickelten FDW-Tests zum Einsatz digitaler Medien

Ziel der vorgestellten Validierungsstudien war es, die Validität im Sinne einer Angemessenheit der mit dem Test geschlossenen Testwertinterpretation zu untersuchen, welche auf eine Erklärung und eine Bewertung von Leistungen abzielt. Es wurden verschiedene Teilstudien zu unterschiedlichen Validitätsaspekten durchgeführt, die im Sinne der Leistungserklärung Hinweise zu folgenden Fragen liefern sollen: (1) Wird universitär erwerbbares FDW zum Einsatz digitaler Medien gemessen? (2) Wird FDW zum Einsatz digitaler Medien gemessen, das sich von allgemein-mediendidaktischem und -pädagogischem sowie rein fachlichem Wissen abgrenzt? (3) Sind die angenommenen Kategorien zur Strukturierung des FDW zum Einsatz digitaler Medien empirisch gestützt?

Im Hinblick auf die erste Frage liefert die Expert:innenbefragung zur inhaltlich-curricularen Validierung Indizien dafür, dass die entwickelten Testaufgaben zum FDW zum Einsatz digitaler Medien an der Universität erwerbbares FDW adressieren (9.4). Alle Aufgaben weisen auf Basis der Einschätzungen von Physikdidaktiker:innen an zehn Hochschulstandorten eine gute bis sehr gute curriculare Passung auf. Nur eine Aufgabe wurde auf Basis der Anmerkungen in der Befragung aus dem Test entfernt, da die Passung zum Konstrukt von den Teilnehmenden in Frage gestellt wurde.

Hinsichtlich der zweiten Frage ergibt sich zunächst aus den Korrelationsanalysen mit verwandten Professionswissensbereichen der Hinweis darauf,

dass das gemessene Wissen einem *fachspezifischen* FDW zu digitalen Medien entspricht und weniger mit fachunabhängigem pädagogischem Wissen zusammenhängt (9.1). Die Studie zur diskriminanten Validität konnte jedoch keinen weiteren Hinweis zur Fachspezifität des gemessenen FDW liefern. Zwar erzielten die befragten Physiklehramtsstudierenden erwartungskonform bessere Testleistungen als die Lehramtsstudierenden der Fächer Chemie und Technik; jedoch zeigte eine Regressionsanalyse unter zusätzlichem Einbezug der Abiturnote, dass die Fachzugehörigkeit (Physik oder kein Physik) keinen signifikanten Einfluss mehr auf die Testleistungen im FDW hat (9.2). Da die betrachteten Fächer Chemie und Technik ebenfalls Teil des Fachbereichs MINT sind, könnte eine Erklärung für dieses Ergebnis sein, dass der Test weniger physikspezifisches als eher MINT-spezifisches FDW zum Einsatz digitaler Medien adressiert. Mit den vorliegenden Daten in dieser Studie kann hierzu jedoch keine Aussage vorgenommen werden. Weitere Indizien zur zweiten Frage geben wiederum die Think-Aloud-Studien zur kognitiven Validierung im Hinblick auf die Abgrenzung des untersuchten FDW zu weiteren Wissensbereichen wie dem Fachwissen. Bei der Beantwortung der entwickelten Testaufgaben wurde von den Teilnehmenden häufiger FDW als Wissensquelle herangezogen als fachliches Wissen oder Wissen basierend auf Schulerfahrungen (9.3). Einschränkend ist hier jedoch anzumerken, dass häufig die Begründungen und Erläuterungen der vier Teilnehmenden in den Think-Aloud-Interviews nicht eindeutig zu einer Wissensquelle zugeordnet werden konnten. Dies könnte einerseits daran liegen, dass die Befragten keine der Wissensquellen nutzten und aufgrund subjektiver Einschätzungen oder Intuition antworteten, was wiederum die Validität einschränken würde. Andererseits ist es auch möglich, dass die Befragten zwar bestimmtes Wissen nutzten, sie dieses jedoch nicht klar genug explizieren konnten (9.3).

Zuletzt wurde im Hinblick auf die dritte Frage die mögliche Strukturierung des FDW zum Einsatz digitaler Medien hinsichtlich der vier modellierten Kategorien empirisch geprüft. Wenngleich der Modellvergleich mittels Differenzentest signifikant ausfällt, sind die Hinweise für eine Überlegenheit des vierdimensionalen Modells nicht sehr deutlich. Zudem fallen die EAP-Reliabilitäten in den möglichen Teilskalen geringer aus als die des Gesamt-

tests (9.5). Das Testinstrument ist damit insbesondere für eine eindimensionale Betrachtung des FDW zum Einsatz digitaler Medien nutzbar. Die Unterscheidung in die vier Kategorien gemäß der Modellierung sollte primär zur inhaltlichen Charakterisierung der Testaufgaben dienen und weniger als Dimensionierung des FDW zum Einsatz digitaler Medien, welches einzeln nutzbare Teilskalen enthält, betrachtet werden.

Insgesamt bieten die vorgestellten Teilstudien verschiedene Indizien, die die Validität einer angemessenen Testwertinterpretation zur Erklärung und Bewertung von Leistungen im Test zum FDW zum Einsatz digitaler Medien unterstützen, wenngleich nicht alle Aspekte gleichermaßen gestützt werden konnten. Das optimierte und im Hinblick auf die verschiedenen Validitätsaspekte untersuchte Testinstrument wurde anknüpfend für die Messung des FDW zum Einsatz digitaler Medien bei Lehramtsstudierenden der Physik eingesetzt, um den Wissenserwerb über die entwickelten Seminare im Verbundprojekt DiKoLeP zu untersuchen.

10 Vorstellung des entwickelten Testinstruments

In diesem Kapitel wird das finale Testinstrument zur Erfassung des FDW zum Einsatz digitaler Medien in Physik vorgestellt. In 10.1 erfolgt zunächst die Beschreibung übergeordneter Merkmale des Testinstruments. Anschließend werden in 10.2 die Verteilung der Aufgaben auf die Kategorien und kognitiven Anforderungen aufgezeigt sowie exemplarisch mehrere Aufgabenbeispiele vorgestellt.

10.1 Übergeordnete Testmerkmale

Der entwickelte Test ist ein Leistungstest zur Erfassung des FDW zum Einsatz digitaler Medien in Physik, welches einen kognitiven Aspekt professioneller Kompetenz von angehenden Lehrkräften darstellt (2.1). Mithilfe des Testinstruments soll dabei universitär erwerbbares FDW zum Einsatz digitaler Medien im Fach Physik gemessen werden. Die Zielgruppe für den Testeinsatz bilden demnach Lehramtsstudierende im Fach Physik für die Sekundarstufe I und II (d. h. primär im Lehramt für Gymnasien und Gesamtschulen oder für Berufskollegs).

Für die Erhebungen in dieser Studie sowie im Gesamtprojekt ist das Testinstrument als Online-Test in *SoSci Survey*¹⁹ implementiert. Dabei ist bei der Bearbeitung der Testaufgaben kein Zurückgehen zu vorherigen Aufgaben möglich. Das Überspringen von Aufgaben bzw. Auslassen von (einzelnen) Antworten ist möglich, jedoch wird vom Umfragesystem beim Weiterklicken ein Fenster geöffnet, das auf die fehlende Beantwortung hinweist und nachfragt, ob die beantwortende Person trotzdem im Test weitergehen möchte, um ein Nichtbeantworten aus Versehen zu verhindern. Die Bearbeitungszeit für die Testaufgaben wurde auf 20-25 Minuten angesetzt, jedoch ist die Zeit in den Erhebungen nicht beschränkt worden. Im Mittel benötigten Teilnehmende etwa 22 Minuten zur Bearbeitung des Tests mit einer Standardabweichung von ca. 11 Minuten²⁰.

¹⁹ <https://www.soscsurvey.de/> (letzter Zugriff am 12.11.2024)

²⁰ Diese Angabe basiert auf einer Stichprobe von 154 beantworteten Datensätzen.

Aufbau der Testaufgaben und Bepunktung

Wie in 8.2 bereits beschrieben, bestehen alle Testaufgaben aus einem Aufgabenstamm mit einer Fragestellung sowie einem anknüpfenden Antwortformat. Der Aufgabenstamm der Aufgaben variiert dabei, je nachdem ob vorab eine (kurze) Unterrichtssituation beschrieben oder ein Medienbeispiel eingebunden wird oder der Stamm nur die Fragestellung enthält. Im Hinblick auf das Antwortformat stellen alle Aufgaben des finalen Testinstruments Mehrfachwahlaufgaben dar. Jede Aufgabe enthält dabei vier bis sieben Antwortoptionen, die jeweils eine Beantwortung mit *ja* oder *nein* einfordern.

Dieses Format der Mehrfachwahlaufgaben sowie die Bewertung dieser Aufgaben sind dabei orientiert an dem Typ K' (Kprim) nach Krebs (1997, 2004, 2019). Bei diesem Aufgabentyp wird für vier Aussagen oder Antwortoptionen zu einer Aufgabenstellung jeweils eine ja/nein- oder richtig/falsch-Entscheidung getroffen. Im Allgemeinen kann die Bewertung nach diesem Typ so erfolgen, dass nur dann ein Punkt in der Aufgabe vergeben wird, wenn alle vier Beurteilungen korrekt beantwortet sind. Es kann jedoch auch ein halber Punkt vergeben werden, wenn bereits drei der vier Beurteilungen korrekt sind.

Der Typ K' eignet sich bei Sachverhalten, bei denen inhaltlich mehrere Aspekte bedeutsam sein können, d. h. mehrere Elemente zu einer richtigen Lösung gehören, wobei die einzelnen Elemente eindeutig als falsch oder richtig zu bewerten sein müssen. In der Aufgabenstellung bleibt dabei offen, wie viele der Aussagen oder Optionen richtig sind (Krebs, 2004, 2019). Demnach passt dieser Aufgabentyp für die geschlossenen Aufgaben, die für den Test in dieser Arbeit angestrebt wurden, da z. B. für die Begründung eines Medieneinsatzes oder die Rekonstruktion von Zielstellungen zu einer Unterrichtssituation mehrere Aspekte bedeutsam und richtig sein können und die Anzahl richtiger Aspekte variieren kann. Zudem werden mit diesem Bewertungsformat die Antwortoptionen, die zu einer Aufgabe und damit zu einem Sachverhalt zählen, inhaltlich verdichtet. Das Bewertungsformat reduziert außerdem den Einfluss des Ratens, der in geschlossenen Aufgaben meist problematisch ist (8.2). Selbst mit der Halbpunkt-Bewertung muss über die Hälfte der Aussagen richtig beantwortet werden, um überhaupt Punkte zu erhalten.

In der Alles-oder-Nichts-Bewertung sind die Aufgaben sogar deutlich schwieriger als beispielsweise Einfachauswahlaufgaben (Krebs, 2004, 2019). Für das vorliegende Testinstrument wurde das Format adaptiert, da auch Aufgaben mit mehr als vier Optionen möglich sein sollten. In Anlehnung an die Halbpunkt-Bewertungsmethode wird bei richtiger Beantwortung von mehr als der Hälfte aller Alternativen bereits ein Punkt vergeben. Die volle Punktzahl von zwei Punkten gibt es bei richtiger Beantwortung aller Antwortalternativen oder (bei Aufgaben ab fünf Alternativen) aller bis auf eine Antwortalternative. Für das Beispiel von vier Antwortoptionen erhält man die volle Punktzahl von zwei Punkten für vier richtige Antworten, einen Punkt für drei richtige Antworten ($> 50\%$) und keinen Punkt bei zwei oder weniger richtigen Antworten ($\leq 50\%$). Übertragen auf eine Aufgabe mit beispielsweise sechs Antwortoptionen werden zwei Punkte bei sechs oder fünf richtigen Antworten, ein Punkt bei vier richtigen Antworten ($> 50\%$) und keine Punkte bei drei oder weniger richtigen Antworten ($\leq 50\%$) vergeben. Die Bewertungsregeln für jede mögliche Anzahl an Antwortalternativen befinden sich in Anhang C.

10.2 Aufgabenbeispiele

In diesem Abschnitt werden nachfolgend verschiedene Aufgabenbeispiele vorgestellt. Das gesamte Instrument zum FDW zum Einsatz digitaler Medien aus 14 Mehrfachwahlaufgaben inklusive Musterlösung und Bewertungsregeln befindet sich im Anhang dieser Arbeit (Anhang B).

Die Verteilung der 14 finalen Testaufgaben auf die Kategorien und kognitiven Anforderungen im Sinne des Itementwicklungsmodells ist in Tabelle 10-1 aufgezeigt. Hier ist zu sehen, dass auch nach dem Entfernen von drei Aufgaben nach der Pilotierung und den Validierungsstudien alle sinnvollen Kombinationen aus kognitiver Anforderung und Kategorie mit mindestens einer Aufgabe vertreten sind (d. h. ausgenommen von der Kombination der weniger physikspezifischen Kategorien *Erklärvideos* und *fachbezogene Grundlagen* mit der kognitiven Anforderung *Reproduzieren*; 8.2).

Tabelle 10-1: Verteilung der 14 Testaufgaben des finalen Testinstruments auf Kategorien und kognitive Anforderungen

| | Reproduzieren | Anwenden | Analysieren | Summe |
|----------------------------|---------------|----------|-------------|-------|
| Fachbezogene Grundlagen | - | 2 | 2 | 4 |
| Digitale Messwerterfassung | 2 | 1 | 1 | 4 |
| Simulationen | 1 | 1 | 1 | 3 |
| Erklärvideos | - | 1 | 2 | 3 |
| <i>Summe</i> | 3 | 5 | 6 | 14 |

In Tabelle 8-2 und Tabelle 8-3 wurde bereits beschrieben, was die einzelnen Kategorien und kognitiven Anforderungen umfassen. Aus diesen Übersichten werden die Erläuterungen, welche Aufgaben zu einer Kategorie oder kognitiven Anforderung zählen, aufgegriffen und im Folgenden um eine Spalte zur Angabe aller zugehörigen Testaufgaben erweitert (Tabelle 10-2 und Tabelle 10-3).

In den nachfolgenden Abschnitten werden die fünf in den Tabellen fett gedruckten Aufgaben exemplarisch vorgestellt und näher erläutert.

Tabelle 10-2: Anwendung der Kategorien und Zuordnung der zugehörigen Aufgaben

| Kategorie | Dazu zählen... | Zugehörige Aufgaben |
|-------------------------|---|--|
| Fachbezogene Grundlagen | <p>Aufgaben, in denen mehrere Beispiele zu verschiedenen Medien aufgegriffen werden.</p> <p>Aufgaben, in denen übergreifende Aspekte zu einem (spezifischen) Medieneinsatz abgefragt werden, die nicht für diese Medienart charakteristisch sind.</p> | <p>Aufgabe 1</p> <p>Aufgabe 3</p> <p>Aufgabe 7</p> <p>Aufgabe 11</p> |

| | | |
|----------------------------|--|--|
| Digitale Messwerterfassung | <p>Aufgaben, die sich ausschließlich mit der Nutzung digitaler Messwerterfassung auseinandersetzen.</p> <p>Aufgaben, die sich mit der Nutzung digitaler Messwerterfassung im Vergleich zu traditioneller Messmethodik auseinandersetzen.</p> <p>Aufgaben, die eine digital-gestützte Auswertung und Analyse von Messwerten mit solchen Systemen adressieren.</p> | <p>Aufgabe 2</p> <p>Aufgabe 5</p> <p>Aufgabe 10</p> <p>Aufgabe 14</p> |
| Simulationen | <p>Aufgaben, die sich ausschließlich mit der Nutzung von Simulationen im Physikunterricht auseinandersetzen.</p> | <p>Aufgabe 4</p> <p>Aufgabe 8</p> <p>Aufgabe 12</p> |
| Erklärvideos | <p>Aufgaben, die sich ausschließlich mit Erklärvideos zur Physik oder deren Einsatzmöglichkeiten im Physikunterricht auseinandersetzen.</p> | <p>Aufgabe 6</p> <p>Aufgabe 9</p> <p>Aufgabe 3</p> |

Tabelle 10-3: Anwendung der kognitiven Anforderungen und Zuordnung der zugehörigen Aufgaben

| Kognitive Anforderung | Dazu zählen... | Zugehörige Aufgaben |
|------------------------------|---|--|
| Reproduzieren | <p>Aufgaben, in denen allgemeine Vorteile, Prinzipien oder Kriterien zum Einsatz digitaler Medien im Physikunterricht unabhängig von einer konkreten Unterrichtssituation erkannt werden sollen.</p> | <p>Aufgabe 2</p> <p>Aufgabe 8</p> <p>Aufgabe 14</p> |
| Anwenden | <p>Aufgaben, in denen ein geeignetes Medium oder ein geeigneter Medieneinsatz zu einem bestimmten didaktischen Einsatzzweck ausgewählt werden soll.</p> <p>Aufgaben, in denen Begründungen oder Argumente für einen gegebenen Medieneinsatz in einer Situation oder zu einem Zweck ausgewählt werden sollen.</p> <p>Aufgaben, in denen Aspekte für einen bestimmten Medieneinsatz berücksichtigt werden sollen.</p> | <p>Aufgabe 3</p> <p>Aufgabe 4</p> <p>Aufgabe 5</p> <p>Aufgabe 9</p> <p>Aufgabe 11</p> |

| | | |
|--------------------|---|--|
| <p>Analysieren</p> | <p>Aufgaben, in denen zu einer beschriebenen Unterrichtssituation geprüft werden soll, ob allgemeine Kriterien für einen Medieneinsatz erfüllt sind oder nicht.</p> <p>Aufgaben, in denen aus einer gegebenen Unterrichtssituation Zielstellungen rekonstruiert werden sollen.</p> <p>Aufgaben, in denen mehrere verschiedene Situationen im Hinblick auf einen Gesichtspunkt bewertet werden sollen.</p> | <p>Aufgabe 1 Aufgabe 6 Aufgabe 7 Aufgabe 10 Aufgabe 12 Aufgabe 13</p> |
|--------------------|---|--|

10.2.1 Aufgabe 3: Interesse am Unterrichtsthema wecken

Die dritte Aufgabe des Testinstruments zählt zur Kategorie *fachbezogene Grundlagen* und der kognitiven Anforderung *Anwenden*. Ausgehend von der groben Beschreibung einer Unterrichtssituation im Aufgabenstamm sollen die Teilnehmenden auswählen, ob die vorgegebenen Antwortalternativen einen besonders geeigneten Medieneinsatz für den im Stamm beschriebenen didaktischen Einsatzzweck darstellen, das Interesse am Unterrichtsthema Plattenkondensator zu wecken (Abbildung 10-1). Für die Auswahl der Antworten genügt keine Wiedererkennungslleistung, da kein reines Faktenwissen abgefragt wird, sondern Wissen zum Einsatz digitaler Medien durch das Auswählen eines Medieneinsatzes für die gegebene Situation angewendet werden soll. Es werden mehrere Medienbeispiele aufgegriffen und nicht ausschließlich ein bestimmtes Medium betrachtet; somit zählt die Aufgabe zur medienübergreifenden Kategorie *fachbezogene Grundlagen*. Durch das vorgegebene Unterrichtsthema Plattenkondensator ist die Aufgabe in den physikalischen Inhaltsbereich Elektrizitätslehre eingebettet. Diese Aufgabe wurde zunächst mit einem offenen Aufgabenformat entwickelt (Abbildung 8-4) und nach der Pilotierung im Austausch im Verbundprojekt zu einer Mehrfachwahlaufgabe geschlossen (8.3), um formal zu den anderen Aufgaben zu passen.

1. Fachdidaktischen Wissen zum Einsatz digitaler Medien – Teil 3 von 14

Sie behandeln in der Sekundarstufe II den Plattenkondensator und wollen im Unterricht Interesse für das Thema wecken.

Handelt es sich bei den folgenden Beschreibungen um einen besonders geeigneten Medieneinsatz für diesen didaktischen Einsatzzweck?

| | Nein | Ja |
|--|-----------------------|-----------------------|
| Ein Erklärvideo zum Plattenkondensator schauen lassen, in dem der Aufbau und die Funktionsweise des Kondensators erklärt werden. | <input type="radio"/> | <input type="radio"/> |
| Eine Simulation zeigen, die die Entstehung des elektrischen Feldes im Kondensator nach Anlegen einer Spannung visualisiert. | <input type="radio"/> | <input type="radio"/> |
| Ein digitales Messwerterfassungssystem verwenden, um das Potenzial des Themas zur Anwendung moderner Technologien im Physikunterricht zu zeigen. | <input type="radio"/> | <input type="radio"/> |
| Ein kurzes Video aus der Tagespresse vorführen, das aufzeigt, inwiefern der Plattenkondensator im Alltag nützlich sein kann. | <input type="radio"/> | <input type="radio"/> |
| Ein digitales Lernquiz zum Einstieg in das Thema durchführen, um mögliche Schülervorstellungen zum Kondensator zu erkunden. | <input type="radio"/> | <input type="radio"/> |

Abbildung 10-1: Aufgabe 3 des finalen Testinstruments

(Kategorie: *fachbezogene Grundlagen*, kognitive Anforderung: *Anwenden*; Musterlösung: nein, nein, nein, ja, nein)

Die einzige richtige Antwort ist die vierte Option, da hier ein Alltagsbezug geschaffen wird, welcher für die Interessensförderung nützlich ist (z. B. Duit & Mikelskis-Seifert, 2010). Die anderen Antwortoptionen (Distraktoren) sind dabei teilweise aus den Überlegungen und Erkenntnissen zur vorausgehenden Version der Aufgabe im offenen Format entstanden (z. B. erste Option zum Erklärvideo oder fünfte Option zum Lernquiz), da diese für andere didaktische Einsatzziele geeignet wären, jedoch nicht typischerweise zur Interessensförderung. Die zweite Option ist sehr innerphysikalisch und damit für einen Einstieg in das Unterrichtsthema wenig interessensfördernd. Der Distraktor an der dritten Position zielt auf eine bloße Motivierung um des Mediums willen ab, in der Aufgabenstellung ist jedoch das Interesse wecken am physikalischen Unterrichtsinhalt gefragt.

10.2.2 Aufgabe 7: Potenzial digitaler Medien zur Berücksichtigung der Heterogenität

Die siebte Aufgabe des entwickelten Tests ist der Kategorie *fachbezogene Grundlagen* und der kognitiven Anforderung *Analysieren* zugehörig. Für mehrere gegebene Situationen zur Berücksichtigung der Heterogenität einer

Lerngruppe soll analysiert und bewertet werden, inwiefern der Einsatz digitaler Medien dafür ein besonders Potential zeigt. Es wird kein spezifisches Medium und auch kein physikalischer Inhaltsbereich fokussiert betrachtet (Abbildung 10-2).

1. Fachdidaktisches Wissen zum Einsatz digitaler Medien – Teil 7 von 14

Im Folgenden sind Situationen im Physikunterricht aufgeführt, in denen auf die Heterogenität einer Lerngruppe eingegangen werden soll.

Zeigen digitale Medien in der jeweiligen Situation dafür ein besonderes Potenzial gegenüber nicht-digitalen Medien?

| | Nein | Ja |
|---|-----------------------|-----------------------|
| Vorbereitende Hausaufgabe zur individuellen Erarbeitung erster Grundlagen eines bevorstehenden Unterrichtsinhalts | <input type="radio"/> | <input type="radio"/> |
| Reduktion mathematischer Anforderungen bei der Auswertung von Experimenten | <input type="radio"/> | <input type="radio"/> |
| Bereitstellung passender Übungsangebote nach individuell diagnostiziertem Lernstand | <input type="radio"/> | <input type="radio"/> |
| Adressierung unterschiedlicher Interessenslagen von Schülerinnen und Schülern | <input type="radio"/> | <input type="radio"/> |

Abbildung 10-2: Aufgabe 7 des finalen Testinstruments

(Kategorie: *fachbezogene Grundlagen*, kognitive Anforderung: *Analysieren*; Musterlösung: nein, ja, ja, nein)

Die beiden richtigen Antworten befinden sich an zweiter und dritter Stelle. Beispielsweise ermöglicht die Nutzung digitaler Messwerterfassungssysteme häufig eine automatisierte Auswertung der Messdaten, sodass bestimmte mathematische Auswertungsschritte ausgelagert (z. B. Bestimmung von Regressionsgeraden) und mögliche Beeinträchtigungen des Physikunterrichts durch mathematische Schwierigkeiten bei den Lernenden vermieden werden können (Lampe et al., 2015)²¹. Mithilfe intelligenter tutorieller Programme ist es möglich, adaptive Übungsangebote je nach Leistungsstand bereitzustellen (Girwidz, 2020c); während eine Lehrkraft dies im regulären Unterrichtsalltag kaum leisten kann. Die erste Option stellt einen Distraktor dar. Man könnte an dieser Stelle vermuten, dass ein Erklärvideo diese Möglichkeit

²¹ Dieser Vorteil, mithilfe digitaler Messwerterfassung mathematische Schwierigkeiten zu umgehen, meint nicht, dass diesen Defiziten gar nicht begegnet werden sollte. Dies sollte an anderer Stelle definitiv geschehen. Für eine Situation im Physikunterricht, in der beispielsweise jedoch die Auswertung und Interpretation eines Experiments aus physikalischer Sicht im Vordergrund steht, kann es nützlich sein, das Hindernis und den Zeitaufwand aufgrund mathematischer Defizite mithilfe automatisierten Auswertungsmöglichkeiten zu vermeiden.

bietet, jedoch liegt im Hinblick auf die Berücksichtigung der Heterogenität hier kein direkter Mehrwert gegenüber der individuellen Vorbereitung beispielsweise mithilfe eines Schulbuches vor. Der Aspekt der Multimodalität (d. h. die Kombination von visuellen und auditiven Inhalten), den ein Erklärvideo gegenüber einem Schulbuch leistet, ist für die Berücksichtigung der Heterogenität weniger bedeutsam. Ebenso besteht die Möglichkeit, unterschiedliche Interessenslagen von Schülerinnen und Schülern zu adressieren (vierte Option), nicht im Besonderen für digitale Medien, sondern wird vielmehr durch die Einbettung der physikalischen Inhalte in interessante und anwendungsbezogene Kontexte erreicht (Duit & Mikelskis-Seifert, 2010).

10.2.3 Aufgabe 8: Vorteile beim Einsatz von Simulationen im Physikunterricht

Die achte Aufgabe im Test zählt zur Kategorie *Simulationen* und der kognitiven Anforderung *Reproduzieren*. Unabhängig von einer Unterrichtssituation sollen die Antwortalternativen bejaht werden, die für einen Einsatz von Simulationen im Physikunterricht sprechen. Demnach wird in der Aufgabe Wissen zu Gestaltungsmerkmalen und typischen Vorteilen von Simulationen im Physikunterricht reproduziert – unabhängig von einem physikalischen Inhaltsbereich oder einer Unterrichtssituation (Abbildung 10-3).

1. Fachdidaktisches Wissen zum Einsatz digitaler Medien – Teil 8 von 14

Handelt es sich im Folgenden um Aspekte, die generell für einen Einsatz von Simulationen im Physikunterricht sprechen?

| | Nein | Ja |
|--|-----------------------|-----------------------|
| Simulationen bieten eine präzise Abbildung der Wirklichkeit. | <input type="radio"/> | <input type="radio"/> |
| Simulationen können nicht direkt sichtbare physikalische Sachverhalte visualisieren. | <input type="radio"/> | <input type="radio"/> |
| Simulationen können als Beispiel eines wissenschaftlichen Modells thematisiert werden. | <input type="radio"/> | <input type="radio"/> |
| Simulationen ermöglichen das Abdecken von Lernzielen durch die Strukturierung des Lernprozesses. | <input type="radio"/> | <input type="radio"/> |
| Simulationen bereichern den Lernprozess durch fachbezogene Übungsaufgaben an. | <input type="radio"/> | <input type="radio"/> |
| Simulationen unterstützen die Modellbildung durch das Ausblenden weniger relevanter Aspekte. | <input type="radio"/> | <input type="radio"/> |

Abbildung 10-3: Aufgabe 8 des finalen Testinstruments

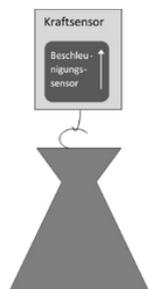
(Kategorie: *Simulationen*, kognitive Anforderung: *Reproduzieren*; Musterlösung: nein, ja, ja, nein, nein, ja)

Die richtigen Alternativen (zweite, dritte und sechste Option) ergeben sich aus typischen Vorteilen und Möglichkeiten von Simulationen für den Physikunterricht, wie sie auch in fachdidaktischer Literatur aufgezeigt werden (z. B. Girwidz, 2020c; Wieman et al., 2010). Da Simulationen meist ein bewusst vereinfachtes Modell eines realen Prozesses oder Phänomens darstellen (Rutten et al., 2012), ist die erste Option falsch. Die vierte Option und fünfte Option stellen ebenfalls Distraktoren dar, da Simulationen selbst weder Lernziele vorgeben noch den Lernprozess strukturieren oder mit Aufgaben anreichern. Diese Aspekte ergeben sich erst über die Einbettung der Simulation in das Unterrichtskonzept und begleitende Aufgabenstellungen und Maßnahmen durch die Lehrkraft (Girwidz, 2020c).

10.2.4 Aufgabe 10: Zielstellung für den Einsatz digitaler Messwerterfassung

Die zehnte Aufgabe des Testinstruments ist Teil der Kategorie *digitale Messwerterfassung* und der kognitiven Anforderung *Analysieren* (Abbildung 10-4).

1. Fachdidaktisches Wissen zum Einsatz digitaler Medien – Teil 10 von 14



Eine Lehrkraft führt in einer Unterrichtsstunde zum Grundgesetz der Mechanik $F = m \cdot a$ ein Experiment durch, in welchem ein Massestück wiederholt angehoben und damit beschleunigt wird. Sie verwendet dazu ein digitales Messwerterfassungssystem mit einem Beschleunigungs- und einem Kraftsensor, um beim dynamischen Vorgang die beschleunigende Kraft und die erzielte Beschleunigung direkt zu messen. Die Messdaten können jeweils als zeitlicher Verlauf sowie in Abhängigkeit zueinander grafisch aufgetragen werden.

Handelt es sich im Folgenden um typische Zielstellungen (didaktische Funktionen) für die digitale Messwerterfassung im Physikunterricht, unter denen dieses Vorgehen sinnvoll ist?

| | Nein | Ja |
|---|-----------------------|-----------------------|
| Transparenz des Messvorgangs für die Lernenden | <input type="radio"/> | <input type="radio"/> |
| Auslagerung von Routinearbeiten zur Fokussierung auf die Dateninterpretation | <input type="radio"/> | <input type="radio"/> |
| Abdeckung des Motivationsaspekts durch die Nutzung digitaler Messtechnik | <input type="radio"/> | <input type="radio"/> |
| Zeiteffiziente Untersuchung des physikalischen Zusammenhangs | <input type="radio"/> | <input type="radio"/> |
| Verständnisförderung durch die simultane Visualisierung der Messdaten in verschiedenen Darstellungsformen | <input type="radio"/> | <input type="radio"/> |

Abbildung 10-4: Aufgabe 10 des finalen Testinstruments

(Kategorie: *digitale Messwerterfassung*; kognitive Anforderung: *Analysieren*; Musterlösung: nein, ja, nein, ja, ja)

Im Aufgabenstamm wird eine Unterrichtssituation aus dem physikalischen Inhaltsgebiet Mechanik beschrieben, in der in einem Experiment Beschleunigung und Kraft mithilfe eines digitalen Messwerterfassungssystems aufgenommen und ausgewertet werden. Die Teilnehmenden sollen die Zielstellung für diese Nutzung digitaler Messwerterfassung aus fachdidaktischer Sicht rekonstruieren und die nachfolgenden Antwortoptionen dahingehend einschätzen (Abbildung 10-4).

Die Idee zu dieser Aufgabe und für das dargestellte Experiment stammt aus einem Themenheft der MNU-Themenreihe Bildungsstandards²² zu Experimenten mit digitaler Messwerterfassung in den Naturwissenschaften (Lampe et al., 2015). Die drei richtigen Antwortalternativen (zweite, vierte und fünfte Option) ergeben sich aus den im Themenheft beschriebenen Aspekten zum fachdidaktischen Mehrwert digitaler Messwerterfassung, für die dieses Experiment ein Beispiel darstellt (einfachere und schnellere Messungen, Messdaten in verschiedenen Darstellungsformen; vgl. Lampe et al., 2015, S. 5). Der erste Distraktor (erste Option) greift den Aspekt auf, dass digitale Messwerterfassung im Vergleich zu manchen analogen Messgeräten weniger transparent ist (Konrad, 2019). So würde beispielsweise eine analoge Messung der Kraft mittels Federkraftmesser anschaulicher darstellen, dass eine größere Auslenkung des Messgerätes einer größeren Kraft entspricht. Die dritte Option greift als Distraktor auf, dass dem Einsatz digitaler Medien ein Motivationsbonus zugesprochen wird (z. B. Hillmayr et al., 2020). Dieser ist jedoch nicht unbedingt langfristig und sollte nicht als einziger Aspekt zur Motivierung betrachtet werden – vielmehr ist das Wecken von Interesse am Inhalt des Experiments oder an der zentralen Fragestellung bedeutsam (Lampe et al., 2015).

10.2.5 Aufgabe 13: Einbettung eines Erklärvideos in den Unterrichtseinstieg

Die dreizehnte Aufgabe im entwickelten Testinstrument zählt zur Kategorie *Erklärvideos* und kognitiven Anforderung *Analysieren*. In dieser Aufgabe wird eine Unterrichtssituation zum Thema Schwingung eines Fadenpendels

²² MNU: Deutscher Verein zur Förderung des mathematischen und naturwissenschaftlichen Unterrichts

(physikalischer Inhaltsbereich Mechanik) beschrieben. Eine Lehrkraft lässt ihre Schülerinnen und Schüler zum Einstieg individuell ein Erklärvideo ansehen, das die physikalischen Zusammenhänge zwischen der Periodendauer und ihren Einflussgrößen erklärt. Anschließend wird in der Stunde ein Schülerexperiment zur Untersuchung der Einflussgrößen auf die Periodendauer durchgeführt. In der Testaufgabe sollen die Befragten den im Aufgabenstamm beschriebenen Medieneinsatz analysieren, indem sie die nachfolgenden Aussagen dazu einschätzen (Abbildung 10-5).

1. Fachdidaktisches Wissen zum Einsatz digitaler Medien – Teil 13 von 14

Eine Lehrkraft bittet die Klasse zum Einstieg in eine Unterrichtsstunde zur Schwingung eines Fadenpendels ein in die Schulcloud hochgeladenes Erklärvideo mit Smartphone und Headset anzusehen, um die Lerngruppe zu motivieren. Das Video gibt einen ersten Überblick über die Zusammenhänge zwischen der Periodendauer eines Fadenpendels und ihren Einflussgrößen. Anschließend werden in der Unterrichtsstunde die Einflussgrößen der Periodendauer in einem Schülerexperiment untersucht.

Beurteilen Sie den Einsatz des Erklärvideos in dieser Unterrichtsstunde. Sind die folgenden Aussagen für die beschriebene Situation zutreffend?

| | Nein | Ja |
|---|-----------------------|-----------------------|
| Der Einsatz des Erklärvideos erhöht die Motivation und steigert die Lernbereitschaft während der gesamten Unterrichtsstunde. | <input type="radio"/> | <input type="radio"/> |
| Das technische Potential des Erklärvideos wird gut genutzt, da die Schülerinnen und Schüler das Video mit individuellem Tempo (zurückspulen, pausieren) ansehen können. | <input type="radio"/> | <input type="radio"/> |
| Das Erklärvideo wird sinnvoll in den Lernprozess eingebettet, indem der zentrale Inhalt der Stunde vorbereitet wird. | <input type="radio"/> | <input type="radio"/> |
| Das Video kann die Abhängigkeit und Unabhängigkeit verschiedener physikalischer Größen auf die Periodendauer des Fadenpendels überzeugend darstellen. | <input type="radio"/> | <input type="radio"/> |
| Die erkenntnisgewinnende Funktion des Experiments geht durch die Nutzung des Erklärvideos zum Einstieg verloren. | <input type="radio"/> | <input type="radio"/> |

Abbildung 10-5: Aufgabe 13 des finalen Testinstruments

(Kategorie: *Erklärvideos*, kognitive Anforderung: *Analysieren*; Musterlösung: nein, ja, nein, nein, ja)

Die Grundidee der Aufgabe war es, eine nicht besonders gelungene Einbettung eines Erklärvideos in die Unterrichtsstunde darzustellen. So gibt das Video bereits Inhalte vor, die im späteren Experiment in der Stunde erarbeitet werden sollen. Der zentrale Inhalt der Stunde wird demnach eher vorweggenommen als vorbereitet (Distraktor dritte Option) und die Funktion der Erkenntnisgewinnung im Experiment aufgelöst (Attraktor fünfte Option). Aus diesem Grund werden die Lernbereitschaft und Motivation während des Experimentierens nicht gesteigert (Distraktor erste Option). Die Abhängigkeit und Unabhängigkeit der Größen auf die Schwingungsdauer über ein Erklär-

video darzustellen, ist zudem wenig überzeugend; insbesondere im Vergleich zur eigenen Erarbeitung der Abhängigkeiten im Realexperiment (Distraktor vierte Option). Lediglich das technische Potenzial eines Erklärvideos wird in dieser Einbettung genutzt, da die Lernenden dieses individuell ansehen (Attraktor zweite Option).

Diese Aufgabe stellte sich in ihrer vorausgehenden Version in der Pilotierung noch als zu schwierig heraus (leichter Bodeneffekt). So wurde eine Präzision des Aufgabenstamms und der Fragestellung vorgenommen und eine Antwortalternative neu formuliert (Tabelle 8-6), um die Eindeutigkeit und Qualität der Attraktoren und Distraktoren im Hinblick auf die Situation zu verbessern.

Das finale Testinstrument wurde schließlich für die Evaluation der Lehrveranstaltungen in der Hauptstudie dieser Arbeit eingesetzt, um die Entwicklung des FDW zum Einsatz digitaler Medien der Studierenden über die besuchten Seminare zu untersuchen. Die Ergebnisse dieser Untersuchung werden im nachfolgenden Kapitel beschrieben.

11 Ergebnisse zur Untersuchung des fachdidaktischen Wissenserwerbs

Dieses Kapitel beschreibt die Ergebnisse der summativen Evaluation der untersuchten Seminare im Verbundprojekt DiKoLeP im Hinblick auf das FDW zum Einsatz digitaler Medien im Fach Physik.

Die Evaluation der Seminare stellte nach der Modellierung des fokussierten FDW sowie der Testentwicklung und den Untersuchungen zur Validität das dritte Ziel der vorliegenden Arbeit dar. In diesem Zusammenhang wurde in FF2 untersucht, inwiefern sich dieses Wissen bei den beteiligten Physiklehramtsstudierenden über den Besuch der Seminare verändert. Gemäß der formulierten Wirkungshypothese H1 wurde erwartet, dass das gemessene FDW über den Untersuchungszeitraum zunimmt (6.2), da die Seminare die Förderung eben dieses Wissensbereichs adressieren und eine entsprechende inhaltliche Passung dazu aufweisen. Zudem deutet die bisherige Forschung darauf hin, dass FDW in fachdidaktischen Lehrveranstaltungen gesteigert werden kann (Joswig-Käfer, 2024; Riese et al., 2017; Sorge et al., 2019; 6.2). Gleichzeitig wurden jedoch über den kurzen Zeitraum von nur einem Semester keine großen Veränderungen im untersuchten FDW erwartet.

Die Befragung zu dieser Evaluation enthielt (im Prätest) zunächst Abfragen zu demografischen Daten (Geschlecht, Alter, Hochschule/Universität, Ausbildungsphase, Fachsemester, Schulabschluss, Abiturnote bzw. Mathematik-Maturanote²³, absolvierte Schulpraktika, Medieneinsatz in Schulpraktika). Das entwickelte Testinstrument zum FDW zum Einsatz digitaler Medien wurde für die Untersuchung im Prä- und Posttest eingesetzt. Da die Prä-Post-Erhebungen zum Teil noch parallel zu den Validierungsstudien stattfanden, konnte nicht für alle Befragungen die finale Testversion aus 14 Mehrfachwahlaufgaben (wie in Kapitel 10 beschrieben) eingesetzt werden. Für die Auswertung der mit einer vorausgehenden Testversion erhobenen Daten

²³ Bei den Befragungen in Graz musste anstelle der Abiturnote (als möglicher Indikator für kognitive Leistungsfähigkeit) auf die Abfrage einer Maturanote ausgewichen werden. Von den Maturanoten in den zu maturierenden Fächern Deutsch, Mathematik und eine lebende Fremdsprache wurde für die Befragung im Verbundprojekt die Mathematik-Maturanote als Pendant zur Abiturnote ausgewählt.

wurden jedoch spätere Änderungen am Testinstrument berücksichtigt²⁴. Im Rahmen des Verbundprojekts DiKoLeP wurden außerdem verschiedene weitere Konstrukte in den Prä-Post-Erhebungen erfasst (z. B. motivationale Orientierungen zum Einsatz digitaler Medien; 5.1); die Ergebnisse dazu werden in dieser Arbeit jedoch nicht betrachtet.

11.1 Stichprobe der Haupterhebung zur Untersuchung des Wissenserwerbs im FDW zum Einsatz digitaler Medien

Die Prä-Post-Erhebungen zum FDW zum Einsatz digitaler Medien erfolgten vom Sommersemester 2021 bis zum Sommersemester 2023 in insgesamt fünf Durchgängen und an insgesamt vier Projektstandorten. Je nach Studienangebot an den einzelnen Standorten wurden je Semester zwei Seminare evaluiert²⁵. Insgesamt wurden in dieser Evaluation $N_1 = 70$ Studierende befragt, die an den untersuchten Seminaren und den freiwilligen Prä- und Post-Befragungen teilnahmen. Eine Übersicht der Personenmerkmale in der untersuchten Stichprobe zeigt Tabelle 11-1.

Tabelle 11-1: Übersicht zur Stichprobe in der Hauptstudie zur Untersuchung der Entwicklung im FDW zum Einsatz digitaler Medien in den beteiligten Seminaren (FF1)

| Stichprobe | Standorte | Geschlecht | Alter | Ausbildungsphase | Fachsemester |
|------------|-------------------------------|----------------------------|--------------------------|------------------------------------|-------------------------|
| $N_1 = 70$ | 10 AC, 30 GR, 11 PB, 19 TÜ | 26 weiblich 44 männlich | $M = 24$ ($SD = 3$) | 38 BA (Lehramt) 32 MA (Lehramt) | $M = 8$ ($SD = 3$) |

AC: Aachen; GR: Graz; PB: Paderborn; TÜ: Tübingen; BA: Bachelor; MA: Master

²⁴ Für das finale Testinstrument entfernte Aufgaben oder Antwortalternativen wurden demnach in der Auswertung dieser Daten nicht einbezogen. Da z. B. in den Aufgaben 3 und 7 jedoch Antwortalternativen für die finale Testversion neu ergänzt wurden, treten an diesen Stellen für die Datensätze, die mit der vorausgehenden Testversion erhoben wurden, fehlende Werte auf. Durch die Nutzung einer IRT-Modellierung und der Schätzung von Personenfähigkeiten sind diese fehlenden Werte jedoch weniger problematisch als es bspw. bei alleiniger Betrachtung der klassischen Summenscores der Fall wäre.

²⁵ Im Sommersemester 2021 wurden die Seminare in Graz und Tübingen analysiert. Die Stichprobe aus Graz wurde in diesem Semester jedoch aus der Hauptstudie entfernt, da das Seminarkonzept pandemiebedingt im praktischen Teil nicht in angemessener Form durchgeführt werden konnte. Im Wintersemester 2021/22 wurden die Seminare in Aachen und Graz evaluiert, im Sommersemester 2022 die Seminare in Graz und Tübingen, im Wintersemester 2022/23 in Aachen und Graz und im Sommersemester 2023 die Seminare in Paderborn und Tübingen.

Hinsichtlich der Abschlussnote zur Hochschulzugangsberechtigung ergab sich für die 40 Studierenden in Deutschland eine mittlere Abiturnote von 1,82 (SD = 0,54) und für die 30 Studierenden in Österreich eine mittlere Mathematik-Maturanote von 2,00 (SD = 0,98).

Von den teilnehmenden Studierenden gaben fast alle an, mindestens ein Schulpraktikum in ihrem bisherigen Lehramtsstudium absolviert zu haben (alle bis auf eine Person aus Österreich). Von den 40 Teilnehmenden aus Deutschland hatten 73 % ein dreiwöchiges Schulpraktikum, 45 % das Praxissemester und 50 % ein sonstiges Praktikum bereits abgeschlossen. Die 29 Teilnehmenden aus Österreich mit Praxiserfahrung gaben an, welche von vier pädagogisch-praktischen Studien (PPS) sie an der Universität Graz²⁶ bereits absolviert hatten (PPS1: 87 %, PPS2: 57 %, PPS3: 27 %, PPS4: 0 %).

Im Hinblick auf den Medieneinsatz in solchen schulpraktischen Erfahrungen gaben ca. 49 % aller Studierenden an, in manchen Stunden den Einsatz digitaler Medien im Unterricht beobachtet zu haben. Etwa 26 % gaben an, mindestens im Großteil aller Stunden einen Einsatz digitaler Medien beobachtet zu haben. In eigenen gehaltenen Unterrichtsstunden haben nach eigenen Angaben ca. 36 % höchstens in manchen Stunden digitale Medien verwendet, ca. 31 % bei der Hälfte der Stunden und 27 % im Großteil aller Stunden oder sogar allen Stunden.

Nachfolgend werden die Ergebnisse hinsichtlich der Testleistungen im FDW zum Einsatz digitaler Medien im Prä-Post-Vergleich dargestellt.

11.2 Ergebnisse zur Untersuchung des Wissenserwerbs im FDW zum Einsatz digitaler Medien

Zur Analyse der Testleistungen im untersuchten FDW wurde eine eindimensionale IRT-Modellierung auf Basis der 140 Testbearbeitungen (je 70 für Prä- und Posttest) vorgenommen, wobei auch hier *Thurstone Thresholds* für die Schätzung der Itemschwierigkeiten der mehrstufigen Testaufgaben verwen-

²⁶ Informationen zu den PPS an der Universität Graz sind z. B. ersichtlich unter: <https://physikdidaktik.uni-graz.at/de/studium/paedagogisch-praktische-studien-pps/> (letzter Zugriff am 13.09.2024)

det wurden (8.4). Die IRT-Modellierung zeigte eine annehmbare Modellpassung. Der MNSQ-Itemfit (Infit) lag für alle Aufgaben auch in dieser Stichprobe in einem guten Bereich ($0,96 < \text{MNSQ} < 1,06$). Mit einem mittleren RMSD von 0,032 (SD = 0,01) und einem SMSR von 0,08 zeigte sich auch ein angemessener Modellfit (Gäde et al., 2020). Die EAP-Reliabilität fiel für diese Stichprobe mit 0,47 recht gering aus.

Die Wright-Map zu dieser IRT-Modellierung ist in Abbildung 11-1 dargestellt.

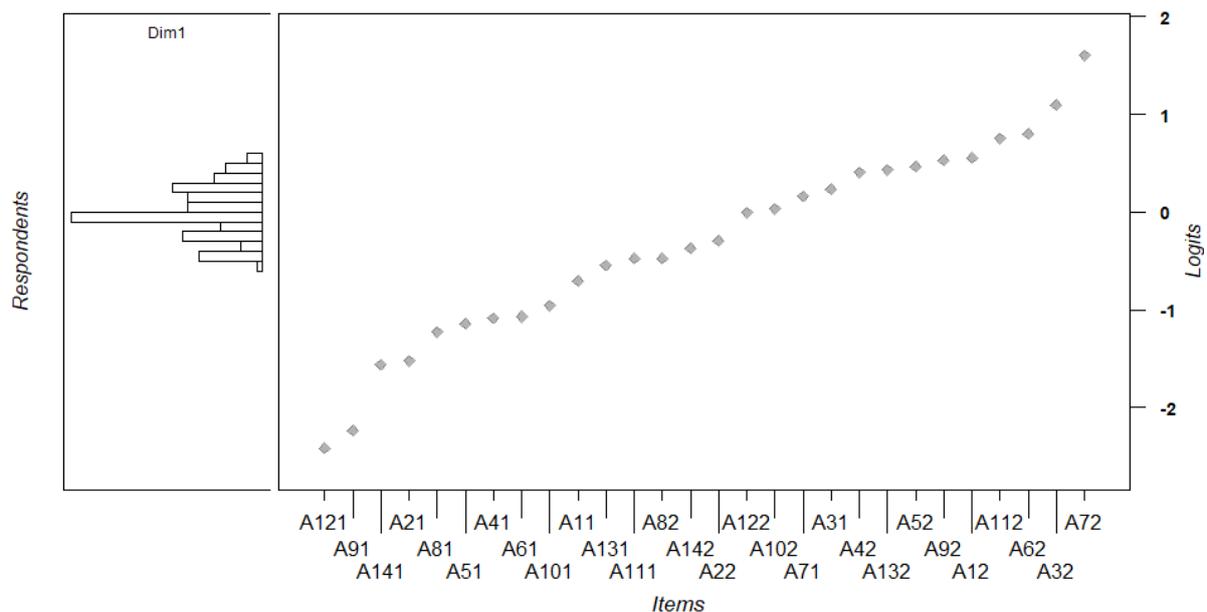


Abbildung 11-1: Wright-Map zur Stichprobe in der Haupterhebung

Beschriftung der Aufgabenschwellen: A11 entspricht dem Threshold-Wert $t_{1,1}$ für das Erreichen von einem Punkt in Aufgabe 1 und A12 dem Wert $t_{1,2}$ für das Erreichen von zwei Punkten in Aufgabe 1. Höhere Werte auf der Logit-Skala entsprechen schwieriger zu erreichenden Aufgabenschwellen sowie höheren Personenfähigkeiten.

Hier zeigt sich, dass der Bereich der Verteilung der Personenfähigkeiten auch durch einige Aufgaben(-schwellen) hinsichtlich ihrer Schwierigkeiten abgedeckt wird. Die Schwierigkeiten der Aufgabenschwellen verteilen sich insgesamt jedoch breiter auf der Logit-Skala.

Tabelle 11-2 zeigt als Ergebnisse die mittleren Personenscores und Personenfähigkeiten der Testteilnehmenden jeweils für Prä- und Posttest. Als Schätzer für Personenfähigkeit und Reliabilität wurde die EAP-Schätzung (*Expected-a-posteriori*; 3.3.1) verwendet.

Tabelle 11-2: Ergebnisse in Prä- und Posttest im FDW zum Einsatz digitaler Medien

| MZP | mittlerer Personenscore (Max. 28) | mittlere Personenfähigkeit |
|------|-----------------------------------|----------------------------|
| Prä | 15,6 (SD = 3,4) | -0,03 (SD = 0,24) |
| Post | 16,4 (SD = 3,7) | 0,03 (SD = 0,27) |

MZP: Messzeitpunkt; Max: Maximum; SD: Standardabweichung

Zum Vergleich der mittleren Personenfähigkeiten im Prä- und Posttest wurde ein t-Test für gepaarte Stichproben (Bortz & Schuster, 2010) durchgeführt. Dieser zeigt einen signifikanten Zuwachs der mittleren Personenfähigkeiten über den Untersuchungszeitraum ($t(69) = 2,14$; $p = 0,036$; $d = 0,26$). Nach Cohen (1988) handelt es sich dabei um einen kleinen Effekt.

Im Hinblick auf FF2 zeigt sich somit wie erwartet eine positive Veränderung im gemessenen FDW zum Einsatz digitaler Medien über die untersuchten Seminare, sodass die Wirkungshypothese H1 angenommen werden kann. Das zugrundeliegende Lehrkonzept im Verbundprojekt scheint demnach geeignet, um das FDW zum Einsatz digitaler Medien bei Physiklehramtsstudierenden zu fördern. Der gemessene Zuwachs in der untersuchten Stichprobe stellt jedoch nur einen kleinen Effekt dar, obwohl eine inhaltliche Passung zwischen den abgestimmten Kerninhalten der Seminare und dem im Test fokussierten FDW vorliegt. Diese unerwartet geringe Effektstärke (trotz vorliegender Passung) soll mithilfe der qualitativen Interviews weiter aufgeklärt werden. Zudem wird dieses Ergebnis zu FF2 im Fazit dieser Arbeit weiter diskutiert (14.1).

In den nachfolgenden Kapiteln 12 und 13 wird als qualitativer Teil dieser Arbeit die retrospektive Interviewstudie erläutert sowie deren Ergebnisse beschrieben. Die bisherigen Erkenntnisse aus der quantitativen Erhebung zur Entwicklung des FDW der Studierenden können mithilfe der qualitativen Interviewstudie ergänzt und vertieft sowie angemessener interpretiert werden.

Qualitativer Teil

12 Gestaltung der retrospektiven Interviews und Vorgehen zur Auswertung

Im Rahmen des vierten Ziels dieser Arbeit zur Erkundung von Gründen für veränderte Testantworten von Prä- zu Posttest sowie zur anknüpfenden Identifikation von lernförderlichen und lernhinderlichen Seminarelementen für die Beantwortung von FF3 (6.2), wurden 19 qualitative retrospektive Interviews mit Seminarteilnehmenden durchgeführt (7.2). Dieses Kapitel stellt dar, wie diese Interviews gestaltet und ausgewertet wurden.

In 12.1 werden dafür zunächst die Stichprobe der Interviewstudie sowie weitere Eckdaten zu den Interviews dargestellt. Zur Verdeutlichung des Ablaufs der Interviews erfolgt in 12.2 die Beschreibung des entwickelten Leitfadens sowie im Interview genutzter Zusatzmaterialien. Das Kategoriensystem zur Auswertung der Interviews im Hinblick auf Gründe für Veränderungen in den Testantworten wird in 12.3 erläutert.

12.1 Stichprobe der qualitativen Interviewstudie

Die retrospektiven Interviews wurden in drei Durchgängen im Wintersemester 2021/22, Sommersemester 2022 und Wintersemester 2022/23 an den Standorten Aachen, Graz und Tübingen durchgeführt. Insgesamt wurden $N_2 = 19$ Studierende befragt, welche zu den $N_1 = 70$ Studierenden der übergeordneten Stichprobe zur quantitativen Prä-Post-Erhebung zählten. Diese 19 Studierenden sind alle diejenigen der 70 Studienteilnehmenden, die sich freiwillig für die Teilnahme an einem Interview bereit erklärten. In Tabelle 12-1 sind Merkmale der Stichprobe zu dieser Interviewstudie sowie der übergeordneten Stichprobe (11.1) dargestellt. Um die Repräsentativität der Interviewstichprobe einschätzen zu können, wird diese im Folgenden mit der übergeordneten Stichprobe verglichen. Dabei werden die Merkmale aus Tabelle 12-1 und weitere Angaben aus den demografischen Daten sowie darüber hinaus die Leistungen in den Prä-Post-Befragungen zum FDW aufgegriffen.

Es zeigen sich bei den Interviewteilnehmenden ähnliche Mittelwerte beim Alter und Fachsemester wie in der übergeordneten Stichprobe. Eine etwas andere Verteilung ergibt sich beim Geschlecht sowie bei der Ausbildungsphase, da an den Interviews überwiegend Masterstudierende teilnahmen, während in der schriftlichen Befragung die Verteilung zwischen Master- und Bachelorstudierenden eher ausgeglichen war (Tabelle 12-1).

Tabelle 12-1: Übersicht zur Stichprobe in den retrospektiven Interviews im Vergleich zur übergeordneten Stichprobe der quantitativen Befragung

| Stichprobe | Standorte | Geschlecht | Alter | Ausbildungsphase | Fachsemester |
|---------------------|----------------------------------|------------|--------------------|------------------|-------------------|
| N ₂ = 19 | 8 AC, 6 GR, 5 TÜ | 9 w, 10 m | M = 25 (SD = 4) | 5 BA 14 MA | M = 9 (SD = 2) |
| N ₁ = 70 | 10 AC, 30 GR, 19 TÜ, 11 PB | 26 w, 44 m | M = 24 (SD = 3) | 38 BA 32 MA | M = 8 (SD = 3) |

AC: Aachen; GR: Graz; TÜ: Tübingen; PB: Paderborn; w: weiblich; m: männlich; M: Mittelwert; SD: Standardabweichung; BA: Bachelor (Lehramt), MA: Master (Lehramt)

Auch hinsichtlich der mittleren Abiturnote (1,78, SD = 0,56; deutsche Studierende) bzw. der mittleren Mathematik-Maturanote (1,67; SD = 1,03; österreichische Studierende) ist die Interviewstichprobe mit der übergeordneten Stichprobe vergleichbar (mittlere Abiturnote: 1,82, SD = 0,54; mittlere Mathematik-Maturanote: 2,0; SD = 0,98; 11.1).

In Bezug auf die Schulerfahrungen und den dort beobachteten Medieneinsatz zeichnet sich ebenfalls ein ähnliches Bild bei beiden Stichproben ab: alle Interviewteilnehmenden gaben mindestens ein Praktikum als Schulerfahrung an (übergeordnete Stichprobe: 97 %) und ca. 41 % gaben an, zwei oder mehr Schulpraktika absolviert zu haben (übergeordnete Stichprobe: 51 %). Einen Einsatz digitaler Medien in manchen Stunden beobachteten 42 % (gegenüber 49 %), während 32 % (gegenüber 26 %) dies für mindestens den Großteil aller Stunden angaben. Eine leichte Verschiebung zeigt sich bei der Häufigkeit, im eigenen gehaltenen Unterricht digitale Medien eingesetzt zu haben: Von den Interviewteilnehmenden gaben etwa 42 % an, höchstens in manchen Stun-

den (übergeordnete Stichprobe: 36 %) und circa 37 % bei der Hälfte der Stunden (übergeordnete Stichprobe: 31 %) digitale Medien verwendet zu haben. Nur 11 % verwendeten digitale Medien im Großteil aller Stunden oder sogar allen Stunden, während dies in der übergeordneten Stichprobe noch 27 % angaben (11.1).

Betrachtet man die Leistungen im FDW zum Einsatz digitaler Medien im Fach Physik, so erreichten die Interviewteilnehmenden mit einem mittleren Summenscore von 16,8 (SD = 3,5) im Prätest und 18,0 (SD = 3,4) im Posttest etwas bessere Testleistungen als die Gesamtstichprobe mit mittleren Summenscores von 15,6 (SD = 3,4) im Prätest und 16,4 (SD = 3,7) im Posttest (11.2).

Die Stichprobe der qualitativen Interviewstudie stellt demnach eine gewisse Positivauswahl gegenüber der Gesamtstichprobe der quantitativen Befragungen in Bezug auf ihre Leistungen im gemessenen FDW in Prä- und Posttest dar. Insgesamt scheint die Stichprobe der Interviewteilnehmenden jedoch hinsichtlich ihrer Personenmerkmale und praktischen Erfahrungen vergleichbar mit der Gesamtstichprobe.

12.2 Interviewleitfaden, Zusatzmaterialien und Ablauf

An den Standorten Graz und Tübingen wurden die Interviews gemeinsam mit einem Projektpartner aus Tübingen geplant und geführt. Dabei wurden dort zusätzliche Fragen gestellt, die für die Forschungsarbeit am Standort Tübingen von Bedeutung waren. Im Folgenden wird sich für diese Interviews nur auf die Interviewfragen bezogen, die für die vorliegende Arbeit relevant sind. Die Interviews dauerten in Aachen zwischen 29 und 76 Minuten und in Graz und Tübingen zwischen 57 und 109 Minuten.

Die retrospektiven Interviews fanden alle per Videokonferenz mithilfe der Videokonferenzsoftware *Zoom* statt. Dabei wurde zur Audio-Aufzeichnung der Interviews die interne Aufzeichnen-Funktion von *Zoom* genutzt. Die freiwilligen Teilnehmenden wurden vorab per Mail über den Inhalt und Zweck des Interviews informiert und es wurde ein Einverständnis zur Aufzeichnung der Interviews und Nutzung der anonymisierten Transkripte zu Forschungszwecken eingeholt. Zudem war es für die individuelle Gestaltung der Interviews

zur Untersuchung von FF3 notwendig, dass die Interviewteilnehmenden ihren persönlichen Code der quantitativen Befragung angaben, damit im Interview auch die individuellen Veränderungen der Testantworten angesprochen werden konnten. Diese Zugehörigkeit von Person und persönlichem Code in der Befragung war allein der Autorin (und in Graz und Tübingen zusätzlich dem Projektpartner) als Interviewführende bekannt, wurde nicht an Dritte weitergegeben und ausschließlich zur Umsetzung der Interviews und Triangulation der quantitativen Testdaten und qualitativen Interviewdaten genutzt. Auch hierzu wurden die Teilnehmenden vorab informiert und das Einverständnis eingeholt (Anhang E).

Interviewablauf anhand des Leitfadens

Der für die vorliegende Arbeit erstellte Leitfaden besteht in seiner Grobstruktur aus einer Einstiegs- und Ausstiegsphase zur Schaffung einer angenehmen Gesprächsatmosphäre sowie einem Hauptteil aus drei Phasen A-C entsprechend den Untersuchungszielen (Tabelle 12-2). An einigen Stellen wurde im Interview zudem auf Begleitmaterialien zurückgegriffen, die den Befragten als Stimulus dienen sollten. Im Folgenden wird der Ablauf des Interviews sowie die genutzten Materialien für die einzelnen Phasen beschrieben. Der gesamte Interviewleitfaden sowie das Begleitmaterial finden sich in Anhang E.

Tabelle 12-2: Grobstruktur des Interviewleitfadens zur Untersuchung von FF3

| Phase | Inhalte | Material |
|--------------|--|--|
| Einstieg | Erfahrung bei der Fragebogenbearbeitung | |
| Hauptphase A | Einschätzungen zum Seminar (Grobstruktur, Theorie-Praxis-Aufteilung, Lernförderlichkeit einzelner Seminarelemente) | Kurzumfrage zur empfundenen Lernförderlichkeit |
| Hauptphase B | Individuelle Vergleiche der Testantworten in Prä- und Posttest (Gründe für Veränderungen in den Testantworten) | Testantworten der Probanden im Prä- und Posttest |

| | |
|--------------|---|
| Hauptphase C | Berücksichtigung individueller Besonderheiten (Lerngelegenheiten im Zweitfach, in den Bildungswissenschaften, durch einen Nebenjob etc.) |
| Ausstieg | Wünsche der Studierenden für ein Seminar |

Vor Beginn der eigentlichen Interviews wurden die Teilnehmenden begrüßt, ihnen für die Teilnahme gedankt und relevante Informationen und Hinweise zum Ablauf des Interviews vorgetragen. Nach Beginn der Aufzeichnung startete das Interview im **Einstieg** mit einer Frage zur schriftlichen Befragung im besuchten Seminar, wie es den Teilnehmenden beim (ersten und zweiten) Durchführen des Tests erging.

Anknüpfend ging es in der **ersten Hauptphase A** um das Seminar als Lerngelegenheit, zu dem die Einschätzung der Studierenden erfragt wurde. Als Material wurde dabei der Seminarplan vorgelegt (Anhang E) sowie auf die Antworten der Studierenden in einer Kurzumfrage (zweiseitige Tabelle) zurückgegriffen, die die Studierenden bereits vor dem Interviewtermin ausgefüllt hatten. Diese Tabelle diente zur Einschätzung der subjektiv wahrgenommenen Lernförderlichkeit verschiedener Seminarinhalte und -tätigkeiten. Ein Auszug der leeren Tabelle zur Kurzumfrage ist exemplarisch in Abbildung 12-1 dargestellt (die vollständigen Kurzumfrage findet sich in Anhang E).

| Schätzen Sie ein, inwiefern Sie diese Inhalte/Tätigkeiten im Seminar beim Erwerb von Kompetenzen zum Einsatz digitaler Medien unterstützt haben. | Dieser Inhalt / diese Tätigkeit hat mich beim Kompetenzerwerb... | | | | | | |
|--|--|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|------------------------------|
| | sehr unterstützt | unterstützt | eher unterstützt | eher nicht unterstützt | nicht unterstützt | gar nicht unterstützt | Ich kann mich nicht erinnern |
| <i>Allgemein, einleitend oder übergreifend</i> | | | | | | | |
| Kennzeichen von Multimedia (Multimodalität, Multicodierung, Interaktivität) erkunden | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Designprinzipien zum Multimedialen Lernen und der Cognitive Load Theory | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Empirische Befunde zum Einsatz digitaler Medien im Unterricht | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Beispiele aus dem Physikunterricht zu Ebenen des SAMR-Modells | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Abgrenzung von digitalen Experimentierformaten (Simulation, IBE, AR-Experiment) | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| <i>Simulationen</i> | | | | | | | |
| Möglichkeiten von Simulationen und Einsatzszenarien für den Physikunterricht | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Empirische Befunde zum Einsatz von Simulationen im Unterricht | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Ausprobieren / Vorstellung einer Simulation im Seminar & Diskussion | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Abbildung 12-1: Ausschnitt aus der Tabelle zur Einschätzung der subjektiven Lernförderlichkeit der Seminarinhalte und -tätigkeiten

Diese vorausgehende schriftliche Kurzumfrage ermöglichte es, einige Fragen aus dem Interview auszulagern (wie es z. B. im Schritt Prüfen des SPSS-Prinzips vorgeschlagen wird; 7.2). Im Interview selbst konnte sich somit auf das Erfragen von Begründungen zu ausgewählten Antworten in der Kurzumfrage beschränkt werden. An dieser Stelle verdeutlicht sich bereits, dass es für die hier durchgeführten Interviews nicht *den einen* Interviewleitfaden gab, sondern dieser für jedes Interview individuell angepasst wurde. Die Analyse dieser ersten Hauptphase des Interviews im Hinblick auf die subjektiv wahrgenommene Lernförderlichkeit steht nicht im Fokus der vorliegenden Arbeit. Stattdessen erfolgte diese im Rahmen einer Bachelorarbeit an der Universität Paderborn (Veltum, 2023), zu der ausgewählte Ergebnisse zu einem späteren Zeitpunkt dargestellt werden (13.2.3).

Die **zweite Hauptphase B** war für die Untersuchung von FF3 am bedeutendsten und stellte auch den längsten Teil der Interviews dar. Der Ausschnitt aus dem Interviewleitfaden zu dieser Phase ist in Abbildung 12-2 dargestellt.

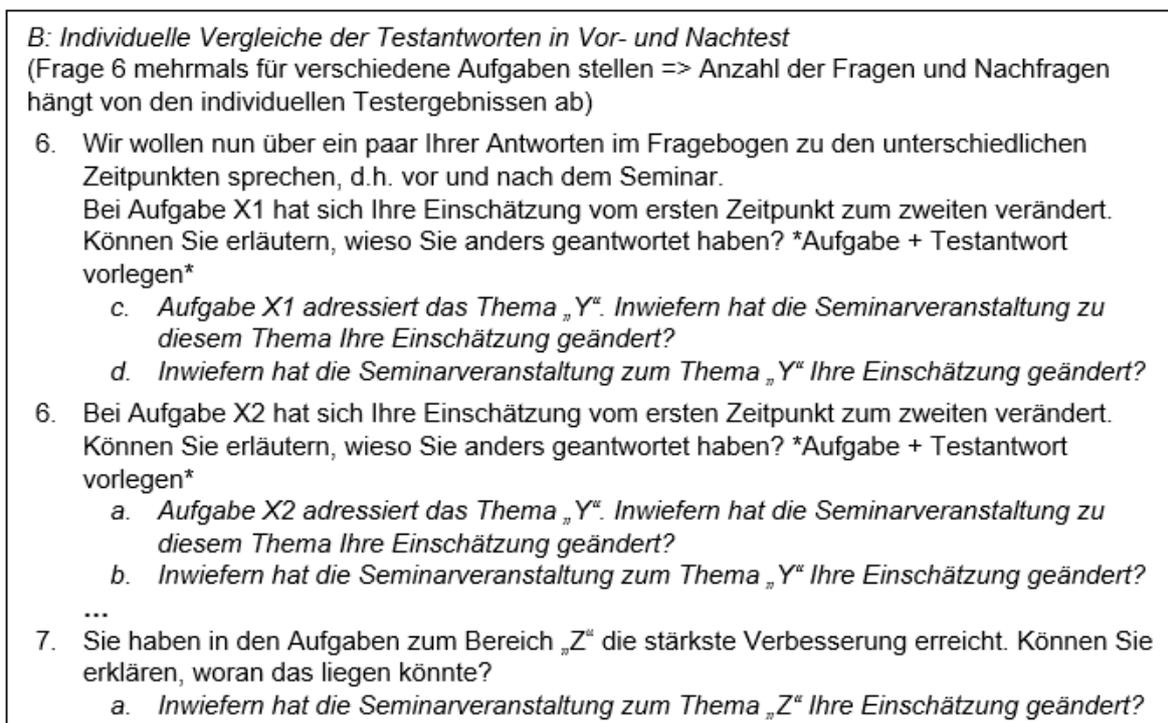


Abbildung 12-2: Ausschnitt aus dem Interviewleitfaden zur Hauptphase B (individuelle Vergleiche der Testantworten in Prä- und Posttest)

Hier wurden den Studierenden ihre beantworteten Testaufgaben in Prä- und Posttest (d. h. vor und nach dem Seminar) als Stimulus vorgelegt. Die Teilnehmenden sollten daraufhin begründen, warum sie die Einschätzung zu einzelnen Antwortalternativen nach dem Seminar verändert haben. Diese Aufforderung bzw. Frage bekamen alle Interviewteilnehmenden für verschiedene Aufgaben wiederholt gestellt (Abbildung 12-2). Pro Befragungsperson wurden dazu in Vorbereitung auf das Interview etwa fünf bis zehn beantwortete Testaufgaben ausgewählt (z. B. bei starken Veränderungen in der erreichten Punktzahl oder weiteren interessanten Veränderungen).

Als Beispiel zeigt Abbildung 12-3 eine beantwortete Testaufgabe, wie sie im Interview einer Befragungsperson vorgelegt wurde. Die veränderten Einzelantworten von Prätest (vorher) zu Posttest (nachher) wurden markiert und im Interview zu diesen Stellen eine Begründung für die veränderte Antwort erfragt. Dabei wurde im Interview nicht thematisiert, welche Antworten gemäß der Musterlösung richtig oder falsch sind.

1. Fachdidaktisches Wissen zum Einsatz digitaler Medien – Teil 8 von 14

Handelt es sich im Folgenden um Aspekte, die generell für einen Einsatz von Simulationen im Physikunterricht sprechen?

| | Nein | Ja | Nein | Ja |
|--|----------------------------------|----------------------------------|-----------------------|----------------------------------|
| Simulationen bieten eine präzise Abbildung der Wirklichkeit. | <input checked="" type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> |
| Simulationen können nicht direkt sichtbare physikalische Sachverhalte visualisieren. | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> |
| Simulationen können als Beispiel eines wissenschaftlichen Modells thematisiert werden. | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> |
| Simulationen ermöglichen das Abdecken von Lernzielen durch die Strukturierung des Lernprozesses. | <input checked="" type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> |
| Simulationen bereichern den Lernprozess durch fachbezogene Übungsaufgaben an. | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> |
| Simulationen unterstützen die Modellbildung durch das Ausblenden weniger relevanter Aspekte. | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> |
| | vorher | | nachher | |

Abbildung 12-3: Beispiel zu einer im Interview vorgelegten Testaufgabe mit Veränderungen von Prä- zu Posttest

Da im Interview das untersuchte Seminar fokussiert wurde, lag es nahe, dass Erfahrungen im Seminar teilweise als Ursachen für die Änderungen in den Testantworten genannt werden. Aus diesem Grund wurde in der **dritten Hauptphase C** abschließend abgegrenzt, inwiefern neben dem Seminar auch mögliche andere individuelle Lernerfahrungen im gleichen Zeitraum die Studierenden bei der Beantwortung im Posttest beeinflusst haben. Auf

diese Weise sollte vermieden werden, die Veränderungen in den Testantworten und im gemessenen FDW zum Einsatz digitaler Medien ausschließlich auf das Seminar zu beziehen, obwohl die Studierenden möglicherweise andere Lerngelegenheiten in diesem Bereich hatten.

Die letzte Phase zum **Ausstieg des Interviews** beinhaltete die Frage, welche Wünsche die Studierenden für ein Seminar zum Einsatz digitaler Medien im Physikunterricht haben. Abschließend erhielten sie zudem die Möglichkeit, Anmerkungen zum Thema zu ergänzen – sei es zum Interview, Seminar oder beantworteten FDW-Test. Danach wurde das Interview sowie die Audio-Aufzeichnung beendet.

Das für die Analyse der Interviews entwickelte Kategoriensystem wird im nachfolgenden Abschnitt beschrieben.

12.3 Kategoriensystem zur Interviewauswertung

Wie in 7.2 angeführt, erfolgte die Auswertung der transkribierten Interviews durch eine inhaltlich-strukturierende qualitative Inhaltsanalyse (Kuckartz, 2018) mit deduktiv-induktiver Kategorienbildung. Relevante Faktoren zur Vorgehensweise der Interviewauswertung (z. B. Kodierregeln) sind als Übersicht in Anhang F aufgeführt.

Das übergeordnete Kategoriensystem in Bezug auf das gesamte Interview beinhaltet dabei sieben Hauptkategorien. Diese sieben Hauptkategorien dienten einer ersten groben Kategorisierung von verschiedenen Aspekten, die im gesamten Interview auftreten konnten und während der drei Hauptphasen des Interviewleitfadens angesprochen wurden. Tabelle 12-3 zeigt eine Auflistung der Hauptkategorien und ihrer kurzen inhaltlichen Beschreibung.

In Tabelle 12-3 sind die Hauptkategorien *Testbewertung* und *Änderungen in Testantworten* hervorgehoben, da diese für die individuellen Vergleiche der Testantworten relevant waren und somit für die Untersuchung von FF3 in der vorliegenden Arbeit am bedeutsamsten sind. Im weiteren Verlauf werden diese Kategorien daher fokussiert und dazu entwickelte Subkategorien beschrieben. Die weiteren Hauptkategorien werden nachfolgend kurz vorgestellt; diese werden in der vorliegenden Arbeit jedoch nicht vertieft betrachtet oder weiter differenziert.

Tabelle 12-3: Übersicht zu den Hauptkategorien zur Interviewauswertung und ihrer inhaltlichen Beschreibung

| Hauptkategorie | Inhaltliche Beschreibung |
|--|---|
| Testbewertung | Aussagen, die die Veränderungen von einzelnen Antworten im Prätest zu Antworten im Posttest aufzeigen. |
| Änderungen in Testantworten | Aussagen, die Veränderungen von Antworten im Prätest zu Antworten im Posttest begründen (oder beschreiben). |
| Seminar als Lerngelegenheit | Aussagen, die einen konkreten oder indirekten Bezug zum untersuchten Seminar "Lehr-Lern-Gelegenheit im Physikunterricht" (Aachen) bzw. "Moderne Medien im Physikunterricht" (Graz, Tübingen) aufzeigen. |
| Subjektiv wahrgenommene Lernförderlichkeit | Aussagen, die in Bezug auf das untersuchte Seminar eine Bewertung der (wahrgenommenen) Lernförderlichkeit vornehmen (aus Sicht der Studierenden). |
| Andere Lerngelegenheiten | Aussagen, die einen konkreten oder indirekten Bezug zu anderen Lerngelegenheiten neben dem untersuchten Seminar aufzeigen. |
| Zukünftiger Unterrichtsalltag/ Lehrberuf | Aussagen, die einen konkreten Bezug zum zukünftigen Unterrichten als Lehrkraft aufzeigen. |
| Testinstrument/ Testaufgaben | Aussagen, die einen direkten Bezug zur Beantwortung des Tests oder den Testaufgaben zum FDW digitale Medien aufzeigen (im allgemeinen Sinne). |

Hervorgehoben sind die beiden Hauptkategorien, die zur Untersuchung von FF3 in dieser Arbeit fokussiert werden.

Die beiden Hauptkategorien *Seminar als Lerngelegenheit* und *subjektiv wahrgenommene Lernförderlichkeit* waren für die Einschätzungen zum Seminar relevant, um Textstellen mit Bezug zu Seminarinhalten sowie mit Bewertungen hinsichtlich der subjektiv wahrgenommenen Lernförderlichkeit

solcher Inhalte kodieren zu können. Eine weitere Differenzierung dieser Kategorien stand nicht im Fokus der vorliegenden Arbeit. Stattdessen erfolgte die Interviewanalyse dahingehend im Rahmen einer Abschlussarbeit zur wahrgenommenen Lernwirksamkeit der Lerngelegenheiten mithilfe eines differenzierteren Kategoriensystems (Veltum, 2023; 13.2.3). Die Hauptkategorie *andere Lerngelegenheiten* diente dazu, Bezüge zu anderen Lerngelegenheiten als den untersuchten Seminaren zu identifizieren. Eine weitere Differenzierung dieser Kategorie wurde nicht vorgenommen, da die zugehörigen Textstellen dies nicht nahelegten und für die Ziele der Interviewstudie ein einfacher Bezug zur Hauptkategorie ausreichend war. Schließlich berücksichtigen die Hauptkategorien *zukünftiger Unterrichtsalltag/Lehrberuf* und *Testinstrument/Testaufgaben* Textstellen, in denen Studierende z. B. auf die Relevanz verschiedener Aspekte aus dem Seminar für ihren späteren Beruf als Lehrkraft hindeuteten oder allgemeine Anmerkungen zu den Testaufgaben oder der Testbearbeitung äußerten. Auch hier genügte für die Interviewauswertung der vorliegenden Arbeit der einfache Bezug zu diesen Kategorien ohne eine weitere Differenzierung in Subkategorien.

Eine Übersicht zu den sieben Hauptkategorien, ihren inhaltlichen Beschreibungen sowie Anwendungsregeln findet sich in Anhang F. Für die weitere Auswertung der Interviews im Hinblick auf FF3 zur Identifikation lernförderlicher und -hinderlicher Elemente werden nun die beiden Hauptkategorien *Testbewertung* und *Änderungen in Testantworten* und die dazu entwickelten Subkategorien vertieft betrachtet. Dazu wird zunächst das Vorgehen zur Entwicklung dieser Kategorien dargestellt; anschließend werden die finalen Haupt- und Subkategorien beschrieben.

12.3.1 Vorgehen zur Entwicklung der Hauptkategorien und Subkategorien zur Untersuchung von FF3

Die fokussierten Hauptkategorien *Testbewertung* und *Änderungen in Testantworten* wurden beide in Hauptphase B des Interviews kodiert (Abbildung 12-2), in der die Teilnehmenden ihre veränderten Testantworten begründeten. Die erste Hauptkategorie *Testbewertung* diente dabei zur Triangulation der Interviewanalyse mit den quantitativen Daten der Prä-Post-Erhebung.

Sie beinhaltet nur zwei deduktive Subkategorien (*Verbesserung* und *Verschlechterung*) und ist mithilfe der Musterlösung zum Testinstrument trivial zu kodieren (12.3.2). Die Subkategorien zur zweiten Hauptkategorie *Änderungen in Testantworten* wurden einerseits induktiv am Material gebildet; andererseits ergaben sich durch Vorüberlegungen zur Studie auch hier deduktive Subkategorien (12.3.2).

Das Interviewmaterial der ersten beiden Durchgänge im Wintersemester 2021/22 und Sommersemester 2022 (elf Interviews) wurde zunächst von der Autorin kodiert. Dabei wurden insbesondere zu Beginn, aber auch fortlaufend induktive Subkategorien (zur Hauptkategorie *Änderungen in Testantworten*) am Material entwickelt und das Kategoriensystem angepasst.

Um die Qualität des (vorläufigen) Kategoriensystems nach der Kodierung dieser ersten elf Interviews zu überprüfen, wurde ein Interrating mit einer zweiten kodierenden Person vorgenommen (studentische Mitarbeiterin in der physikdidaktischen Arbeitsgruppe an der RWTH Aachen). Als Maß für die Übereinstimmung zwischen zwei auswertenden Personen diente hierbei die Interrater-Reliabilität, welche ein spezifisch inhaltsanalytisches Gütekriterium darstellt (z. B. Kuckartz, 2018; Mayring, 2015; 7.2). Dazu wurde die Funktion der Intercoder-Übereinstimmung in MAXQDA verwendet und als Übereinstimmungsart die Codeüberlappung an Segmenten gewählt. In dieser Variante wird verglichen, ob die Kodierungen einzelner Segmente (zu einem Mindest-Prozentsatz) übereinstimmen. So können im Sinne eines Toleranzbereichs auch Fälle eingeschlossen werden, in denen nicht absolut identisch mit einer Codeüberlappung von 100 % kodiert wurde, wenn z. B. wenige Worte von einer Person mehr oder weniger kodiert wurden. Rädiker und Kuckartz (2019) schlagen vor, mit einer Einstellung von 95 % zu beginnen und diesen Wert schrittweise zu erniedrigen, wenn viele fehlerhafte Nicht-Übereinstimmungen gemeldet werden. Für die vorliegende Arbeit hat sich nach diesem Verfahren ein Prozentsatz von 85 % als sinnvolle Einstellung für das Interrating ergeben.

Die Ergebnisse zum Interrating für die Hauptkategorie *Änderung in Testantworten* werden nachfolgend erläutert; da die Hauptkategorie *Testbewertung* mithilfe der Musterlösung zum FDW-Test trivial zu kodieren war, ergab das

zur Absicherung durchgeführte Interrating eine perfekte Übereinstimmung von 100 %.

Eine erste Doppelkodierung der ersten elf Interviews im Rahmen des Interratings (zur Hauptkategorie *Änderungen in Testantworten*) ergab eine prozentuale Übereinstimmung von 74 %. Im Sinne eines konsensuellen Kodierens (7.2) wurden anschließend auftretende Nicht-Übereinstimmungen zwischen den beiden Kodierenden diskutiert, (systematische) Fehler behoben und Anpassungen am Kategoriensystem bzw. den Kategoriendefinitionen vorgenommen.

In einem anschließenden zweiten Interrating zu den ersten elf Interviews zeigte sich eine prozentuale Übereinstimmung von 84 %. Zudem wurde für dieses finale Interrating der zufallskorrigierte Koeffizient Kappa (κ) nach Brennan und Prediger (1981) mithilfe der Software MAXQDA bestimmt²⁷. Dabei ergab sich ein κ -Wert von $\kappa_n = 0,83$. Die Ergebnistabelle zur Berechnung des κ -Wertes aus der Anzahl der Übereinstimmungen und Nicht-Übereinstimmungen im zweiten Interrating ist in Anhang F dargestellt. Nach Kuckartz (2018) oder Rädiker und Kuckartz (2019) gelten κ -Werte zwischen 0,6 und 0,8 als gut und ab 0,8 als sehr gut, sodass hier im Anschluss an die Ausschärfung der Kategorien nach dem ersten Interrating von einer sehr guten Übereinstimmung gesprochen werden kann.

Nachdem durch dieses Verfahren somit ein angemessener Konsens zur Anwendung der Kategorien geschaffen werden konnte, wurden die übrigen acht Interviews aus dem letzten Durchgang im Wintersemester 2022/23 ausschließlich von der studentischen Hilfskraft kodiert. Dabei bestand stets die Möglichkeit, die Kodierungen bei Unsicherheiten mit der Autorin zu diskutieren und im Sinne einer konsensuellen Validierung (7.2) abzusichern.

Weitere Gütekriterien wie eine Regelgeleitetheit (7.2) wurden in der Kategorienbildung ebenfalls beachtet, indem beispielsweise das Vorgehen bei der

²⁷ Die Software MAXQDA greift auf den Vorschlag von Brennan und Prediger (1981) zur Berechnung der zufälligen Übereinstimmung zurück, weil diese sich intensiv mit den Einsatzmöglichkeiten des Koeffizienten Cohens Kappa und dessen Problemen bei ungleichen Randsummenverteilungen auseinandersetzen (VERBI GmbH, 2022). Die Formel für den Koeffizienten ist $\kappa_n = (P_o - P_c) / (1 - P_c)$, wobei P_o die prozentuale Übereinstimmung und P_c die zufällige Übereinstimmung darstellt, welche die Anzahl der Kategorien mit einbezieht.

Interviewanalyse die Schritte der inhaltlich-strukturierenden qualitativen Inhaltsanalyse berücksichtigte und die Kodierung systematisch anhand festgelegter Kodierregeln erfolgte. Durch die umfassende Beschreibung der gebildeten Kategorien (im Folgenden und in Anhang F) werden zudem die Kriterien der Verfahrensdokumentation und Nachvollziehbarkeit berücksichtigt (7.2).

12.3.2 Beschreibung der relevanten Hauptkategorien und Subkategorien zur Untersuchung von FF3

Nachfolgend werden die Hauptkategorien *Testbewertung* und *Änderungen in Testantworten* sowie die dazu gehörenden Subkategorien dargestellt.

Testbewertung

Die erste Hauptkategorie *Testbewertung* wurde für die Triangulation der Interviewanalyse mit den quantitativen Daten der Prä-Post-Erhebung gebildet. Dabei wurde für jedes Interviewsegment, in dem eine Veränderung einer einzelnen Testantwort (Antwortalternative) von Prä- zu Posttest begründet wird, jeweils kodiert, ob es sich bei dieser veränderten Antwort um eine Verbesserung oder Verschlechterung (im Sinne der Musterlösung) handelt. Die Kategorie ist demnach mithilfe der Musterlösung zum Testinstrument trivial zu kodieren und umfasst zwei Subkategorien (*Verbesserung* und *Verschlechterung*).

Änderungen in Testantworten

Die zweite Hauptkategorie *Änderungen in Testantworten* beinhaltet alle Aussagen, die mögliche Veränderungen von Antworten im Prätest zu Antworten im Posttest beschreiben oder begründen (Tabelle 12-3). Diese Hauptkategorie enthält mehrere Subkategorien und z. T. weitere Unterkategorien, welche im Folgenden vorgestellt werden. Eine hierarchische Übersicht zu allen Subkategorien zeigt Abbildung 12-4.

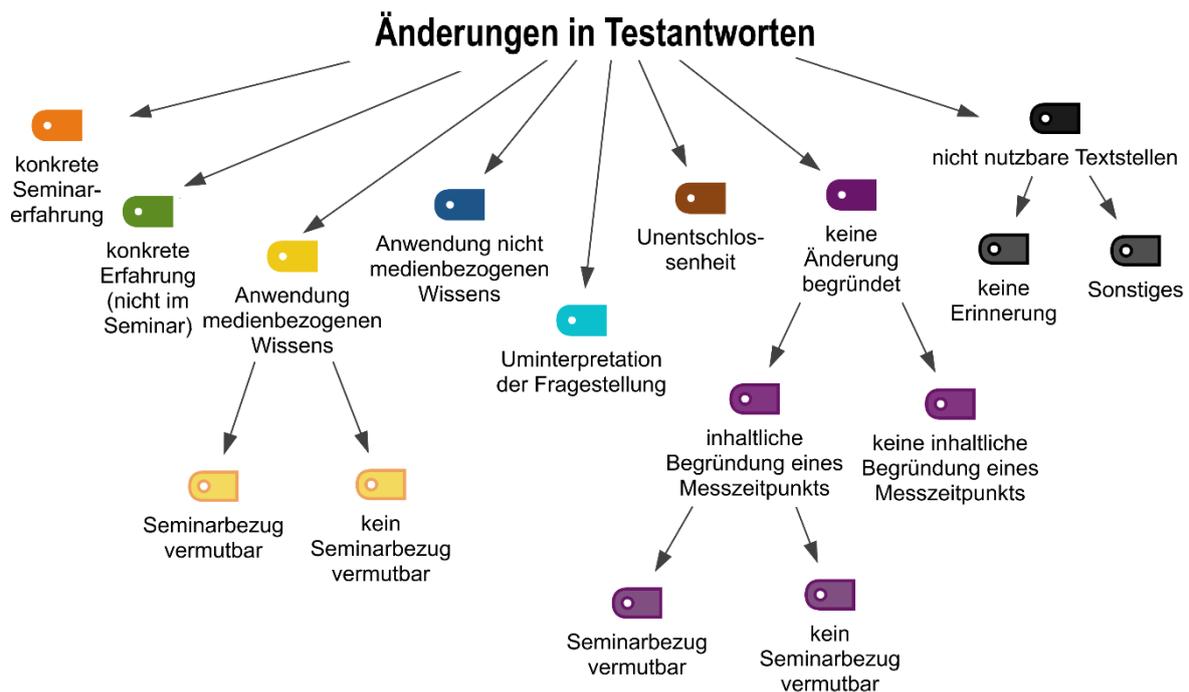


Abbildung 12-4: Übersicht der entwickelten Subkategorien zur Hauptkategorie *Änderungen in Testantworten*

Zur Subkategorie *konkrete Seminarerfahrung* zählten dabei Begründungen zu den veränderten Testantworten, in denen direkt Bezug zum untersuchten Seminar genommen wurde, also z. B. ein konkreter Inhalt oder eine Erfahrung aus dem Seminar als Begründung genannt wurde. Beispielsweise wurde das folgende Interviewzitat mit dieser Subkategorie kodiert, da hier ein Bezug zur praktischen Erprobung (Unterricht) im Rahmen der Lehrveranstaltung deutlich wird:

Zitat 1: A21a02 (Pos. 61)

„Deswegen habe ich jetzt hier nein angekreuzt, beim anderen ja und vor allen Dingen dadurch, dass wir jetzt nochmal auch Simulationen eingesetzt haben im Unterricht, das hat quasi so den Eindruck nochmal verändert, was man im Unterricht machen kann, was nicht und wie lange Dinge dauern. Und ja, Dinge dauern länger als ich eigentlich angenommen habe.“

Ähnlich wurde die Subkategorie *konkrete Erfahrung (nicht im Seminar)* angewendet, wenn bei der Begründung einer Veränderung ein konkreter Inhalt oder eine konkrete Erfahrung aus einer anderen Lehrveranstaltung oder Lerngelegenheit, die nicht das untersuchte Seminar darstellt, genannt wird.

Die Anwendung der Kategorien *konkrete Seminarerfahrung* und *konkrete Erfahrung (nicht im Seminar)* erfolgte nur bei eindeutigen und konkreten Bezügen zu den Lerngelegenheiten.

Ansonsten wurde die Subkategorie *Anwendung medienbezogenen Wissens* kodiert, wenn kein konkreter Aspekt des untersuchten Seminars (→ *konkreter Seminarinhalt*) oder einer anderen Lehrveranstaltung (→ *konkrete Erfahrung (nicht im Seminar)*) als Begründung für das Verändern der Antwort genannt wurde, aber dennoch eine reflektierte Entscheidungsänderung unter Nutzung von Wissen zum Einsatz digitaler Medien erfolgte. Dabei konnte bei solchen Begründungen ein indirekter Seminarbezug naheliegend sein (kein konkreter Inhalt), musste es aber nicht. Dementsprechend wurden zu dieser Subkategorie zwei weitere Unterkategorien gebildet (*Seminarbezug vermutlich* und *kein Seminarbezug vermutlich*). Die Unterkategorie *Seminarbezug vermutlich* wurde dabei eher offensiv kodiert, sodass im Zweifel für den Bezug kodiert wurde, wobei bei Unsicherheiten – wie in 12.3.1 beschrieben – stets die Möglichkeit zur Diskussion und konsensuellen Abstimmung der Kodierungen zwischen der Autorin und der studentischen Hilfskraft bestand. Das nachfolgende Zitat 2 stellt ein Beispiel für die Subkategorie *Anwendung medienbezogenen Wissens* (mit *Seminarbezug vermutlich*) dar, da Aspekte genannt werden, die im Seminar besprochen wurden (technische und didaktische Qualität eines Erklärvideos), aber nicht das Seminar konkret benannt wird. Das Zitat bezieht sich auf die zweite Antwortoption in Aufgabe 13 (Abbildung 10-5):

Zitat 2: A21a02 (Pos. 50)

„Ich habe es jetzt beim zweiten Mal- und auch hier kann ich mich daran erinnern, weil ich überlegt habe, ob ich nein oder ja ankreuzen soll. Das ist das technische Potential des Erklärvideos insofern, als dass die Schüler das Video halt dann- ehm das ist halt einfach das technische Potential, nicht das didaktische Potential. Ich glaube beim ersten Mal habe ich das nicht getrennt und beim zweiten Mal schon.“

Die Subkategorie *Anwendung nicht-medienbezogenen Wissens* wurde hingegen kodiert, wenn die Veränderung zwar inhaltlich, nicht jedoch über medienbezogenes Wissen begründet wurde, sondern beispielsweise durch Anwendung von FDW ohne Bezug zu digitalen Medien.

Teilweise wurden die veränderten Antworten jedoch auch ohne inhaltliches Wissen oder inhaltlichen Bezug begründet. So beinhaltet die Subkategorie *Uminterpretation der Fragestellung* Begründungen zu veränderten Antworten, die eher auf das Verständnis oder die sprachliche Auffassung der Fragestellung oder Antwortoption zurückzuführen waren. Zu der Subkategorie *Unentschlossenheit* zählten weiterhin Aussagen, in denen die veränderte Antwort durch Unentschlossenheit, Raten oder Zufall begründet wurde.

Die Subkategorie *keine Änderung begründet* wurde induktiv zum Kategoriensystem ergänzt, da die Befragten in ihren Ausführungen auf die Frage, warum sie an der jeweiligen Stelle ihre Einschätzung von Prä- zu Posttest verändert haben, teilweise keine wirkliche Begründung zur Änderung gaben. Stattdessen begründeten sie häufig die gegebene Antwort zu einem Messzeitpunkt, nicht jedoch die Veränderung vom ersten zum zweiten Messzeitpunkt. Daher enthält die Subkategorie eine weitere Unterkategorie *inhaltliche Begründung eines Messzeitpunkts*, die in solchen Fällen vergeben wurde. Als Dummy-Kategorie dazu wurde die Unterkategorie *keine inhaltliche Begründung eines Messzeitpunkts* immer genau dann kodiert, wenn *inhaltliche Begründung eines Messzeitpunkts* nicht kodiert wurde. Im Falle einer inhaltlichen Begründung wurden zudem in einer weiteren Ebene die beiden Unterkategorien *Seminarbezug vermutlich* und *kein Seminarbezug vermutlich* ergänzt, um einen möglichen Seminarbezug zu kennzeichnen. Bei einem vermutetem Seminarbezug konnten dadurch ebenfalls relevante Hinweise für die Analyse von FF3 gewonnen werden, wenngleich hier strenggenommen keine Begründung der *Veränderung* der Testantwort erfolgte.

Zuletzt umfasst die Subkategorie *nicht nutzbare Textstellen* alle Aussagen zur Hauptkategorie *Änderungen in Testantworten*, die sich keiner anderen Subkategorie zuordnen lassen und für die weitere Analyse nicht nutzbar sind. So war dies der Fall, wenn die Befragungsperson beispielsweise einen Fehler in der eigenen Beantwortung feststellte oder keine Erinnerung an die Testantwort oder des Grunds dafür vorlag. Daher wurde hier eine Unterkategorie *keine Erinnerung* ergänzt. Die weitere Unterkategorie *Sonstiges* enthält als Dummy-Kategorie alle Textstellen, die nicht zu *keine Erinnerung* zählen.

Die Kategoriendefinitionen aller Subkategorien gemäß des Schemas nach Kuckartz (2018; Tabelle 7-1) befinden sich in Anhang F. Im folgenden Kapitel wird beschrieben, wie die Kodierungen der vorgestellten (Sub-)Kategorien zur Erkundung von Gründen für die Veränderungen im gemessenen FDW sowie zur Identifikation lernförderlicher oder -hinderlicher Seminarelemente genutzt wurden.

13 Ergebnisse der Interviewstudie

Dieses Kapitel stellt die Ergebnisse zur qualitativen Inhaltsanalyse der retrospektiven Interviews vor. Dazu werden in 13.1 zunächst die aus der Inhaltsanalyse identifizierten Gründe für veränderte Testantworten sowie deren Häufigkeiten dargestellt. In 13.2 wird im Hinblick auf FF3 beschrieben, inwiefern die Veränderungen mit Bezug zum Seminar mit Verbesserungen oder Verschlechterungen im Test zusammenhängen, um Indizien für lernförderliche und -hinderliche Seminarelemente zu gewinnen. Abschließend erfolgt aus diesen Erkenntnissen die Ableitung von Hypothesen zur Gestaltung entsprechender Lerngelegenheiten zu digitalen Medien im Physikunterricht (13.3).

13.1 Gründe für die Veränderungen in den Testantworten zum FDW zum Einsatz digitaler Medien

Im Rahmen des vierten Ziels dieser Arbeit zur Erkundung von Gründen für die Veränderungen im gemessenen FDW zum Einsatz digitaler Medien (6.2) wurde in der qualitativen Inhaltsanalyse die Hauptkategorie *Änderungen in Testantworten* fokussiert. Aus der Analyse von 19 Interviews ergaben sich insgesamt 314 kodierte Segmente zu dieser Kategorie. Abbildung 13-1 stellt dar, wie viele Segmente davon mit den in 12.3 beschriebenen Subkategorien kodiert wurden. Dabei kann es vorkommen, dass ein Segment mehreren Subkategorien zugeordnet wurde, wenn z. B. in der Begründung der Änderung durch die Befragungsperson mehrere mögliche Gründe deutlich wurden.

Aus der Übersicht in Abbildung 13-1 wird ersichtlich, dass in den meisten Fällen ein konkreter Seminarbezug bei der Begründung veränderter Antworten erfolgte (107 kodierte Segmente). Weiterhin wurde häufig medienbezogenes Wissen zur Begründung herangezogen, wobei sich teilweise auch dort ein Bezug zum Seminar vermuten lässt (45 kodierte Segmente). Eine ebenfalls häufig kodierte Kategorie ist *keine Änderung begründet*, die Aussagen enthält, in denen nicht die Veränderung der Testantwort begründet wurde, sondern häufig nur eine Testantwort – i. d. R. die Posttest-Antwort (Kategorie

inhaltliche Begründung eines Messzeitpunkts, 46 kodierte Segmente). Wenn gleich hier strenggenommen keine Begründung der Veränderung von Prä- zu Post-Antwort vorliegt, können nichtsdestotrotz relevante Hinweise für die Analyse entnommen werden, da auch hier z. T. ein indirekter Seminarbezug angenommen werden kann (16 kodierte Segmente).

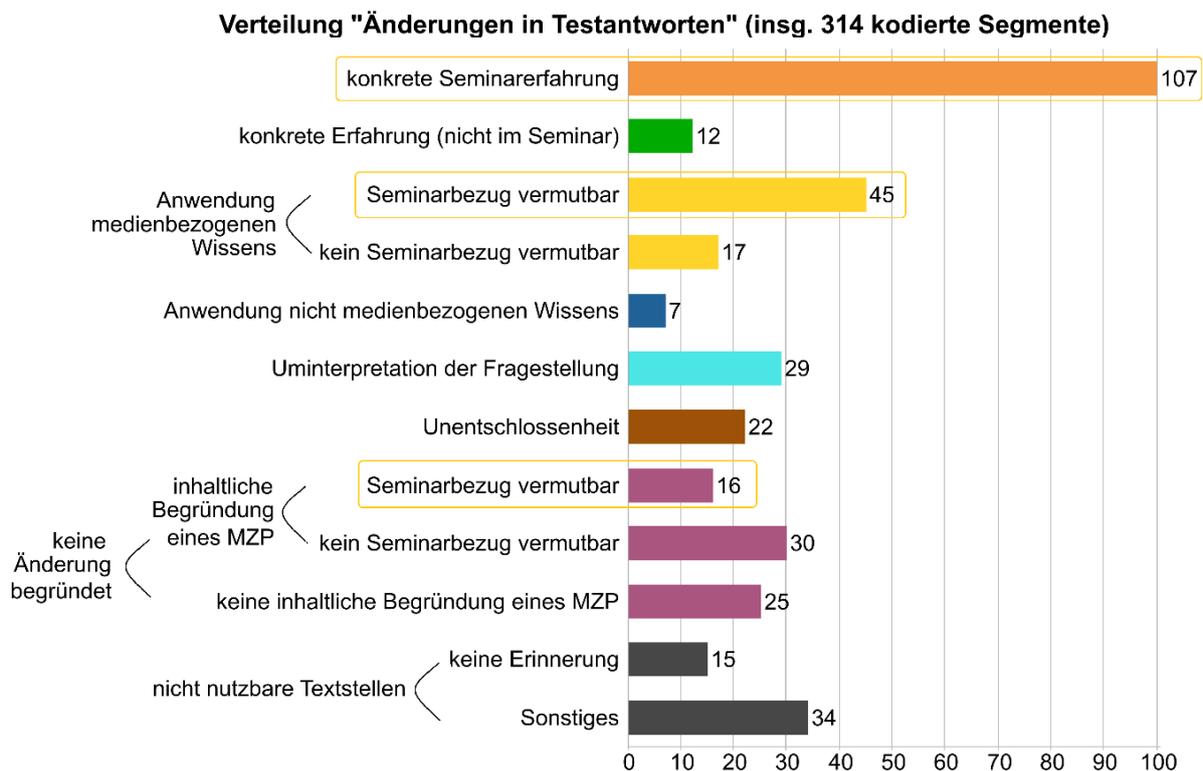


Abbildung 13-1: Übersicht zu identifizierten Gründen für Veränderungen in den Testantworten im FDW zum Einsatz digitaler Medien im Fach Physik (hervorgehoben sind die Subkategorien zu direktem oder vermutetem Seminarbezug, da diese für FF3 bedeutsam sind)

Seltener wurden veränderte Antworten durch andere Lerngelegenheiten oder Lehrveranstaltungen (12 kodierte Segmente) oder unter Nutzung von (fachdidaktischem) Wissen ohne Medienbezug (7 kodierte Segmente) begründet. Etwas häufiger, aber insgesamt auch vergleichsweise selten zeigten sich eine *Uminterpretation der Fragestellung* (29 kodierte Segmente) oder *Unentschlossenheit* hinsichtlich der Antwortoptionen (22 kodierte Segmente) als Gründe für Änderungen in den Antworten. Die Kategorie *nicht nutzbare Textstellen* tritt mit 49 kodierten Segmenten wiederum etwas häu-

figer auf (Abbildung 13-1). Sie enthält jedoch Aussagen, die nicht für eine Interpretation oder weitere Analyse nutzbar sind, da meist keine wirkliche Änderung im Wissen bzw. der Testantwort (z. B., wenn die Testperson einen Fehler in der Beantwortung zu einem Messzeitpunkt äußert) oder keine Erinnerung mehr an den Grund für die einzelnen Antworten oder die Veränderung vorliegt.

Insgesamt ist somit erkennbar, dass ein Großteil der veränderten Antworten im Test zum FDW zum Einsatz digitaler Medien im Fach Physik inhaltlich begründet wurde und dabei auch häufig mit konkretem oder indirektem Bezug zu den untersuchten Seminaren (bei insgesamt 168 von 314 kodierten Segmenten bzw. 168 von 265 inhaltlich nutzbaren kodierten Segmenten; in Abbildung 13-1 hervorgehoben).

13.2 Identifizierung lernförderlicher und lernhinderlicher Elemente des Lehrkonzepts (FF3)

Auch wenn Abbildung 13-1 verdeutlicht, dass ein Großteil der Veränderungen in den Testantworten mit (vermutetem) Bezug zum untersuchten Seminar begründet wurde, ist dadurch noch nicht ersichtlich, ob diese Veränderungen auch zu einer Verbesserung oder sogar zu einer Verschlechterung im gemessenen FDW zum Einsatz digitaler Medien beigetragen haben. Um ebensolche Hinweise zur Beantwortung von FF3 zu erhalten, welche Elemente des Lehrkonzepts lernförderlich oder lernhinderlich für den Erwerb von FDW zum Einsatz digitaler Medien sind, wurden Kombinationen bzw. Überschneidungen mehrerer Kodierungen aus den beiden Hauptkategorien *Änderungen in Testantworten* und *Testbewertung* (12.3.2) betrachtet.

So diente die Kombination der Kodierung *Änderungen in Testantworten*\ *konkrete Seminarerfahrung* mit *Testbewertung*\ *Verbesserung* dazu, Indizien für lernförderliche Seminarelemente zu erhalten. Umgekehrt wurde die Kombination mit der Kodierung *Testbewertung*\ *Verschlechterung* betrachtet, um Indizien für lernhinderliche Seminarelemente sowie für mögliche „typische Fehler“ oder Probleme in der Argumentation der Studierenden zu gewinnen. Das Verfahren wurde zudem erweitert auf Kombinationen der beiden Subkategorien in *Testbewertung* mit den Kodierungen zu vermutetem Seminarbe-

zug (Kodierungen *Änderungen in Testantworten*\Anwendung medienbezogenen Wissens\Seminarbezug vermutbar und *Änderungen in Testantworten*\keine Änderung begründet\inhaltliche Begründung eines Messzeitpunkts\Seminarbezug vermutbar). Da hier keine expliziten Nennungen des Seminars vorliegen, sondern der Bezug implizit vermutet wurde, sind daraus gewonnene Indizien weniger belastbar und sollten entsprechend vorsichtiger betrachtet werden. Diese drei Subkategorien zum direkten bzw. vermuteten Seminarbezug sind in Abbildung 13-1 hervorgehoben.

Die Kombinationen konnten jeweils über das Analysetool MAXQDA (Funktion „Komplexe Segmentsuche → Überschneidung“) ausgegeben werden sowie als tabellarische Übersicht zu den kodierten Segmenten dieser Kombinationen in *Excel* exportiert werden (jeweils für *Verbesserung* und *Verschlechterung*). Die Anzahlen der Überschneidungen für die beschriebenen Kombinationen sind in Tabelle 13-1 dargestellt.

Tabelle 13-1: Anzahl der Interviewsegmente, die Überschneidungen zwischen den Subkategorien zur Testbewertung und Subkategorien mit (vermutetem) Seminarbezug darstellen

| | Verbesserung (insg. 175) | Verschlechterung (insg. 143) |
|---|-----------------------------|---------------------------------|
| konkrete Seminarerfahrung | 63 | 46 |
| Anwendung medienbezogenen Wissens\ Seminarbezug vermutbar | 31 | 12 |
| keine Änderung begründet\ inhaltliche Begründung eines Messzeitpunkts\ Seminarbezug vermutbar | 13 | 3 |
| Summe direkter und vermuteter Seminarbezug | 107 (61 %) | 61 (43 %) |

Zur Identifikation lernförderlicher Elemente (Kombination mit *Testbewertung**Verbesserung*) konnten aus der Analyse aller Interviewtranskripte 63 Überschneidungen mit *konkreter Seminarerfahrung* gefunden werden. Bei einer Erweiterung des Verfahrens auf Textstellen mit vermutetem indirektem

Bezug ergaben sich weitere 44 Überschneidungen. Damit weisen von insgesamt 175 kodierten Segmenten zur *Verbesserung* mehr als die Hälfte (107 Segmente) einen konkreten oder möglichen indirekten Bezug zum untersuchten Seminar auf (Tabelle 13-1). Analog ergaben sich zur Identifikation lernhinderlicher Seminarelemente (Kombination mit *Testbewertung*\|*Verschlechterung*) 46 Überschneidungen mit *konkreter Seminarerfahrung* und weitere 15 mit vermutetem Seminarbezug (Tabelle 13-1). Von insgesamt 143 kodierten Segmenten zur *Verschlechterung* weisen hier weniger als die Hälfte (61 Segmente) einen konkreten oder vermuteten Bezug zum Seminar auf.

13.2.1 Identifikation lernförderlicher Seminarelemente

Die 107 Interviewsegmente zur *Verbesserung* (Tabelle 13-1) wurden zunächst ausgehend von der tabellarischen Übersicht hinsichtlich ihrer Kernaussagen paraphrasiert. Diese Paraphrasen wurden anschließend inhaltlichen Kategorien zugeordnet, die diese Inhalte und Erfahrungen im Seminar, welche zu einer Verbesserung in den Testantworten geführt haben, als Indizien für lernförderliche Seminarinhalte gruppieren.

Für die Prüfung dieser Zuordnungen der 107 Segmente erfolgte ebenfalls eine Doppelkodierung durch die Unterstützung der studentischen Hilfskraft. Vergleichbar zum Vorgehen beim Interrating in 12.3.1 wurden auch hier Änderungen an den Kategorien und ihren Definitionen auf Basis einer ersten Doppelkodierung und dem konsensuellen Austausch darüber vorgenommen. Anschließend ergab sich eine prozentuale Übereinstimmung von 88 % für die Segmente mit direktem Seminarbezug und von 73 % für die Segmente mit vermutetem Seminarbezug. Dies entspricht κ -Werten nach Brennan und Prediger (1981) von $\kappa_n = 0,88$ und $\kappa_n = 0,71$, welche im sehr guten und guten Bereich liegen (Kuckartz, 2018; Rädiker & Kuckartz, 2019).

Tabelle 13-2 zeigt die finale Zuordnung der 107 Segmente mit Kombination zur *Verbesserung* zu den verschiedenen Kategorien. Dabei waren Mehrfachzuweisungen möglich. So wurden beispielsweise *praktische Erfahrungen zum Einsatz digitaler Medien im Seminar*, *Designprinzipien (zur CTML und CLT)* oder *konkrete Seminarinhalte zu einzelnen Medien aus dem theoretischen Seminarteil* als Kategorien aufgenommen, welche jeweils differenzierende Unterkategorien enthalten. Eine Beschreibung aller (Unter-)Kategorien

findet sich in Anhang G. In der Tabelle ist jeweils die Häufigkeit für Segmente mit konkretem Seminarbezug eingetragen und zusätzlich in Klammern die Häufigkeit für jene Segmente mit vermutetem Seminarbezug. Diese separate Auftragung soll den unterschiedlichen Graden der Sicherheit des Seminarbezugs bei der Identifikation lernförderlicher Seminarelemente gerecht werden.

Tabelle 13-2: Kategorien zur Identifikation lernförderlicher Seminarelemente und Angabe der Häufigkeiten bei Überschneidungen *konkreter Seminarerfahrung (Seminarbezug vermutbar)* mit *Verbesserung*

| Oberkategorie | Unterkategorien | Häufigkeit |
|--|--|----------------|
| | Ausprobieren (im theoretischen Seminar- teil) | 13 (7) |
| Praktischer Einsatz/ praktische Erfah- rung im Seminar | eigene praktische Erprobung | 11 (1) |
| | praktische Erprobung anderer Gruppen | 3 (2) |
| | nur Oberkategorie | 2 (1) |
| | Gesamt | 29 (11) |
| Seminarinhalt zu einem Medium im Theorieteil | Erklärvideos | 12 (12) |
| | Digitale Messwerterfassung | 11 (6) |
| | Simulationen | 5 (9) |
| | Gesamt | 28 (27) |
| Designprinzipien (CTML/CLT) | Anwendung auf Erklärvideos | 12 (2) |
| | Allgemein | 2 (0) |
| | Gesamt | 14 (2) |
| Allgemeines, unkonkreter Bezug | kritischere Sichtweise nach dem Seminar | 12 (21) |
| | erhöhte Akzeptanz für Vorteile | 7 (4) |
| | Diskussionen im Seminar | 3 (0) |
| | Gesamt | 19 (25) |

Die angegebenen Häufigkeiten beziehen sich auf die 63 Interviewsegmente mit konkretem Seminarbezug; zusätzlich in Klammern sind die Häufigkeiten in Bezug auf die 44 Segmente mit vermutetem indirektem Bezug gegeben. Da Mehrfachzuordnungen möglich waren, ergibt die Summe der Häufigkeiten aller Unterkategorien nicht immer die Gesamthäufigkeit der Oberkategorie.

Aus dieser Übersicht in Tabelle 13-2 wird deutlich, dass praktische Erfahrungen zum Einsatz digitaler Medien, z. B. beim Ausprobieren oder Kennenlernen einzelner Medien im theoretischen Teil des Seminars (insg. 20 Zuordnungen) oder in den Phasen der eigenen praktischen Auseinandersetzung oder Erprobung (insg. 12 Zuordnungen) zu einer Verbesserung in den Testantworten führten. Das nachfolgende Zitat 3 stellt ein Beispiel dar, in dem das Kennenlernen von Medienbeispielen im Seminar als Begründung für eine (zum Positiven) veränderte Antwort genannt wurde. Das Zitat bezieht sich dabei auf Aufgabe 8 (Abbildung 10-3, dritte Antwortoption).

Zitat 3: T22c04 (Pos. 99)

„Ich glaube, ich hab an sehr geführte Simulationen gedacht und habe dann später in den Beispielen, die wir im Seminar gemacht haben, gesehen, dass viele Simulationen zwar irgendwie so eine so eine kleine Aufgabe dabei haben oder so aber trotzdem relativ frei sind, und dass man eben viel damit rumspielen kann und ich habe dann so diese Strukturierung des Lernprozesses irgendwie nicht mehr so gesehen also ich glaube nicht, dass es nur durch eine Simulation gegeben werden kann, ich bin mir ziemlich sicher, dass man da extra noch Anleitung braucht, oder ein Arbeitsblatt, oder sonst irgendwas, um das zu strukturieren.“

Ein Beispiel, in dem die positiv veränderte Antwort mit einer Erfahrung in der praktischen Erprobung begründet wurde, zeigt das in 12.3.2 dargestellte Zitat 1.

Neben den praktischen Erfahrungen werden zudem auch Seminarinhalte zu einzelnen Medien aus dem Theorieteil (z. B. typische Vorteile, Gestaltungsmerkmale oder Qualitätskriterien) als Gründe für zum Positiven veränderte Testantworten angeführt. Dabei werden Inhalte zur digitalen Messwerterfassung und zu Erklärvideos häufiger (konkret) genannt als Inhalte zu Simulationen. Diese konkreten Nennungen hängen jedoch auch davon ab, welche Aufgaben im Interview als Stimulus vorgelegt und gemeinsam mit den Befragungspersonen besprochen wurden (12.2). Zwar wurde bei der Gestaltung der Interviews versucht, Aufgaben von allen Kategorien einzubinden und keine der Kategorien bewusst zu fokussieren. Jedoch beziehen sich die 107 untersuchten Segmente zur Kombination von Seminarbezug und Verbesserung häufiger auf Aufgaben der Kategorien *digitale Messwerterfassung* und

Erklärvideos (37 und 33 Segmente) als auf Aufgaben der Kategorien *fachbezogene Grundlagen* und *Simulationen* (17 und 20 Segmente). Somit können entsprechend auch mehr lernförderliche Seminarelemente zur digitalen Messwerterfassung und zu Erklärvideos identifiziert werden, wenn mehr Aufgaben zu diesen Themen Gegenstand dieser Untersuchung zu möglichen lernförderlichen Seminarinhalten waren.

Zudem sind die Designprinzipien zum Multimedialen Lernen (bzw. der CTML und CLT; 2.3.3) und insbesondere deren Anwendung auf die Bewertung von Erklärvideos eine häufig genannte Ursache für positive Veränderungen in den Testantworten (insg. 16 Zuordnungen). Ein Beispiel zeigt das nachfolgende Zitat 4.

Zitat 4: A21a01 (Pos. 74)

„Ja da habe ich meine ich an die Cognitive Load Theory gedacht und daran, dass es ja so einen Redundanzeffekt gibt, wenn dieselbe Information gesprochen und geschrieben dasteht, was dann, meiner Erinnerung nach, eher lernhinderlich ist. Mit dem Wissen habe ich da dann meine Meinung geändert bei dem Punkt.“

Weiterhin deutete sich in einigen der betrachteten positiven Veränderungen in den Testantworten an, dass die Teilnehmenden in ihrer Begründung eine kritischere Sichtweise (z. B. bei der Einschätzung eines Medieneinsatzes oder -beispiels) aufzeigten (insg. 33 Zuordnungen). Hier ist jedoch zu beachten, dass sich ein Großteil dieser Zuordnungen auf Segmente mit *vermutetem* Bezug zum Seminar bezieht. Zudem wurde die Kategorie *kritischere Sichtweise* selten als einziges vergeben, sondern meist in Kombination mit weiteren Kategorien zu den Indizien für lernförderliche Inhalte. Ein Beispiel dazu (mit vermutetem, indirektem Seminarbezug) zeigt das folgende Zitat 5 zu Aufgabe 7 (Abbildung 10-2; erste Antwortoption):

Zitat 5: T22c05 (Pos. 96-98) (B: Befragungsperson, I: Interviewende)

B: „Ja ich glaube, dass ich da einfach gedacht habe beim Ersten ja, analoge Medien, also nicht digitale können genauso gut. Also da kann man genauso gut eine vorbereitende Hausaufgabe aufgeben, da muss man- also es ist jetzt kein besonderes Potential von einem digitalen Medium. [...]

I: „Ja ok und was glauben Sie woran das gelegen haben könnte, dass Sie es vorher noch mit ja angekreuzt haben aber dann doch eben anders eingeschätzt haben später?“

B: „Ehm ich glaube, weil ich da nicht so über das besondere Potential nachgedacht habe und dachte digitale Medien sind immer besser.“

13.2.2 Identifikation lernhinderlicher Seminarelemente oder Probleme

Durch analoges Vorgehen ließen sich auch die 61 Fälle mit (vermutetem) Seminarbezug zur Verschlechterung (Tabelle 13-1) betrachten. Diese 61 Interviewsegmente wurden ebenfalls paraphrasiert und inhaltlichen Kategorien zugeordnet, um hier Indizien für mögliche lernhinderliche Seminarinhalte zu erhalten.

Auch diese Zuordnungen zu Verschlechterungen wurden mit der Unterstützung der studentischen Hilfskraft doppelkodiert (im analogen Vorgehen wie die zur Verbesserung; 13.2.1). Hierbei ergaben sich prozentuale Übereinstimmungen und κ -Werte (nach Brennan & Prediger, 1981) von 78 % und $\kappa_n = 0,77$ (Segmente mit direktem Seminarbezug) und 48 % und $\kappa_n = 0,44$ (Segmente mit vermutetem Seminarbezug), wobei nur die Werte für den direktem Seminarbezug im guten Bereich liegen (Kuckartz, 2018; Rädiker & Kuckartz, 2019). Der geringere κ -Wert bei Segmenten mit vermutetem Seminarbezug ergibt sich möglicherweise dadurch, dass diese Segmente keine konkreten Seminarinhalte aufzeigen, sondern den vermuteten Bezug nur indirekt liefern. Eine inhaltliche Kategorisierung dieser weniger konkreten Textstellen ist damit schwieriger und offensichtlich fehlerbehafteter bzw. weniger eindeutig.

Bei der Zuordnung zeigte sich, dass durch diese Analyse zudem Indizien für mögliche Probleme oder typische Fehler von Studierenden in der Beantwortung der Testaufgaben oder ihrer Argumentation dazu identifiziert werden konnten. So wurde neben einer Kategorie zu *konkreten Seminarinhalten*, die i. d. R. unpassend angewendet wurden, z. B. eine Kategorie *Eigenschaft eines Mediums in einem Kontext verallgemeinert* für die Fälle gebildet, in denen die Teilnehmenden Erfahrungen unpassend verallgemeinert und/oder die Antwortoptionen im Test nicht ausreichend kritisch betrachteten. Tabelle 13-3 zeigt die finale Zuordnung der 61 Segmente zur Verschlechterung auf die verschiedenen Kategorien und Unterkategorien zu Indizien für lernhinderliche Elemente oder mögliche Probleme und Fehler. Auch hier sind die Häufigkeiten zu Segmenten mit vermutetem Seminarbezug in Klammern ausgewiesen.

Die Beschreibungen zu den Kategorien und Unterkategorien sind in Anhang G aufgeführt.

Tabelle 13-3: Kategorien zur Identifikation lernhinderlicher Seminarelemente oder möglicher Probleme und Angabe der Häufigkeiten bei Überschneidungen *konkreter Seminarerfahrung (Seminarbezug vermutlich)* mit *Verschlechterung*

| Oberkategorie | Unterkategorien | Häufigkeit |
|--|---|----------------|
| konkrete Seminarinhalte | Praxis/ praktischer Einsatz | 6 (2) |
| | Erklärvideos | 5 (0) |
| | Digitale Messwerterfassung | 3 (2) |
| | Multimedialernen (CTML / CLT) | 3 (1) |
| | Simulationen | 2 (0) |
| | Gesamt | 18 (5) |
| Eigenschaft eines Mediums in einem Kontext verallgemeinert | Distraktor nicht kritisch genug bewertet | 19 (10) |
| | Verallgemeinerung einer positiven Erfahrung (zu typischem Vorteil) | 19 (5) |
| | „negative“ Erfahrung verallgemeinert (Attraktor zu streng bewertet) | 8 (1) |
| | vorher skeptische Einstellung | 6 (5) |
| | unpassende Verallgemeinerung | 6 (5) |
| | Gesamt | 37 (15) |
| Fehlverständnis zu einem Medium | Vermischung von Medien | 4 (1) |
| | Fehlverständnis zu einer Medieneigenschaft | 1 (0) |
| | Gesamt | 5 (1) |
| erster Kontakt mit einem Medium | fehlendes ‚technisches‘ Wissen vorher | 6 (1) |
| | Erwartungen (zu Vorteilen) nicht erfüllt | 2 (0) |
| | Gesamt | 7 (1) |

| | | |
|-----------|--|-------|
| | Fehlverständnis zur Frage | 2 (2) |
| Sonstiges | bessere Beispiele im Seminar als im Test | 1 (0) |
| | Gesamt | 3 (2) |

Die angegebenen Häufigkeiten beziehen sich auf die 46 Interviewsegmente mit konkretem Seminarbezug; zusätzlich in Klammern sind die Häufigkeiten in Bezug auf die 15 Segmente mit vermutetem indirektem Bezug gegeben. Mehrfachzuordnungen möglich. Da Mehrfachzuordnungen möglich waren, ergibt die Summe der Häufigkeiten aller Unterkategorien nicht immer die Gesamthäufigkeit der Oberkategorie.

Die Auswertung zeigt für die Fälle, bei denen es zu einer Verschlechterung gekommen ist, dass häufig eine Erfahrung oder ein Inhalt aus dem Seminar unpassend verallgemeinert wurde. Bei positiven Erfahrungen mit den Medien wurden diese zu typischen Vorteilen verallgemeinert (insg. 24 Zuordnungen) und/oder Distraktoren im Posttest nicht ausreichend kritisch bewertet und als zutreffend ausgewählt (insg. 29 Zuordnungen). Das nachfolgende Zitat 6 stellt exemplarisch einen Fall dar, in dem eine (zum Negativen) veränderte Testantwort durch eine positive Erfahrung im Rahmen des besuchten Seminars begründet wurde. Es handelt sich dabei um den Distraktor „Simulationen ermöglichen das Abdecken von Lernzielen durch die Strukturierung des Lernprozesses“ in Aufgabe 8 (Abbildung 10-3; vierte Antwortoption).

Zitat 6: T22c02 (Pos. 96)

„Die erste Änderung ist tatsächlich einfach dem Seminar geschuldet, weil ich mir nicht, also ich konnte mir am Anfang nicht vorstellen, inwieweit eine Simulation Lernprozesse strukturiert, aber dadurch, dass ich ja den Fokus auf relevante Größen setzen kann und sage ‚das lassen wir mal konstant und wir verändern jetzt nur die Temperatur‘, kann ich ja den Wissenserwerb immer weiter staffeln.“

Die Person in diesem Zitat bewertete demnach den Distraktor im Posttest nicht kritisch, da sie nicht abgrenzte, dass die Strukturierung des Lernprozesses durch passende Aufgabenstellungen von der Lehrperson – und nicht von der Simulation allein – geleistet wird. Interessanterweise gelingt die Abgrenzung bei anderen Personen jedoch: das im vorigen Abschnitt dargestellte Zitat 3 bezieht sich auf denselben Distraktor, welcher dort im Posttest rich-

tigerweise abgelehnt wird, da diese Person erkennt, dass eine Aufgabenstellung zur Strukturierung des Lernprozesses bei der Arbeit mit Simulationen notwendig ist.

Das nächste Zitat 7 bezieht sich auf die fünfte Antwortoption aus derselben Aufgabe (Distraktor: *„Simulationen bereichern den Lernprozess durch fachbezogene Übungsaufgaben an.“*). Hier wird bei der Bewertung des Distraktors im Posttest nicht kritisch abgegrenzt, dass die Übungsaufgaben von den Mitstudierenden erstellt wurden und nicht durch die Simulation selbst.

Zitat 7: T22c05 (Pos. 94)

„Ja also da dachte ich dann halt, also wir haben zumindest im Seminar eben dann die Simulation genutzt oder es hat eben eine andere Gruppe das Beispiel genommen und Übungsaufgaben zu dieser Simulation eben erstellt und dadurch wurde meiner Meinung nach der Lernprozess angereichert, ja.“

Umgekehrt traten auch Fälle auf (jedoch seltener), in denen eine eher negative Erfahrung (z. B. als ein Experiment mit dem Smartphone im Seminar nicht direkt funktionierte) verallgemeinert wurde und bei der Beantwortung des Tests zur Ablehnung von Attraktoren führte (insg. 9 Zuordnungen). Weiterhin wurde in manchen Antworten erkennbar, dass die im Seminar besprochenen digitalen Medien oder ihre Eigenschaften von den Teilnehmenden vermischt bzw. verwechselt oder nicht verstanden wurden (insg. 6 Zuordnungen).

13.2.3 Zusammenfassung und Vergleich mit Erkenntnissen zur subjektiv eingeschätzten Lernförderlichkeit

Insgesamt lassen sich aus der vorgestellten Analyse einerseits Indizien für lernförderliche Seminarinhalte ableiten. Es zeigten sich praktische Erfahrungen zum Einsatz digitaler Medien – sei es beim Ausprobieren im eher theoretischen Seminarteil oder im eigenen Einsatz der praktischen Erprobung bzw. Auseinandersetzung – als häufig genannte Gründe für zum Positiven veränderte Testantworten und dementsprechend vermutlich lernförderliche Seminarinhalte. Ebenso scheinen (theoretische) Seminarinhalte zu digitaler Messerwerterfassung, Erklärvideos und zu den Designprinzipien zum Multimedialen Lernen (CTML) und der CLT sowie insbesondere deren Anwendung

auf Erklärvideos lernförderlich, da auch hier häufig eine Verbesserung in den Testantworten mit diesen Begründungen zur Veränderung zusammenhing.

Die Analyse ergab andererseits weniger Hinweise für konkrete Seminarinhalte, die als lernhinderlich zu deuten sind. Hingegen wurden eher Probleme hinsichtlich der Lernprozesse und der Argumentation in den Studierendenantworten deutlich. So hingen die meisten Verschlechterungen mit einer unpassenden Verallgemeinerung oder Übergeneralisierung von Einzelerfahrungen oder -aspekten im Seminar zusammen.

Im Rahmen der retrospektiven Interviews wurde zudem die Einschätzung der Studierenden zur Lernförderlichkeit einzelner Seminarinhalte und -tätigkeiten über eine Kurzumfrage erhoben und anschließend in Hauptphase A des Interviews die Begründungen zu einigen Einschätzungen erfragt (12.2). Die Analyse dieses Interviewteils steht nicht im Fokus der vorliegenden Arbeit, sie wurde jedoch im Rahmen einer Bachelorarbeit an der Universität Paderborn vorgenommen (Veltum, 2023). Die dort gewonnen Erkenntnisse zur subjektiv wahrgenommenen Lernförderlichkeit einzelner Seminarelemente werden im nachfolgenden kurz dargestellt und mit der Analyse in dieser Arbeit zur Identifikation lernförderlicher Seminarelemente (über die veränderten Testantworten) verglichen.

Die Auswertung der Kurzumfrage in der Bachelorarbeit zeigte ebenfalls, dass Tätigkeiten zum Ausprobieren und Kennenlernen verschiedener Medien als sehr unterstützend wahrgenommen wurden (Seminarinhalte/-tätigkeiten: Ausprobieren und Vorstellung einer Simulation, Ausprobieren einer digitalen Messwerterfassung und Ausprobieren einer mobilen Videoanalyse, Analyse eines Erklärvideos aus dem Internet). Auch die in dieser Arbeit als lernförderlich identifizierten Designprinzipien zum Multimedialen Lernen und zur CLT stellten nach der Einschätzung der Studierenden ein förderliches Seminarelement dar. Hinsichtlich der Erfahrungen aus dem praktischen Seminarteil wurde deutlich, dass die Betreuung in der Schulerprobung und die anschließende Reflexion dieser von den Studierenden (in Aachen) als unterstützend wahrgenommen wurden, während die Einschätzungen zur Vorbereitung, Durchführung und Reflexion der praktischen Auseinandersetzungen in Tübingen und Graz im Vergleich etwas weniger förderlich bewertet

wurden (Veltum, 2023). In der Bachelorarbeit wurden zudem die Begründungen für einzelne Einschätzungen identifiziert. In den Begründungen der Studierenden äußerte sich dabei insbesondere das Lernen durch eigenes Handeln und die Relevanz für den eigenen Unterricht. In einigen Fällen wurde die Lernförderlichkeit zudem mit Bezug zu individuellen Voraussetzungen begründet (z. B. persönliches Interesse an einem Inhalt oder umgekehrt persönliche Abneigung). Teilweise hing die angegebene Einschätzung zur Lernförderlichkeit auch mit dem Vorwissen zusammen, wenn z. B. ein Inhalt aus anderen Lehrveranstaltungen schon bekannt war und daher das entsprechende Element als weniger hilfreich eingeschätzt wurde. Als weitere Gründe wurden für als nicht hilfreich wahrgenommene Seminarelemente z. B. eine zu geringe Lernzeit geäußert und für als hilfreich wahrgenommene Elemente das Erhalten von Feedback (Veltum, 2023). Insgesamt zeigt der Vergleich ein stimmiges Bild im Hinblick darauf, welche Seminaraspekte und -inhalte Verbesserungen im gemessenen FDW begünstigen und auch von den Studierenden als hilfreich wahrgenommen werden.

Die gewonnenen Indizien zu lernförderlichen Seminarelementen und möglichen Problemen können nun für die Hypothesengenerierung für lernwirksame Lerngelegenheiten zum Einsatz digitaler Medien im Rahmen des fünften Ziels dieser Arbeit dienen (6.2). Im folgenden Abschnitt werden diese entwickelten Hypothesen als Folgerungen für die Gestaltung entsprechender Lerngelegenheiten zu digitalen Medien im Physikunterricht erläutert.

13.3 Hypothesenentwicklung: Folgerungen für die Seminargestaltung

Die vorgestellte Analyse der Interviews deutet an, welche Elemente des Lernkonzepts förderlich für den Wissenserwerb sind, aber auch welche Probleme bei der Beantwortung der Testaufgaben auftreten, die im Zusammenhang mit Seminarerfahrungen stehen.

So wurden durch die Interviews Fälle deutlich, in denen aufgrund einer Erfahrung im Seminar eine Testantwort zum Negativen verändert wurde. Dabei äußerte sich bei den Teilnehmenden häufig eine fehlende kritische Reflexion der betrachteten Medienbeispiele oder -einsätze. Digitale Medien werden zum Teil als eine Art „Allheilmittel“ gesehen, das sämtliche Aufgaben einer Lehrkraft übernimmt oder ersetzt (z. B. Strukturierung des Lernprozesses,

Angebot von Übungsaufgaben; 13.2.2). Aufgrund positiver Einzelerfahrungen im Seminar mit einzelnen digitalen Medien wird diesen mehr zugesprochen und nicht über mögliche Grenzen zum Einsatz digitaler Medien reflektiert. Überwiegend positive Beispiele zu Medien oder Medieneinsätzen erzeugen bei den Teilnehmenden wohlmöglich den Eindruck, dass der Einsatz digitaler Medien per se gut für den Unterricht ist, sodass sie erfahrene Vorteile in einer Situation unreflektiert auf andere Kontexte verallgemeinern. Zitat 6 und Zitat 7 verdeutlichen diesen Aspekt exemplarisch. Aufgrund dieser Erkenntnisse lautet die erste Folgerung oder Hypothese zur Gestaltung von Lerngelegenheiten zum Einsatz digitaler Medien im Physikunterricht:

1. Umfassendes kritisches Reflektieren von Medienbeispielen/-einsätzen in der Lehrveranstaltung gezielt ermöglichen.

Durch ein gezieltes kritisches Reflektieren und gemeinsames Diskutieren (von positiven sowie negativen Medienbeispielen) soll ein Positivismus oder Über-Enthusiasmus vermieden werden. Zudem sollte in einer Lehrveranstaltung für eine Unterscheidung zwischen anekdotischer und empirischer Evidenz sensibilisiert werden, um das unpassende Verallgemeinern von Einzelerfahrungen auf andere Situationen zu vermeiden.

Da die Praxiserfahrungen beim Ausprobieren und bei den Erprobungen zum Einsatz digitaler Medien häufig mit Verbesserungen verbunden waren und demnach hilfreich scheinen (13.2.1), sollten ebensolche praktischen Erfahrungen in einer Lehrveranstaltung ermöglicht werden. Einige Studierende gaben an, dadurch Möglichkeiten oder Vorteile hinsichtlich des Einsatzes digitaler Medien erkannt zu haben, die ihnen vorher nicht bewusst waren. Andere berichteten, dass die Erfahrungen mit Medien im Seminar dazu beitragen, ihre (zuvor eher negativen) Erfahrungen aus der eigenen Schulzeit umzukehren. Gemäß den Erläuterungen der Studierenden hinsichtlich der subjektiv wahrgenommenen Lernförderlichkeit der Seminarinhalte sollte solchen Phasen zum Ausprobieren ausreichend Zeit in der Lehrveranstaltung eingeräumt werden (13.2.3). Eine eigene Implementation digitaler Medien in Unterrichtsszenarien (Vorbereitung sowie Durchführung) unterstützt zudem eine realistischere und reflektierte Einschätzung, da auch Anforderungen und Grenzen deutlich werden. Ein Beispiel dazu stellt Zitat 1 dar.

Jedoch zeigte die Interviewanalyse, dass Praxiserfahrungen auch mit Verschlechterungen im Test zusammenhingen, da teilweise die kritische Reflexion der gemachten Erfahrung fehlt (siehe oben). Wichtig ist daher stets, die praktischen Erfahrungen mit digitalen Medien auch (gemeinsam) zu reflektieren und diskutieren. Die zweite Folgerung lautet demnach:

2. Eigenes Ausprobieren und praktische Implementierung der Medien ermöglichen (in Verbindung mit anschließender Reflexion).

An einigen Stellen wurde in den Interviews zudem eine mangelnde Verknüpfung des Wissens zu verschiedenen digitalen Medien oder die Vermischung einzelner digitaler Medien deutlich. Die Versuche der Teilnehmenden, Wissen zu vernetzen oder Wissen zu einem Medium auf ein anderes zu übertragen, gelingen zum Teil nicht. Aus der quantitativen Analyse zeigt sich zudem die Tendenz, dass Aufgaben, die übergreifendes oder vernetzendes Wissen oder Vergleiche zu digitalen Medien erfordern, den Teilnehmenden schwerer fallen (Aufgaben 3, 7 und 11; Abbildung 11-1). Als dritte Folgerung wird daher formuliert:

3. Entwicklung von vernetztem Wissen zu Medien fördern, ohne individuelle Gestaltungsmerkmale der einzelnen Medien zu vernachlässigen.

Eine entsprechende Lerngelegenheit soll demnach die Möglichkeit bieten, erlerntes Wissen zu einzelnen digitalen Medien zu vernetzen, damit dieses nicht bloß als isoliertes Wissen vorliegt. Neben der Vorstellung einzelner Medien und ihrer Gestaltungsmerkmale in (isolierten) Einheiten sind dazu anknüpfende, übergreifende Phasen bedeutsam, um die Eigenschaften, Möglichkeiten und Grenzen der verschiedenen Medien hinsichtlich Gemeinsamkeiten und Unterschieden gegenüberstellen zu können. In Anwendungsübungen z. B. zur Gestaltung von Unterrichtssequenzen oder Planung von Medieneinsätzen sollten – anstelle einer Fokussierung auf ein Medium – die Möglichkeiten von verschiedenen Medien gemeinsam betrachtet und diskutiert werden.

Weiterhin deutete sich bei mehreren Teilnehmenden an, dass ihre Vorerfahrungen und -einstellungen die Wahrnehmung und Reflexion der Medienbeispiele im Seminar beeinflussen. So kam es vor, dass Studierenden aufgrund

ihrer eigenen Schulerfahrung hinsichtlich digitaler Medien eine skeptische Einstellung hatten, die sich aufgrund der Erfahrungen im Seminar positiv gewandelt hat (Tabelle 13-3). Umgekehrt zeigten sich auch Fälle, in denen vor dem Seminar eine äußerst positive und eher unreflektierte Einstellung vorlag, welche über das Seminar zu einer kritischeren Sichtweise verändert wurde (Tabelle 13-2). Daher wird als vierte Folgerung aufgenommen:

4. Vorerfahrungen und -einstellungen der Seminarteilnehmenden explizit adressieren.

Die eigenen Vorerfahrungen und möglichen Einstellungen zum Einsatz digitaler Medien sollten den Teilnehmenden bewusst gemacht werden, indem Raum gegeben wird, diese zu diskutieren und neue Erfahrungen im Seminar in Bezug auf die eigenen Vorerfahrungen zu reflektieren. Eine gemeinsame Diskussion mit allen Teilnehmenden reduziert zudem, dass mögliche unreflektierte Extrem-Ansichten entstehen oder gefestigt werden.

Diese Schlussfolgerungen aus der Interviewanalyse, welche als Hypothesen zur Gestaltung von Lerngelegenheiten zum fachdidaktisch sinnvollen Einsatz digitaler Medien im Physikunterricht dienen, sind in Abbildung 13-2 zusammengefasst. Dabei sind diese Hypothesen als Gestaltungsmerkmale nicht unabhängig voneinander und sollten auch gemeinsam berücksichtigt werden. Insbesondere die kritische Reflexion und gemeinsame Diskussion ist für alle vier Gestaltungsmerkmale bedeutsam. Im Anhang ist zudem in einer Übersicht dargestellt, aus welchen Indizien und Erkenntnissen die Hypothesen entwickelt wurden und welche (weiteren) Textstellen aus den Interviews diese stützen (Anhang H).

- 1. Umfassendes kritisches Reflektieren von Medienbeispielen/-einsätzen in der Lehrveranstaltung gezielt ermöglichen**
- 2. Eigenes Ausprobieren und praktische Implementierung der Medien ermöglichen (in Verbindung mit anschließender Reflexion)**
- 3. Entwicklung von vernetztem Wissen zu Medien fördern, ohne individuelle Gestaltungsmerkmale einzelner Medien zu vernachlässigen**
- 4. Vorerfahrungen und -einstellungen der Seminarteilnehmenden explizit adressieren**

Abbildung 13-2: Entwickelte Hypothesen zur Gestaltung von Lerngelegenheiten zum fachdidaktisch sinnvollen Einsatz digitaler Medien im Physikunterricht

14 Fazit

In diesem Kapitel werden zunächst die Erkenntnisse der vorliegenden Arbeit hinsichtlich der formulierten Ziele und Forschungsfragen zusammengefasst und diskutiert (14.1). Daraufhin werden in 14.2 die Grenzen der Arbeit beleuchtet, um die Belastbarkeit der Ergebnisse einschätzen zu können. In 14.3 wird abschließend ein Ausblick zu Anknüpfungspunkten in der Lehrkräftebildung und Lehrkräftebildungsforschung gegeben.

14.1 Diskussion der Ergebnisse

Die vorliegende Arbeit adressierte die Entwicklung von FDW zum Einsatz digitaler Medien bei Physiklehramtsstudierenden. Das Forschungsanliegen wurde durch die zunehmende Bedeutung digitaler Medien in Schule und Fachunterricht motiviert. Daraus ergibt sich die Notwendigkeit, angehende Lehrkräfte für einen fachdidaktisch sinnvollen Einsatz digitaler Medien im Unterricht zu professionalisieren und hinsichtlich des Erwerbs fachdidaktischer digitalisierungsbezogener Kompetenzen zu fördern (Kapitel 1). Ausgehend vom Verständnis über das FDW als Teil der professionellen Handlungskompetenz von (angehenden) Lehrkräften (2.1) wurde das FDW zum Einsatz digitaler Medien (als Facette des FDW in Physik; 2.2) fokussiert, welche wiederum dem zentralen Wissensbereich TPACK des TPACK-Modells ähnlich ist (2.3). Daher beschäftigte sich die Arbeit insbesondere mit den Fragen, wie dieses Wissen angemessen erfasst und in der universitären Lehrkräftebildung effektiv gefördert werden kann.

Ausgehend von der Beschreibung theoretischer Grundlagen (Kapitel 2) wurden bisherige Forschungsergebnisse zur Erfassung von FDW und TPACK (Kapitel 3) sowie deren Förderung in der Lehrkräftebildung (Kapitel 4) beschrieben. In Kapitel 5 erfolgte die Vorstellung der in dieser Arbeit untersuchten Lerngelegenheiten zur Förderung des FDW zum Einsatz digitaler Medien. Auf Basis des bisherigen Forschungsstandes wurden in Kapitel 6 vier Desiderate herausgestellt (6.1) zu Aspekten, die (nach Kenntnisstand der Autorin) bis zu diesem Zeitpunkt nicht vorlagen und in dieser Arbeit adressiert wurden:

1. Es existiert keine konkrete fachspezifische Beschreibung des FDW zum Einsatz digitaler Medien (bzw. TPACK) im Fach Physik.

2. Es existieren keine validen Messverfahren zur fachspezifischen Erfassung des FDW zum Einsatz digitaler Medien (bzw. TPACK) im Fach Physik.
3. Es existieren keine Untersuchungen zur Evaluation physikdidaktischer Lerngelegenheiten im Lehramtsstudium hinsichtlich der Entwicklung des FDW zum Einsatz digitaler Medien im Physikunterricht (bzw. TPACK).
4. Es fehlen Erkenntnisse über Elemente von fachdidaktischen Lerngelegenheiten, die förderlich für den Erwerb von FDW zum Einsatz digitaler Medien (bzw. TPACK) sind.

Anknüpfend an diese Desiderate wurden für die Arbeit fünf Ziele abgeleitet und drei zugehörige Forschungsfragen sowie eine Hypothese formuliert (6.2):

Ziel 1: Modellierung des FDW zum Einsatz digitaler Medien im Physikunterricht.

Ziel 2: Entwicklung und Validierung eines Testinstruments zur Erfassung des FDW zum Einsatz digitaler Medien von Physiklehramtsstudierenden.

FF1: Inwiefern lässt sich das physikdidaktische Wissen zum Einsatz digitaler Medien mithilfe des entwickelten Testinstruments valide messen?

Ziel 3: Untersuchung des Wissenserwerbs im FDW zum Einsatz digitaler Medien über die im Verbundprojekt DiKoLeP beteiligten Seminare.

FF2: Inwieweit verändert sich das gemessene physikdidaktische Wissen zum Einsatz digitaler Medien über die Seminare der kooperierenden Standorte im Verbundprojekt DiKoLeP?

H1: Das FDW zum Einsatz digitaler Medien der teilnehmenden Studierenden nimmt über die untersuchten Seminare zu.

Ziel 4: Erkundung von Gründen für Veränderungen im Testverhalten zum FDW zum Einsatz digitaler Medien und Identifikation lernförderlicher und lernhinderlicher Seminarelemente.

FF3: Welche Elemente des gemeinsamen Kerns bzw. der standortspezifischen Teile des Lehrkonzepts sind besonders lernförderlich oder eher

lernhinderlich und stehen mit Verbesserung oder Verschlechterung im gemessenen FDW zum Einsatz digitaler Medien in Verbindung?

Ziel 5: Generierung von Hypothesen für wirksame Lerngelegenheiten zum Einsatz digitaler Medien im Physikunterricht.

Im Folgenden werden die Erkenntnisse aus der Arbeit hinsichtlich dieser Ziele und Forschungsfragen jeweils zusammenfassend dargestellt und diskutiert.

Modellierung, Testentwicklung und Untersuchungen zur Validität (FF1)

Das erste Ziel der Arbeit stellte die Modellierung des fokussierten Wissensbereichs FDW zum Einsatz digitaler Medien im Fach Physik dar, da bisher keine konkrete physikspezifische Beschreibung zu medienbezogenem FDW bzw. TPACK vorlag (6.1; 6.2). Die Modellierung erfolgte ausgehend von der Definition zur Facette (Digitale) Medien des FDW in Physik (Gramzow, 2015; Gramzow et al., 2013; 8.1). Unter Berücksichtigung fachspezifischer sowie übergreifender Beschreibungen zu digitalisierungsbezogenen Kompetenzen und (fach-)didaktischer Literatur zum Einsatz digitaler Medien im Unterricht wurde in Zusammenarbeit mit den Projektpartnern im Verbundprojekt DiKoLeP eine ausdifferenzierte Beschreibung entwickelt. Die finale Beschreibung oder Modellierung differenziert die vier Kategorien *fachbezogene Grundlagen*, *digitale Messwerterfassung*, *Simulationen* und *Erklärvideos* (8.1). Somit konnte das erste Ziel zur Modellierung des FDW zum Einsatz digitaler Medien erreicht werden.

Ebenso fehlte es bisher an fachspezifischen Messverfahren für das FDW zum Einsatz digitaler Medien. Etablierte physikdidaktische Leistungstests adressierten bislang nicht den Einsatz digitaler Medien, während die Erfassung des TPACK meist über Selbsteinschätzungen erfolgte, die in ihrer Validität eingeschränkt sind (3.5.3). Die wenigen leistungsbasierten Messverfahren zum TPACK beziehen sich auf andere naturwissenschaftliche Fächer als Physik (3.5.2). Deshalb stellte das zweite Ziel der Arbeit die Entwicklung eines Leistungstests zur validen Erfassung des FDW zum Einsatz digitaler Medien im Fach Physik in den Mittelpunkt. Für das Testinstrument erfolgte die Aufgabenentwicklung in Anlehnung an das Vorgehen zur Entwicklung eines

etablierten Leistungstests zum FDW in Physik im Projekt ProfiLe-P (Gramzow, 2015; Riese et al., 2017; 3.4). Demnach wurden systematisch Testaufgaben zu den vier modellierten Kategorien sowie zu den drei kognitiven Anforderungen *Reproduzieren*, *Anwenden* und *Analysieren* entwickelt. Als Aufgabenformat wurden für das Testinstrument in dieser Arbeit geschlossene Mehrfachwahlaufgaben gewählt, um eine zeitökonomische Auswertung zu ermöglichen, wobei wichtige Anforderungen zur Gestaltung von Mehrfachwahlaufgaben berücksichtigt wurden (8.2). Eine erste Testversion aus 17 Aufgaben wurde im Rahmen einer Pilotierung erprobt und anschließend überarbeitet (8.3). Durch weitere Erhebungen mit dem überarbeiteten Testinstrument konnten zudem Analysen zur Testgüte auf Basis der IRT vorgenommen werden. Diese ergaben gute Kennwerte hinsichtlich der Modellpassung und der Aufgabenschwierigkeiten sowie eine für den Kontext eines Leistungstests zu einem inhaltlich anspruchsvollen Konstrukt annehmbare EAP-Reliabilität (8.4). So konnte im Hinblick auf das zweite Ziel ein Leistungstest zum FDW zum Einsatz digitaler Medien im Fach Physik entwickelt werden, welcher in ersten Erhebungen angemessen nutzbar erschien.

Im Rahmen von FF1 erfolgte anknüpfend die Untersuchung der Validität des entwickelten Testinstruments. Mithilfe mehrerer Teilstudien zur Untersuchung der Validität sollten verschiedene Hinweise zur Konstruktvalidität im Sinne eines argumentbasierten Validitätsverständnisses (3.3.2) abgeleitet werden. Dabei wurden drei Aspekte untersucht, inwiefern mit dem Test (1) universitär erwerbbares Wissen, (2) Wissen, das sich von allgemein-didaktischem bzw. pädagogischem Wissen (zum Medieneinsatz) sowie von rein fachlichem Wissen abgrenzt und (3) Wissen mit einer angenommenen Strukturierung gemäß den vier modellierten Kategorien erfasst werden kann (Kapitel 9).

Eine Expert:innenbefragung zur curricularen Passung der entwickelten Testaufgaben ergab gute bis sehr gute Passungen für alle Aufgaben (9.4) und liefert damit einen wichtigen Hinweis für den ersten Aspekt zur Erwerbbarkeit des gemessenen FDW in der universitären Lehrkräftebildung.

Im Hinblick auf den zweiten Aspekt zeigen die Korrelationsanalysen mit Testleistungen zu anderen Wissensbereichen des Professionswissens (FDW

ohne Bezug zu digitalen Medien und PW) erwartungskonform, dass die Leistung im entwickelten Test stärker mit der Testleistung im FDW als mit der im PW zusammenhängt (9.1). Der entwickelte Test scheint demnach eher fachspezifisches FDW zu messen als allgemein-pädagogisches Wissen (zu digitalen Medien). Die geringeren Testleistungen der Lehramtsstudierenden in Chemie und Technik (ohne Kombinationsfach Physik) deuten ebenso erwartungskonform darauf hin, dass mit dem Test ein fachspezifisches Wissen gemessen wird. Jedoch zeigt sich in der Regressionsanalyse unter Einbezug der Abiturnote, dass diese einen signifikanten Einflussfaktor auf die Testleistung darstellt und der Einfluss der Fachzugehörigkeit (Physik und kein Physik) nicht mehr signifikant ist (9.2). Das Ergebnis könnte darauf hindeuten, dass mit dem entwickelten Leistungstest eher MINT-spezifisches als physikspezifisches Wissen zum Einsatz digitaler Medien erfasst wird. Zur weiteren Aufklärung wären zusätzliche Untersuchungen notwendig, da Erhebungen mit Studierenden aus weiteren Fachbereichen ggfs. andere Ergebnisse aufzeigen würden. Die Ergebnisse der Think-Aloud-Befragungen zur kognitiven Validierung unterstützen wiederum den zweiten Aspekt, da – sofern eine inhaltliche Wissensressource zuzuordnen war – vor allem FDW zur Beantwortung der Testaufgaben herangezogen wurde und weniger rein fachliches Wissen oder Schulerfahrungen (9.3).

Hinsichtlich des dritten Aspekts zur Untersuchung der angenommenen Modellstruktur aus vier Kategorien ergab der Vergleich einer eindimensionalen IRT-Modellierung mit einer vierdimensionalen Modellierung keine deutliche Überlegenheit des mehrdimensionalen Modells – beide Modellierungen weisen gute Modellfitwerte auf. Für die vier angenommenen Teilskalen zeigten sich jedoch geringere EAP-Reliabilitäten als für die Gesamtskala (9.5). Dieses Ergebnis spricht dafür, dass das entwickelte Testinstrument vor allem als Gesamtskala nutzbar ist.

Aus den durchgeführten Studien zur Beantwortung von FF1 lassen sich insgesamt zu mehreren Aspekten Indizien ableiten, die für eine geeignete Nutzung des entwickelten Testinstruments zur validen Messung von FDW zum Einsatz digitaler Medien bei Physiklehramtsstudierenden sprechen (9.6). Auch wenn nicht aus allen Studien gleichermaßen eindeutige Hinweise ab-

geleitet werden können, so wurden in der vorliegenden Arbeit mehr Validitätsaspekte zum Testinstrument untersucht und eine höhere Anzahl Indizien dafür gewonnen, als es in den meisten Studien zur Untersuchung des TPACK der Fall ist (3.5). Insofern konnte auch das zweite Ziel der Arbeit erreicht werden, wenngleich für die Nutzung des entwickelten Testinstrument auch gewisse Grenzen zu beachten sind (14.2).

Untersuchung der Entwicklung des FDW zum Einsatz digitaler Medien (FF2)

Im Rahmen des dritten Ziels der Arbeit wurden die im Verbundprojekt DiKoLeP entwickelten und implementierten physikdidaktischen Lehrveranstaltungen evaluiert. Dazu wurde im Rahmen von FF2 die Entwicklung des FDW zum Einsatz digitaler Medien bei den teilnehmenden Studierenden über die Seminare mithilfe des optimierten Leistungstests im Prä-Post-Design untersucht. Aufgrund der inhaltlichen Passung der Seminare zum untersuchten FDW sowie der Orientierung an Schlüsselaspekten des SQD-Modells, die zur Vorbereitung auf den Einsatz digitaler Medien als förderlich gelten (Tondeur et al., 2012; 5.2.1), wurde ein Zuwachs im untersuchten FDW zum Einsatz digitaler Medien über die Seminare erwartet (Wirkungshypothese H1).

Die standortübergreifende Evaluation ergab einen signifikanten Zuwachs im untersuchten FDW mit einer kleinen Effektstärke ($d = 0,26$), sodass H1 angenommen werden kann (11.2). Demnach scheinen das Lehrkonzept und die danach implementierten Seminare grundsätzlich geeignet, um das FDW zum Einsatz digitaler Medien zu fördern. Einschränkend ist an dieser Stelle jedoch anzumerken, dass die EAP-Reliabilität für die untersuchte Stichprobe von $N_1 = 70$ mit 0,47 gering ausfiel, wodurch die Belastbarkeit der Testergebnisse in der Untersuchung limitiert ist.

Dass der gemessene Zuwachs im FDW zum Einsatz digitaler Medien nur einen kleinen Effekt darstellt, ist insofern verwunderlich, da die untersuchten Seminare gerade zur Förderung dieses FDW entwickelt worden sind und entsprechend eine inhaltliche Passung zwischen den Kerninhalten der Seminar-konzepte und dem im Test fokussierten FDW vorliegt. Möglicherweise sind mit dem entwickelten Test jedoch über den kurzen Zeitraum eines Semesters keine größeren Wissensveränderungen messbar, weil das Testinstrument

ein breiteres Wissensspektrum im FDW abdeckt. In anderen Studien, die die Entwicklung von TPACK (oder TPACK-Wissensbereichen) über besuchte Lehrveranstaltungen untersuchten, werden mittlere bis große Effekte im Wissenszuwachs berichtet (Henne et al., 2022; Max, 2022; Stinken-Rösner et al., 2023; Zimmermann, 2021; 4.2.1), die demnach höher ausfallen als in der vorliegenden Studie. Jedoch basieren diese Untersuchungen teils auf Selbsteinschätzungen zum TPACK (Henne et al., 2022; Stinken-Rösner et al., 2023), was die Vergleichbarkeit mit der leistungsbasierten Messung in dieser Arbeit einschränkt. Zudem lässt sich das Ergebnis der vorliegenden Arbeit nicht mit einer Evaluation einer physikspezifischen Lehrveranstaltung vergleichen, da die aufgeführten Studien in anderen naturwissenschaftlichen Fächern stattfanden und keine vergleichbare Studie im Fach Physik bekannt ist (4.2.1). Schließlich bieten die Erkenntnisse aus den retrospektiven Interviews eine weitere mögliche Erklärung für den unerwartet kleinen Effekt beim Wissenserwerb, die im nachfolgenden Abschnitt erläutert wird.

Erkundung von Gründen für veränderte Testantworten und Identifikation lernförderlicher sowie lernhinderlicher Seminarelemente (FF3)

Anknüpfend an die quantitative Prä-Post-Erhebung zum FDW zum Einsatz digitaler Medien adressierte das vierte Ziel der Arbeit die Erkundung von Gründen für Veränderungen in den Testantworten sowie die Identifikation lernförderlicher oder lernhinderlicher Seminarelemente. Die zugehörige FF3 wurde im Rahmen einer explorativen Interviewstudie nach den schriftlichen Prä-Post-Erhebungen untersucht.

Die Erkundung der Gründe für die Veränderungen in den Testantworten erfolgte über eine inhaltlich-strukturierende qualitative Inhaltsanalyse der retrospektiven Interviews bzw. des Interviewteils, in dem die Befragten die veränderten Testantworten vorgelegt bekamen und sie zum Grund für die Änderungen befragt wurden (12.3). Diese Inhaltsanalyse zeigte zunächst, dass ein Großteil der veränderten Antworten entweder explizit mit dem untersuchten Seminar begründet wurde oder ein indirekter Bezug zum Seminar in der Begründung nahelag. Vergleichsweise seltener wurde eine Veränderung durch

die (sprachliche) Uminterpretation der Fragestellung, durch Unentschlossenheit oder unter Anwendung nicht-medienbezogenen Wissens begründet (13.1).

Aus diesen geringeren Häufigkeiten von nicht-inhaltlichen Begründungen oder Begründungen ohne Medienbezug lässt sich ein weiterer Validitätshinweis für die Angemessenheit der Testnutzung zur Erfassung des FDW zum Einsatz digitaler Medien ableiten. Nichtsdestotrotz wurde durch die Interviews deutlich, dass eine Veränderung in den Testantworten nicht ausschließlich auf verändertes Wissen oder auf eine Lerngelegenheit zurückzuführen ist, da auch andere Faktoren die Testbearbeitung beeinflussen. So kommt es bei der Bearbeitung des Tests scheinbar durchaus vor, dass Antworten geraten werden (insbesondere durch das Antwortformat Ja/Nein) oder sprachliche Aspekte der Aufgabenstellung einbezogen werden – wie das Auftreten der Begründungen in den Kategorien *Unentschlossenheit* und *Uminterpretation der Fragestellung* nahelegt. Jedoch verdeutlichen die Interviews, dass dies nicht sehr häufig der Fall ist, sondern bei den Befragten größtenteils eine begründete Entscheidungsänderung von Prä- zu Posttest vorzuliegen scheint. In der Arbeit von Joswig-Käfer (2024), die ebenfalls Ursachen für veränderte Testantworten über retrospektive Interviews nach Prä-Post-Erhebungen zum FDW erkundete, ergab sich in ähnlicher Weise, dass die Studierenden häufig Wissen aus universitären (insbesondere fachdidaktischen) Lerngelegenheiten heranzogen. Ebenso zeigten sich auch z. T. Einflüsse durch das Testinstrument, wie z. B. die Formulierung der Aufgabenstellungen. Der dort identifizierte Grund des veränderten Zeitmanagements bei der Testbearbeitung (Joswig-Käfer, 2024) trat in der vorliegenden Arbeit jedoch nicht auf, was möglicherweise durch die unterschiedlichen Antwortformate (überwiegend offene Aufgaben gegenüber ausschließlich Mehrfachwahlaufgaben) und die dadurch abweichenden Bearbeitungszeiten (ca. 60 Minuten gegenüber ca. 25 Minuten) bedingt sein könnte.

Für die Begründungen zu veränderten Antworten mit (vermutetem) Seminarbezug wurde schließlich betrachtet, inwiefern es sich bei diesen Veränderungen um eine Verbesserung oder eine Verschlechterung im Sinne der Testbewertung handelte, um somit Indizien für lernförderliche (bei Verbesse-

rung) und lernhinderliche (bei Verschlechterung) Seminarelemente zu gewinnen (13.2). Im Hinblick auf die Identifikation lernförderlicher Elemente stellte sich heraus, dass häufig praktische Erfahrungen im Seminar (z. B. beim Ausprobieren der Medien oder beim eigenen Einsatz in der praktischen Erprobung) mit einer Verbesserung in den Testantworten zusammenhängen und somit lernförderlich zu sein scheinen (13.2.1). Dieses Ergebnis unterstützt die Relevanz praktischer Erfahrungen und geht zudem mit den Schlüsselaspekten 4 und 6 des SQD-Modells (5.2.1), die bei der Gestaltung des Lehrkonzepts berücksichtigt wurden, sowie mit der Vermutung zur Wirksamkeit des Seminars einher (6.2). Gleichzeitig wurden auch eher theoretische Seminarinhalte (wie typische Vorteile oder Qualitätsmerkmale) zu einzelnen Medien (v. a. Erklärvideos und digitale Messwerterfassung) oder zur CLT und CTML als lernförderlich identifiziert, da diese häufig mit verbesserten Testantworten in Verbindung standen (13.2.1). Neben praktischen Phasen scheinen demnach auch theoretische Grundlagen im Seminar bedeutsam, was wiederum zum Schlüsselaspekt 1 des SQD-Modells (Verknüpfung von Theorie und Praxis; 5.2.1) passt. Im Rahmen der Identifikation lernhinderlicher Seminarelemente konnten insbesondere mögliche Probleme oder typische Fehler in der Argumentation der Studierenden identifiziert werden. Häufig wurde in den Fällen, in denen einen Antwort zum Negativen verändert wurde, dies durch eine unpassende Verallgemeinerung von Einzelerfahrungen oder -inhalten aus dem Seminar begründet. So wurden meist positive Erfahrungen mit einem Medium zu typischen Vorteilen verallgemeinert und/oder Distraktoren im Posttest aufgrund dessen nicht ausreichend kritisch eingeschätzt (13.2.2). Aus diesem Ergebnis lässt sich die Relevanz kritischer Reflexionen zu Medieneinsätzen in einer solchen Lehrveranstaltung ableiten, um ein Übergeneralisieren positiver Einzelerfahrungen zu mindern. Die vorliegenden Ergebnisse stützen damit Erkenntnisse bisheriger Forschung hinsichtlich des Erwerbs oder der Entwicklung von TPACK in anderen Fächern. So deutete sich aus verschiedenen Studien bereits an, dass ein deutlicher Praxisbezug (z. B. Max et al., 2022; Tondeur et al., 2012; Zimmermann, 2021) und das Anknüpfen an fachdidaktisches Grundlagenwissen (Backfisch et al., 2021; Zimmermann, 2021) sowie das Sammeln authentischer Erfahrungen mit digitalen Medien (Tondeur et al., 2012) nützlich zur

Förderung von TPACK sind (4.2.2). Die Ergebnisse von Tondeur et al. (2012) bzw. das SQD-Modell legen zudem nahe, dass die Reflexion hinsichtlich des Einsatzes digitaler Medien im Unterricht lernförderlich ist. In der vorliegenden Arbeit wurde dazu passend deutlich, dass eine *fehlende* Reflexion lernhinderlich zu sein scheint (13.2.2). Umgekehrt wurde zwar im Rahmen der Identifikation lernförderlicher Seminarelemente ebenfalls eine kritischere Sichtweise als mögliches lernförderliches Element herausgestellt, jedoch waren ein Großteil dieser Zuordnungen auf einen vermuteten Seminarbezug zurückzuführen und sind aufgrund dieses weniger sicheren Seminarbezugs weniger belastbar (13.2.1). Eine Förderlichkeit der Zusammenarbeit mit Mitstudierenden, welche sich in bisheriger Forschung andeutete (Tondeur et al., 2012; Zimmermann, 2021), konnte in dieser Arbeit nicht direkt identifiziert werden. Möglicherweise steckt dieser Aspekt jedoch indirekt in den (als lernförderlich identifizierten) praktischen Erfahrungen, da diese im Seminar vorwiegend in Partner- oder Gruppenarbeiten erfolgten. Zudem wurden teilweise die Diskussionen im Seminar als lernförderlich deutlich, wobei diesbezüglich im Vergleich zu anderen Seminarelementen wenig Nennungen auftraten (Tabelle 13-2). Die Ergebnisse zur Identifikation lernförderlicher und lernhinderlicher Elemente auf Basis der veränderten Testantworten in dieser Arbeit waren zudem vergleichbar mit den Ergebnissen einer Bachelorarbeit (Veltum, 2023) zur subjektiv wahrgenommenen Lernwirksamkeit einzelner Seminarelemente (13.2.3).

In der vorliegenden Arbeit zeigte sich ein Mehrwert des Triangulationsansatzes (3.1) durch die Kombination der quantitativen Befragung mit den retrospektiven Interviews. So konnte der gemessene Wissenserwerb über das Seminar vertieft betrachtet und aus einer weiteren Perspektive beleuchtet werden, indem die Interviews einen Einblick in die Denkprozesse der Studierenden hinsichtlich ihrer Testantworten und den Veränderungen über die beiden Messzeitpunkte ermöglichten. Der geringe Effekt beim Zuwachs der mittleren Fähigkeiten im gemessenen FDW (11.2) liegt demzufolge nicht unbedingt daran, dass keine Veränderungen stattgefunden haben. Die Interviews verdeutlichen hingegen, dass auf individueller Ebene viele Veränderungen in den Testantworten auftreten, aber dabei eben auch Verschlechterungen in Einzelantworten (z. B. wegen unpassender Verallgemeinerungen; 13.2.2)

vorkommen. Dies erklärt möglicherweise die insgesamt nur geringen Leistungszuwächse, die mit dem schriftlichen Test gemessen werden konnten.

Schlussfolgerungen für die Gestaltung von Seminaren zum Einsatz digitaler Medien im Physikunterricht

Die Verknüpfung der Interviewanalyse mit den quantitativen Testdaten ermöglichte somit auch die Identifikation lernförderlicher und lernhinderlicher Seminarelemente. Auf dieser Basis konnten anschließend Schlussfolgerungen für die Gestaltung entsprechender Lerngelegenheiten zum fachdidaktisch sinnvollen Einsatz digitaler Medien abgeleitet werden (13.3). Diese Hypothesenentwicklung adressierte das fünfte und letzte Ziel der vorliegenden Arbeit (6.2) und stellte auch ein übergeordnetes Ziel im Verbundprojekt DiKoLeP dar (5.1). Es konnten vier Schlussfolgerungen als Gestaltungsmerkmale abgeleitet werden, welche (1) die umfassende kritische Reflexion von Medieneinsätzen und -beispielen, (2) die Möglichkeit des eigenen Ausprobierens und praktischen Implementierens der Medien, (3) die Entwicklung von vernetztem Wissen zu digitalen Medien sowie (4) das Berücksichtigen von Einstellungen und Vorerfahrungen der Studierenden umfassen. Die Gestaltungsmerkmale sind jedoch nicht unabhängig voneinander und insbesondere ist die kritische Reflexion und gemeinsame Diskussion für alle vier Aspekte relevant (13.3).

Im Folgenden werden die Grenzen der vorliegenden Arbeit diskutiert, welche eine Limitation hinsichtlich der Belastbarkeit der beschriebenen Ergebnisse darstellen.

14.2 Grenzen der Arbeit

Zunächst werden Limitationen im Hinblick auf die **Erstellung und den Einsatz des entwickelten Testinstruments** zum FDW zum Einsatz digitaler Medien betrachtet. Die Bearbeitungszeit des Testinstruments ist mit 20-25 Minuten im Vergleich zu anderen TPACK-Messinstrumenten, die Selbsteinschätzungen verwenden, recht lang (3.5.1). Für einen fachdidaktischen Leistungstest hingegen ist eine solche Bearbeitungszeit jedoch gängig bzw. fällt bei offenen Aufgabenformaten häufig noch länger aus (3.4). In der vorliegen-

den Arbeit wurde die längere Bearbeitungszeit bewusst als Kompromiss gewählt, um dafür proximale Leistungstestaufgaben entwickeln zu können und Validitätsprobleme von bisher häufig genutzten Selbsteinschätzungen zu umgehen (6.2).

Die Aufgabenstellungen der entwickelten Testaufgaben sind teilweise textlastig, insbesondere wenn konkrete Unterrichtssituationen als Vignetten eingebunden sind. Dadurch kann es sein, dass einzelne Aufgaben sehr komplex für die Bearbeitenden sind. Ebenso können die einzelnen Antwortalternativen als komplex oder teilweise abstrakt wahrgenommen werden, wodurch es für Teilnehmende schwierig sein kann, eine Ja- oder Nein-Entscheidung zu treffen. Dieses Ja/Nein-Antwortformat kann Teilnehmende zum Raten bzw. zur zufälligen Beantwortung aus Unentschlossenheit verleiten. Mit der gewählten Schwellenbepunktung (10.1) wurde dieser Aspekt adressiert und der Einfluss des Ratens reduziert. Dennoch können die Testergebnisse dadurch hinsichtlich der Validität verzerrt sein. Die Interviewanalyse deutete mit dem Auftreten der Kategorie *Unentschlossenheit* als Grund für veränderte Testantworten darauf hin, dass die Beantwortung der Antwortalternativen teilweise zufällig erfolgt, was jedoch insgesamt selten vorkam (13.1).

Im Hinblick auf das Ergebnis zu den häufig unpassenden Verallgemeinerungen seitens der Studierenden bei der Beantwortung der Antwortalternativen (13.2.2) lässt sich eine weitere mögliche Limitation zum Test ableiten. Möglicherweise verleitet der Test bzw. manche Testaufgaben auch zu (unpassenden) Verallgemeinerungen, da z. B. gerade die allgemeineren Aufgaben ohne Einbettung in eine Unterrichtssituation typische Eigenschaften, Vorteile oder Qualitätsmerkmale eines Mediums adressieren, die nicht unbedingt in jeder Situation allgemeingültig sind. An dieser Stelle lässt sich erneut das Ja/Nein-Antwortformat kritisieren, da möglicherweise nicht für jede einzelne Alternative eine eindeutige und allgemeingültige Ja/Nein-Entscheidung zu treffen ist. Ein alternatives Antwortformat bspw. mit offener Beantwortung und Begründung würde umgekehrt jedoch auch ein Kodiermanual mit eindeutigen Bewertungsregeln erfordern, wann eine Antwort als richtig oder falsch zu bewerten ist, und wäre zusätzlich mit einem hohen zeitlichen und personellen Aufwand verbunden. Insgesamt ist festzuhalten, dass der entwickelte Test

keinem perfekten Messverfahren entspricht. Stattdessen weist er wie sämtliche andere Testinstrumente auch gewisse Grenzen auf, die sich jedoch auch teilweise daraus ergeben, bewusst andere Herausforderungen oder Probleme vermeiden zu können.

Zuletzt ist hinsichtlich der stetigen Weiterentwicklung des Inhalts *digitale Medien* anzumerken, dass die Testinhalte (die Kategorien, aber auch einzelne Aufgabeninhalte) unter Umständen an Aktualität verlieren und somit die Musterlösung zu einzelnen Antwortalternativen an Eindeutigkeit verlieren kann, wenn sich digitale Technologien und ihre Möglichkeiten verändern. Gleichzeitig könnte es in Zukunft notwendig sein, die Testaufgaben zu erweitern, um neue digitale Anwendungen für den Physikunterricht abdecken zu können.

Weiterhin ist auch die **quantitative Evaluation der Seminare** durch verschiedene Aspekte limitiert. Zunächst sollte im Hinblick auf die Implementation des Lehrkonzepts berücksichtigt werden, dass das Lehrkonzept als reine Präsenzveranstaltung geplant wurde, jedoch teilweise pandemiebedingt als Online- oder hybride Lehrveranstaltung stattgefunden hat. Die Umsetzung variierte zudem an den verschiedenen Standorten aufgrund unterschiedlicher Vorgaben während der Pandemie. Somit ist die Vergleichbarkeit der Seminardurchführungen für die betroffenen Semester vom Sommersemester 2021 bis zum Sommersemester 2022 eingeschränkt.

Zusätzlich konnte das Lehrkonzept in den betroffenen Semestern dadurch nicht in dem Format evaluiert werden, für das es ursprünglich designt wurde. So konnten insbesondere die Aspekte der Verzahnung von Theorie und Praxis oder der authentischen Erfahrungen mit digitalen Medien²⁸ des SQD-Modells (5.2.1) in Online-Sitzungen nicht in geplant angemessener Weise umgesetzt werden.

²⁸ Man könnte hier vermuten, dass gerade diese Erfahrungen mit digitalen Medien in Online-Veranstaltungen gut umsetzbar sind. Für rein digitale Anwendungen wie Simulationen oder Interaktive Bildschirmexperimente ist dies auch weitestgehend möglich. Jedoch gibt es auch Situationen z. B. beim Experimentieren mit digitaler Messwerterfassung oder Augmented Reality, in denen der Medieneinsatz eine Unterstützung oder Erweiterung darstellt, aber dennoch physische Experimentiermaterialien und -geräte erforderlich sind. Erfahrungen mit diesen Medien lassen sich demnach nicht kaum in Online-Veranstaltungen umsetzen.

In der quantitativen Prä-Post-Befragung der Hauptstudie konnte mit $N_1 = 70$ vollständigen Datensätzen nur eine recht kleine Stichprobe untersucht werden. Zwar besuchten insgesamt mehr Studierende die untersuchten Lehrveranstaltungen, jedoch nahmen nicht alle an der freiwilligen Befragung bzw. an beiden Messzeitpunkten der Befragung teil. Diese Stichprobengröße erlaubte dadurch auch keine systematischen Vergleiche zwischen den Studierenden der einzelnen Standorte. Dass Studierende an mehreren Standorten und in verschiedenen Semestern teilnahmen, erhöht die Repräsentativität der Stichprobe gegenüber einer Befragung an nur einem Standort. Zudem wiesen die Personenmerkmale der Teilnehmenden keine unerwarteten Auffälligkeiten auf. Allerdings können darüber hinaus keine Aussagen zur Repräsentativität der untersuchten Stichprobe im Hinblick auf die Grundgesamtheit im Lehramt Physik getroffen werden.

Die fehlenden oder unvollständigen Teilnahmen an der Befragung deuten auf eine geringe Testmotivation hin, die sich durch die lange Bearbeitungszeit ergeben haben könnte (neben dem FDW zum Einsatz digitaler Medien wurden im Verbundprojekt mehrere weitere Konstrukte in den Prä-Post-Befragungen erhoben; 5.1). Ein weiterer Grund dafür kann ein ungünstiger, jedoch für die Erhebung sinnvoller Zeitpunkt des Posttests zum Ende der Vorlesungszeit sein, wenn sich die Studierenden üblicherweise auf Klausuren vorbereiten müssen. Unterschiede der Teilnehmenden hinsichtlich ihrer Testmotivation könnten zudem die Ergebnisse verzerrt haben. Weiterhin zeigte sich in der befragten Stichprobe mit 0,47 eine geringe EAP-Reliabilität für den eingesetzten Leistungstest (11.2). Dass die Reliabilität bei dieser kleineren Stichprobengröße geringer ausfällt als in den weiteren Erhebungen mit dem Testinstrument in einer deutlich größeren Stichprobe (8.4) ist zwar nicht verwunderlich, nichtsdestotrotz muss diese Feststellung als Einschränkung der Ergebnisse in der Hauptstudie berücksichtigt werden. Die IRT-Modellierung wies jedoch auch in dieser kleineren Stichprobe eine annehmbare Modellpassung auf und in der Wright-Map wurde die Verteilung der Personenfähigkeiten durch die meisten Aufgaben(-schwellen) mit passender Schwierigkeit abgedeckt (11.2).

Abschließend muss für die quantitative Evaluation angemerkt werden, dass es keine Kontrollgruppe zum Vergleich gab, die ohne Besuch des Seminars

im Untersuchungszeitraum befragt wurde. Der gemessene Zuwachs im FDW ist demnach auch nicht ausschließlich auf das Seminar zurückzuführen, da noch andere Lerngelegenheiten in diesem Zeitraum für die Wissensentwicklung bedeutsam gewesen sein könnten. Die retrospektiven Interviews geben an dieser Stelle jedoch vertiefende Hinweise darauf, dass ein Großteil der Veränderungen in den Testantworten auch mit dem Seminar zusammenzuhängen scheint.

Schließlich sollten auch die Grenzen der **qualitativen Interviewstudie** in dieser Arbeit berücksichtigt werden. Die 19 Interviewteilnehmenden stellen alle diejenigen Personen aus der Stichprobe zur quantitativen Befragung dar, die sich zusätzlich freiwillig für ein Interview bereit erklärten (12.1). Dementsprechend liegt es nahe, dass es sich bei der Stichprobe zu den Interviews möglicherweise um eine weniger repräsentative Auswahl handeln könnte. Zur Bewertung der Repräsentativität der Interviewstichprobe hinsichtlich der übergeordneten Stichprobe der Evaluation wurden die deskriptiven Personenmerkmale der Gruppen gegenübergestellt. In dieser Hinsicht waren die Interviewteilnehmenden in den meisten Aspekten – bis auf die Verteilung zum Master- bzw. Bachelorstudium – vergleichbar mit der übergeordneten Stichprobe. Jedoch erreichten die Interviewteilnehmenden im Mittel bessere Leistungen im gemessenen FDW, sodass dahingehend eine gewisse Positivauswahl in der Interviewstichprobe vorlag (12.1). Die Repräsentativität der Interviewteilnehmenden bzgl. der Grundgesamtheit im Lehramt Physik kann in der vorliegenden Arbeit wiederum nicht bewertet werden.

Als weitere Limitation ist anzuführen, dass die Erkenntnisse aus den Interviews auf subjektiven Einschätzungen der Studierenden beruhen. Diese müssen nicht zwingend der wahren Ursache für eine Veränderung entsprechen, sondern können bspw. durch soziale Erwünschtheit beeinflusst oder verzerrt worden sein. So könnte das Seminar häufiger als Grund für eine veränderte Antwort genannt worden sein, weil es den Kontext des Interviews darstellte und somit – bewusst oder unbewusst – als möglicher Grund von den Studierenden wahrgenommen wurde. Ebenso kann der Einfluss durch die Person der Autorin, welche die Interviews geführt hat, eine Rolle in den Interviews gespielt haben, wenngleich dies nicht bewusst erfolgte. Die Stu-

dierenden wurden zu Beginn der Interviews daher explizit darauf hingewiesen, dass sie im Interview ehrlich antworten können und sollen (Anhang E). Nichtsdestotrotz lässt sich eine solche Beeinflussung der Interviewteilnehmenden nicht ganz ausschließen.

Die Interviews zeigten in Bezug auf die Gründe für Testveränderungen außerdem ein Problem hinsichtlich des Testinstruments auf. Es gab Fälle, in denen eine Antwortalternative im Prätest richtig beantwortet wurde, z. B. wurde ein Distraktor richtigerweise abgelehnt. Im Interview wurde aus der Begründung der veränderten Antwort jedoch deutlich, dass diese richtige Beantwortung im Prätest nicht auf einer fachdidaktischen reflektierten Überlegung beruhte, sondern eher auf ein fehlendes technisches Wissen zurückzuführen war, was mit diesem Medium (aus technischer Sicht) möglich ist. Dies stellt eine gewisse Einschränkung in der Validität dar, wenngleich diese Fälle insgesamt selten auftraten (Tabelle 13-3). Umgekehrt erlaubt die Verknüpfung der quantitativen Befragung mit den retrospektiven Interviews in dieser Arbeit, ebensolche Fälle aufzudecken und Hinweise dafür zu erhalten, dass es sich nicht um ein systematisch und/oder häufig auftretendes Problem handelt.

In der Analyse erster Interviews wurde ein weiteres Problem deutlich, welches sich in der Häufigkeit der Subkategorie *keine Änderung begründet* äußert, wenn Studierende Aussagen zu veränderten Testantworten machten, in denen sie letztendlich nicht die *Veränderung* der Antwort begründeten (13.1). Diese fehlenden Begründungen traten v. a. in den Interviews im ersten Durchgang häufiger auf. In späteren Interviews konnte durch gezieltes Nachfragen etwas nachgesteuert werden, sodass wirklich die Änderung begründet wurde. Weiterhin zeigte sich auch eine hohe Häufigkeit in der Subkategorie *nicht nutzbare Textstellen*, wozu die Unterkategorien *keine Erinnerung* (an die Testbeantwortung) und *Sonstiges* zählten. Das Auftreten solcher „Sonstiges-Kategorien“ ist in der qualitativen Inhaltsanalyse kaum vermeidbar. Anders als bei den Zuordnungen zur Subkategorie *keine Änderung begründet* konnte jedoch kein systematisches Problem herausgestellt werden, welches in späteren Interviews hätte adressiert werden können. Bei diesen nicht nutzbaren Textstellen war auffällig, dass deren Häufigkeit sich zwischen einzelnen Befragungspersonen stark unterschied. Demnach stellte

dies kein Phänomen war, das über alle Interviews gleichermaßen auftrat. So gab es drei Befragungspersonen, in denen diese Subkategorie gar nicht auftrat und weitere zwei Personen, bei denen die Subkategorie maximal 10 % bezogen auf die Kodierungen in der Hauptkategorie ausmachte. Umgekehrt gab es drei andere Teilnehmende, bei denen über 50 % der Kodierungen in der Hauptkategorie als nicht nutzbare Textstellen kodiert wurden. Es scheint demnach so, dass die Interviewteilnehmenden sich teilweise unterschiedlich gut an ihre Beantwortung der Testaufgaben erinnern konnten sowie unterschiedlich gut mit dem Setting im Interview zurechtkamen, Begründungen für ihre veränderten Antworten zu explizieren.

Die aus der Interviewanalyse gewonnenen Ergebnisse zur Identifikation lernförderlicher oder lernhinderlicher Seminarelemente sind auch durch die Testinhalte sowie insbesondere die Auswahl der Testaufgaben für das Interview beeinflusst. Wenn in einem Interview mehr Aufgaben zur Kategorie *digitale Messwerterfassung* vorgelegt wurden, konnten zu diesem Themenbereich auch mehr lernförderliche Seminarelemente identifiziert werden (13.2.1). Ebenso ist die Identifikation der Seminarelemente davon abhängig, wie häufig es überhaupt Verbesserungen in den einzelnen Kategorien gab. So erklärt die Tatsache, dass die Segmente zur Kombination von Seminarbezug und Verbesserung häufiger zu Aufgaben der Kategorien *digitale Messwerterfassung* oder *Erklärvideos* zählten, womöglich auch, warum zu diesen Themen mehr lernförderliche Seminarelemente identifiziert wurden (13.2.1). Zu berücksichtigen ist für die Analyse in dieser Arbeit auch, dass es nicht möglich war, Seminarelemente oder -inhalte als lernförderlich zu identifizieren, die nicht in den Testaufgaben thematisiert werden. Dies liegt daran, dass das entwickelte Testinstrument zum FDW zum Einsatz digitaler Medien nicht alles abdecken kann, was im Seminar thematisiert und möglicherweise gelernt wird.

Insgesamt werden zu verschiedenen Aspekten und Teilstudien der vorliegenden Arbeit mehrere Limitationen deutlich. Nichtsdestotrotz leistet die vorliegende Arbeit auch unter Berücksichtigung der beschriebenen Grenzen einen Beitrag zur Adressierung der herausgestellten Desiderate im Hinblick auf die Messung von FDW zum Einsatz digitaler Medien im Fach Physik und der Evaluation entsprechender fachdidaktischer Lerngelegenheiten. Im folgenden

Abschnitt wird ein Ausblick auf mögliche Anknüpfungspunkte zu dieser Arbeit für die Forschung und Lehrkräftebildung gegeben.

14.3 Ausblick

Die gewonnenen Erkenntnisse der vorliegenden Arbeit sind einerseits hinsichtlich der Gestaltung und Implementation von Lerngelegenheiten zum Einsatz digitaler Medien nutzbar sowie andererseits für die Erfassung des FDW von Physiklehramtsstudierenden in diesem Bereich.

Das in dieser Arbeit untersuchte Lehrkonzept zur Förderung des FDW zum Einsatz digitaler Medien im Fach Physik wurde an mehreren Standorten erprobt und evaluiert, wobei der Wissenszuwachs über die Seminare auf die Wirksamkeit des Lehrkonzepts hindeutet. Die gelungene Implementation des Konzepts an die Rahmenbedingungen von insgesamt fünf Standorten legt nahe, dass eine Übertragung auf weitere Standorte möglich wäre. Außerdem konnten im Rahmen der Hypothesenentwicklung in dieser Arbeit Merkmale für lernförderliche Lerngelegenheiten zu digitalen Medien herausgearbeitet werden, die als Gestaltungswissen für solche Lerngelegenheiten in der fachdidaktischen Lehrkräftebildung genutzt werden können.

Als Anknüpfungspunkt für weitere Forschung könnten diese gewonnen Hypothesen bzw. Gestaltungsmerkmale systematisch untersucht werden. Es könnten Lehrveranstaltungen zum Einsatz digitaler Medien implementiert und beforscht werden, die die abgeleiteten Gestaltungsmerkmale explizit berücksichtigen (z. B. durch systematische Einbindung von Phasen zur kritischen Reflexion und Diskussion sowie Einbettung von Negativbeispielen für Medieneinsätze) im Vergleich zu Lehrveranstaltungen, die z. B. keine Reflexionsphasen beinhalten.

Für das Testinstrument sind weiterhin Untersuchungen zur Übertragbarkeit auf andere Zielgruppen denkbar, da bisher nur Lehramtsstudierende den Test bearbeitet haben. Beispielsweise könnte die Nutzbarkeit des Tests bei Lehramtsanwärter:innen oder praktizierenden Lehrkräften in weiteren Studien untersucht werden. Im Hinblick auf die Ergebnisse zur Untersuchung der diskriminanten Validität in dieser Arbeit (9.2) könnte im Rahmen weiterer

Forschung untersucht werden, inwiefern eine erneute Erhebung in einer größeren Stichprobe mögliche andere Erkenntnisse liefert. Insbesondere könnten auch Lehramtsstudierende weiterer Fächer (außer Physik) eingebunden werden, um mögliche Gruppenunterschiede in den Testleistungen von Physiklehramtsstudierenden gegenüber anderen MINT- oder nicht-MINT-Lehramtsstudierenden zu untersuchen. Die Ergebnisse könnten Hinweise darauf liefern, ob mit dem Test womöglich ein MINT-spezifisches Wissen erfasst wird. Weiterhin könnte im Hinblick auf weitere Fächer untersucht werden, inwiefern eine Adaption der Testaufgaben z. B. auf die naturwissenschaftlichen Fächer Biologie oder Chemie möglich ist. Die modellierten Kategorien scheinen zunächst auch für diese Fächer bedeutsam, sodass eine grundsätzliche Passung des Tests und seiner Struktur denkbar ist. Auf diese Weise könnte in anknüpfender Forschung untersucht werden, wie umfangreich sich die Adaption einzelner Aufgaben gestaltet (z. B. Anpassung der Unterrichtssituation in der Aufgabenstellung, Umformulierung einzelner Antwortalternativen) und wie gut die adaptierten Aufgaben in Erhebungen in den entsprechenden Zielgruppen funktionieren. Auch wenn in diesen beiden Fächern im deutschsprachigen Raum bereits leistungsorientierte Messverfahren (zur Erfassung von TPACK) entwickelt und eingesetzt wurden (3.5.2), stellten diese recht bearbeitungs- und auswertungsaufwändige Verfahren wie die Untersuchung von TPACK-Wissensbereichen über Unterrichtsplanungen (Zimmermann, 2021; Chemie) oder Leistungstests mit offenen Aufgabenformaten (Kotzebue, 2022a, 2022b; Max, 2022; Max et al., 2020; Biologie) dar. Die Adaption des vorliegenden geschlossenen Testinstruments auf diese Fächer könnte dementsprechend einen Mehrwert im Sinne einer zeitökonomischeren Erfassung und Auswertung von FDW zum Einsatz digitaler Medien bzw. TPACK bieten.

Durch den in dieser Arbeit entwickelten Leistungstest steht nun ein proximales Messverfahren zur Erfassung des FDW zum Einsatz digitaler Medien im Fach Physik als Alternative zu üblicherweise genutzten Selbsteinschätzungen zur Verfügung. Das Testinstrument ist hinsichtlich verschiedener Validitätsaspekte untersucht worden und stellt mit dem geschlossenen Aufgabenformat einen Kompromiss im Hinblick auf die Bearbeitungsdauer und objek-

tive Auswertbarkeit bei gleichzeitig möglichst valider Messung dar. Der entwickelte Leistungstest wurde bereits an anderen Hochschulstandorten außerhalb des Verbundprojekts DiKoLeP eingesetzt und kann auch zukünftig bspw. zur Forschung oder zur Evaluation der Lehre genutzt werden.

15 Literaturverzeichnis

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211.
[https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Akyuz, D. (2018). Measuring technological pedagogical content knowledge (TPACK) through performance assessment. *Computers & Education*, 125, 212–225. <https://doi.org/10.1016/j.compedu.2018.06.012>
- Anderson, L. W. & Krathwohl, D. R. (Hrsg.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Complete ed.). Longman.
- Aufschnaiter, C. von & Blömeke, S. (2010). Professionelle Kompetenz von (angehenden) Lehrkräften erfassen–Desiderata. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 361–367.
- Backfisch, I., Lachner, A., Hische, C., Loose, F. & Scheiter, K. (2020a). Professional knowledge or motivation? Investigating the role of teachers' expertise on the quality of technology-enhanced lesson plans. *Learning and instruction*, 66, 101300. <https://doi.org/10.1016/j.learninstruc.2019.101300>
- Backfisch, I., Lachner, A., Stürmer, K. & Scheiter, K. (2021). Gelingensbedingungen beim Einsatz digitaler Medien im Unterricht – Kognitive und motivationale Voraussetzungen von Lehrpersonen. In *Vielfältig herausgefordert. Forschungs- und Entwicklungsfelder der Lehrerbildung auf dem Prüfstand*.
- Backfisch, I., Schneider, J., Lachner, A., Scheiter, K. & Scherer, R. (2020b). *Another jingle-jangle fallacy? Examining the validity of Technological Pedagogical and Content Knowledge (TPACK) self-report assessments*. <https://doi.org/10.23668/PSYCHARCHIVES.4226>
- Baer, M., Dörr, G., Fraefel, U., Kocher, M., Küster, O., Larcher, S., Müller, P., Sempert, W. & Wyss, C. (2007). Werden angehende Lehrpersonen durch das Studium kompetenter? Kompetenzaufbau und Standardreichung in der berufswissenschaftlichen Ausbildung an drei Pädagogischen Hochschulen in der Schweiz und in Deutschland. *Unterrichtswissenschaft*, 35(1), 15–47. <https://doi.org/10.25656/01:5485>

- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520. <https://doi.org/10.1007/s11618-006-0165-2>
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M. & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American educational research journal*, 47(1), 133–180.
- Baur, N. & Blasius, J. (2014). Methoden der empirischen Sozialforschung: Ein Überblick. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 41–62). Springer VS.
- Becker, S., Bruckermann, T., Finger, A., Huwer, J., Kremser, E., Meier, M., Thoms, L.-J., Thyssen, C. & Kotzebue, L. von. (2020). Orientierungsrahmen Digitale Kompetenzen für das Lehramt in den Naturwissenschaften - DiKoLAN. In S. Becker, J. Meßinger-Koppelt & C. Thyssen (Hrsg.), *Digitale Basiskompetenzen: Orientierungshilfe und Praxisbeispiele für die universitäre Lehramtsausbildung in den Naturwissenschaften* (S. 14–43). Joachim Herz Stiftung.
- Becker, S., Klein, P., Göbbling, A. & Kuhn, J. (2019). Förderung von Konzeptverständnis und Repräsentationskompetenz durch Tablet-PC-gestützte Videoanalyse. *Zeitschrift für Didaktik der Naturwissenschaften*, 25(1), 1–24. <https://doi.org/10.1007/s40573-019-00089-4>
- Bernsteiner, A., Schubatzky, T., Haagen-Schützenhöfer, C. & Spitzer, P. (2024). Impact of working with Arduino on mathematics and science teacher students' self-assessment of TPACK and self-efficacy. *Journal of Physics: Conference Series*, 2750(1), 12043. <https://doi.org/10.1088/1742-6596/2750/1/012043>
- Bilici, S. C., Yamak, H., Kavak, N. & Guzey, S. S. (2013). Technological Pedagogical Content Knowledge Self-Efficacy Scale (TPACK-SeS) for Pre-Service Science Teachers: Construction, Validation, and Reliability. *Eurasian Journal of Educational Research*, 52, 37–60.
- Blake, C. & Scanlon, E. (2007). Reconsidering simulations in science education at a distance: features of effective use. *Journal of Computer Assisted Learning*, 23(6), 491–502. <https://doi.org/10.1111/j.1365-2729.2007.00239.x>

- Blömeke, S. (2013). *Validierung als Aufgabe im Forschungsprogramm „Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor“*. (KoKoHs Working Papers, 2). Humboldt-Universität & Johannes Gutenberg-Universität.
- Blömeke, S., Gustafsson, J.-E. & Shavelson, R. J. (2015). Beyond dichotomies. *Zeitschrift für Psychologie*.
- Blömeke, S., Jentsch, A., Ross, N., Kaiser, G. & König, J. (2022). Opening up the black box: Teacher competence, instructional quality, and students' learning progress. *Learning and instruction*, 79, 101600. <https://doi.org/10.1016/j.learninstruc.2022.101600>
- Blömeke, S., Kaiser, G. & Lehmann, R. (2008). *Professionelle Kompetenz angehender Lehrerinnen und Lehrer: Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare: erste Ergebnisse zur Wirksamkeit der Lehrerausbildung*. Waxmann.
- Blömeke, S. & König, J. (2010). Messung des pädagogischen Wissens: Theoretischer Rahmen und Teststruktur. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *TEDS-M 2008: Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich* (S. 239–264). Waxmann.
- Boone, W. J., Staver, J. R. & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer Netherlands. <https://doi.org/10.1007/978-94-007-6857-4>
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7., vollständig überarbeitete und erweiterte Auflage). Springer. <https://doi.org/10.1007/978-3-642-12770-0>
- Brandt, H. & Moosbrugger, H. (2020). Planungsaspekte und Konstruktionsphasen von Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3., vollständig neu bearbeitete, erweiterte und aktualisierte Auflage, S. 39–66). Springer. https://doi.org/10.1007/978-3-662-61532-4_3
- Brennan, R. L. & Prediger, D. J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and psychological measurement*, 41(3), 687–699.
- Buchholtz, N. (2021). Voraussetzungen und Qualitätskriterien von Mixed-Methods-Studien in der mathematikdidaktischen Forschung. *Journal*

für *Mathematik-Didaktik*, 42(1), 219–242.

<https://doi.org/10.1007/s13138-020-00173-0>

Bundesministerium für Bildung und Forschung (2016). *Neue Wege in der Lehrerbildung: Die Qualitätsoffensive Lehrerbildung*. Zugegriffen am 27.05.2024 unter https://www.qualitaetsoffensive-lehrerbildung.de/lehrerbildung/de/aktuelles/publikationen-der-programmbegleitung/publikationen-der-programmbegleitung_node.html

Bundesministerium für Bildung und Forschung (2018). *Bekanntmachung: Richtlinie zur Förderung von Projekten in der „Qualitätsoffensive Lehrerbildung“ mit den Schwerpunkten „Digitalisierung in der Lehrerbildung“ und/oder „Lehrerbildung für die beruflichen Schulen“*. Bundesministerium für Bildung und Forschung (BMBF). Zugegriffen am 17.07.2023 unter https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2018/11/2097_bekanntmachung

Bundesministerium für Bildung und Forschung (2023). *Qualitätsoffensive Lehrerbildung: Ein Blick zurück, ein Blick nach vorn*. Zugegriffen am 27.05.2024 unter https://www.qualitaetsoffensive-lehrerbildung.de/lehrerbildung/de/aktuelles/publikationen-der-programmbegleitung/publikationen-der-programmbegleitung_node.html

Carlson, J., Cooper, R., Daehler, K. R., Friedrichsen, P. J., Heller, J. I., Kirschner, S., Elliott, N. L., Marangio, K. & Wong, N. (2019a). Vignettes Illustrating Practitioners' and Researchers' Applications of the Refined Consensus Model of Pedagogical Content Knowledge. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 95–115). Springer. https://doi.org/10.1007/978-981-13-5898-2_3

Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Kam Ho Chan, K., Cooper, R., Friedrichsen, P. J., Gess-Newsome, J., Henze-Rietveld, I., Hume, A., Kirschner, S., Liepertz, S., Loughran, J., Mavhunga, E., Neumann, K., Nilsson, P., . . . Wilson, C. D. (2019b). The Refined Consensus Model of Pedagogical Content Knowledge in Science Education. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 77–94). Springer. https://doi.org/10.1007/978-981-13-5898-2_2

- Cauet, E., Liepertz, S., Borowski, A. & Fischer, H. E. (2015). Does it matter what we measure? Domain-specific professional knowledge of physics teachers. *Schweizerische Zeitschrift für Bildungswissenschaften*, 37(3), 462–479. <https://doi.org/10.25656/01:12746>
- Chai, C. S., Koh, J. H. L. & Tsai, C.-C. (2013a). A review of technological pedagogical content knowledge. *Journal of Educational Technology & Society*, 16(2), 31–51.
- Chai, C. S., Koh, J. H. L., Tsai, C.-C. & Lee Wee Tan, L. (2011). Modeling primary school pre-service teachers' Technological Pedagogical Content Knowledge (TPACK) for meaningful learning with information and communication technology (ICT). *Computers & Education*, 57(1), 1184–1193. <https://doi.org/10.1016/j.compedu.2011.01.007>
- Chai, C. S., Ng, E. M. W., Li, W., Hong, H.-Y. & Koh, J. H. L. (2013b). Validating and modelling technological pedagogical content knowledge framework among Asian preservice teachers. *Australasian Journal of Educational Technology*, 29(1). <https://doi.org/10.14742/ajet.174>
- Chan, K. K. H. & Hume, A. (2019). Towards a Consensus Model: Literature Review of How Science Teachers' Pedagogical Content Knowledge Is Investigated in Empirical Studies. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 3–76). Springer. https://doi.org/10.1007/978-981-13-5898-2_1
- Chi, M. T. H. & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37–46.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Lawrence Erlbaum Associates.
- Conrads, T. (2022). *Kognitive Validierung eines physikdidaktischen Leistungstests zum Einsatz digitaler Medien und Vorbereitung einer Adaption für das Fach Biologie* [Unveröffentlichte Masterarbeit]. RWTH Aachen University.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.

- Davis, F. D., Bagozzi, R. P. & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management science*, 35(8), 982–1003.
- Dollny, S. (2011). *Entwicklung und Evaluation eines Testinstruments zur Erfassung des fachspezifischen Professionswissens von Chemielehrkräften. Studien zum Physik- und Chemielernen: Bd. 127*. Logos Verlag.
- Döring, N. & Bortz, J. (2016a). Datenerhebung. In N. Döring & J. Bortz (Hrsg.), *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (S. 321–577). Springer.
https://doi.org/10.1007/978-3-642-41089-5_10
- Döring, N. & Bortz, J. (2016b). Operationalisierung. In N. Döring & J. Bortz (Hrsg.), *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (S. 222–289). Springer.
- Duit, R. & Mikelskis-Seifert, S. (2010). Kontextorientierter Physikunterricht: PiKo-Brief Nr. 5. *PiKo-Briefe. Der fachdidaktische Forschungsstand kurzgefasst*, 1–4.
- Ericsson, K. A. & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215–251. <https://doi.org/10.1037/0033-295X.87.3.215>
- Finkenbergh, F. (2018). *Flipped Classroom im Physikunterricht. Studien zum Physik- und Chemielernen: Bd. 260*. Logos Verlag.
<https://doi.org/10.30819/4737>
- Flick, U. (2011). *Triangulation: Eine Einführung* (3., aktualisierte Auflage). *Qualitative Sozialforschung: Bd. 12*. VS Verlag für Sozialwissenschaften; Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-531-92864-7>
- Förtsch, C., Werner, S., Kotzebue, L. von & Neuhaus, B. J. (2016). Effects of biology teachers' professional knowledge and cognitive activation on students' achievement. *International Journal of Science Education*, 38(17), 2642–2666. <https://doi.org/10.1080/09500693.2016.1257170>
- Frey, A. (2006). Methoden und Instrumente zur Diagnose beruflicher Kompetenzen von Lehrkräften. Eine erste Standortbestimmung zu bereits publizierten Instrumenten. *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern, Zeitschrift für Pädagogik, Beiheft 51*, 30–46. <https://doi.org/10.25656/01:7369>

- Gäde, J. C., Schermelleh-Engel, K. & Brandt, H. (2020). Konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3., vollständig neu bearbeitete, erweiterte und aktualisierte Auflage, S. 615–660). Springer.
- Gess-Newsome, J. (1999). Pedagogical Content Knowledge: An Introduction and Orientation. In J. Gess-Newsome & N. G. Lederman (Hrsg.), *Examining pedagogical content knowledge* (S. 3–17). Kluwer Academic Publishers.
- Gess-Newsome, J. (2015). A Model of Teacher Professional Knowledge and Skill Including PCK: Results of the thinking from the PCK Summit. In A. Berry, P. J. Friedrichsen & J. Loughran (Hrsg.), *Teaching and Learning in Science Series. Re-examining pedagogical content knowledge in science education* (S. 28–42). Routledge.
- Ghassemi, N. (2023). *Evaluation eines Lehramtmasterstudiengangs mit dem Profil Quereinstieg im Fach Physik. Studien zum Physik- und Chemielernen: Bd. 376*. Logos Verlag.
- Girwicz, R. (2013). Lernen mit bewegten Bildern: Animationen und dynamische Visualisierungen im Physikunterricht. *Naturwissenschaften im Unterricht: Physik*, 24(137), 4–9.
- Girwicz, R. (2020a). Experimente im Physikunterricht. In E. Kircher, R. Girwicz & H. E. Fischer (Hrsg.), *Physikdidaktik: Grundlagen* (4. Aufl., S. 263–291). Springer.
- Girwicz, R. (2020b). Medien im Physikunterricht. In E. Kircher, R. Girwicz & H. E. Fischer (Hrsg.), *Physikdidaktik: Grundlagen* (4. Aufl., S. 293–335). Springer.
- Girwicz, R. (2020c). Multimedia und digitale Medien im Physikunterricht. In E. Kircher, R. Girwicz & H. E. Fischer (Hrsg.), *Physikdidaktik: Grundlagen* (4. Aufl., S. 457–527). Springer.
- Gläser-Zikuda, M. (2013). *Qualitative Inhaltsanalyse in der Bildungsforschung–Beispiele aus diversen Studien*. Peter Lang.
- Gonzalez, O., MacKinnon, D. P. & Muniz, F. B. (2021). Extrinsic Convergent Validity Evidence to Prevent Jingle and Jangle Fallacies. *Multivariate behavioral research*, 56(1), 3–19.
<https://doi.org/10.1080/00273171.2019.1707061>

- Gramzow, Y. (2015). *Fachdidaktisches Wissen von Lehramtsstudierenden im Fach Physik: Modellierung und Testkonstruktion. Studien zum Physik- und Chemielernen: Bd. 181*. Logos Verlag.
- Gramzow, Y., Riese, J. & Reinhold, P. (2013). Modellierung fachdidaktischen Wissens angehender Physiklehrkräfte. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 7–30.
- Gronlund, N. E. (1974). *Die Anlage von Leistungstests: constructing achievement tests*. aus dem Englischen übersetzt. *Diesterwegs rote Reihe*. Diesterweg.
- Große-Heilmann, R., Burde, J.-P., Riese, J., Schubatzky, T. & Weiler, D. (2024). Entwicklung fachdidaktischen Wissens zum Einsatz digitaler Medien bei Lehramtsstudierenden im Fach Physik. In B. Herzig, B. Eickelmann, F. Schwabl, J. Schulze & J. Niemann (Hrsg.), *Lehrkräftebildung in der digitalen Welt*. Waxmann.
<https://doi.org/10.31244/9783830998372>
- Große-Heilmann, R., Riese, J., Burde, J.-P., Schubatzky, T. & Weiler, D. (2021). Erwerb und Messung physikdidaktischer Kompetenzen zum Einsatz digitaler Medien. *PhyDid B - Didaktik der Physik - Beiträge zur DPG-Frühjahrstagung 2021 virtuell*, 171–178.
- Große-Heilmann, R., Riese, J., Burde, J.-P., Schubatzky, T. & Weiler, D. (2022). Fostering Pre-Service Physics Teachers' Pedagogical Content Knowledge Regarding Digital Media. *Education Sciences*, 12(7), 440.
<https://doi.org/10.3390/educsci12070440>
- Hamilton, E. R., Rosenberg, J. M. & Akcaoglu, M. (2016). The Substitution Augmentation Modification Redefinition (SAMR) Model: a Critical Review and Suggestions for its Use. *TechTrends*, 60(5), 433–441.
<https://doi.org/10.1007/s11528-016-0091-y>
- Harris, J., Grandgenett, N. & Hofer, M. (2010). Testing a TPACK-Based Technology Integration Assessment Rubric. In D. Gibson (Hrsg.), *Society for Information Technology & Teacher Education 2010 21st international conference* (S. 3833–3840). Association for the Advancement of Computing in Education (AACE).
- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 143–171). Springer. https://doi.org/10.1007/978-3-642-20072-4_7

- Hartig, J., Frey, A. & Jude, N. (2020). Validität von Testwertinterpretationen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3., vollständig neu bearbeitete, erweiterte und aktualisierte Auflage, S. 529–546). Springer.
- Hartig, J. & Jude, N. (2007). Empirische Erfassung von Kompetenzen und psychometrische Kompetenzmodelle. *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik*, 17–36.
- Helfferrich, C. (2011). *Die Qualität qualitativer Daten: Manual für die Durchführung qualitativer Interviews* (4. Auflage). VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-92076-4>
- Henne, A., Möhrke, P., Thoms, L.-J. & Huwer, J. (2022). Implementing Digital Competencies in University Science Education Seminars Following the DiKoLAN Framework. *Education Sciences*, 12(5), 356. <https://doi.org/10.3390/educsci12050356>
- Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I. & Reiss, K. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, 153, 103897. <https://doi.org/10.1016/j.compedu.2020.103897>
- Hoyer, C. & Girwidz, R. (2019). Digitale Medien - Werkzeuge beim Experimentieren: Schlüsselexperimente und digitale Medien. *Naturwissenschaften im Unterricht: Physik*, 30(171/172), 13–19.
- Hume, A., Cooper, R. & Borowski, A. (Hrsg.). (2019). *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science*. Springer. <https://doi.org/10.1007/978-981-13-5898-2>
- Joo, Y. J., Park, S. & Lim, E. (2018). Factors influencing preservice teachers' intention to use technology: TPACK, teacher self-efficacy, and technology acceptance model. *Journal of Educational Technology & Society*, 21(3), 48–59.
- Jordans, M., Zeller, J., Große-Heilmann, R. & Riese, J. (2022). Weiterentwicklung eines physikdidaktischen Tests zum Online-Assessment. In S. Habig & H. van Vorst (Hrsg.), *Gesellschaft für Didaktik der Chemie und Physik Jahrestagung virtuell 2021*.

- Joswig, A.-K. & Riese, J. (2019). Veränderungen des physikdidaktischen Wissens im Verlauf eines Lehr-Lern-Seminars. *PhyDid B - Didaktik der Physik - Beiträge zur DPG-Frühjahrstagung 2019 in Aachen*, 183–187.
- Joswig-Käfer, A.-K. (2024). *Analysen zu Veränderungen des physikdidaktischen Wissens von Lehramtsstudierenden: Qualitative und quantitative Untersuchungen im Verlauf eines Lehr-Lern-Seminars* [Dissertation]. RWTH Aachen University. <https://doi.org/10.18154/RWTH-2024-01330>
- Joswig-Käfer, A.-K. & Riese, J. (2021). Längsschnitt Physikdidaktischen Wissens: Ursachen für Veränderungen. In S. Habig (Hrsg.), *Naturwissenschaftlicher Unterricht und Lehrerbildung im Umbruch? Gesellschaft für Didaktik der Chemie und Physik Online-Jahrestagung 2020*.
- Kaiser, G., Bremerich-Vos, A. & König, J. (2020). Professionswissen. In C. Cramer, J. König, M. Rothland & S. Blömeke (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 811–818). Verlag Julius Klinkhardt.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological bulletin*, 112(3), 527–535.
- Kane, M. T. (2001). Current Concerns in Validity Theory. *Journal of Educational Measurement*, 38(4), 319–342. <https://doi.org/10.1111/j.1745-3984.2001.tb01130.x>
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73.
- Kelava, A. & Moosbrugger, H. (2020a). Deskriptivstatistische Itemanalyse und Testwertbestimmung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3., vollständig neu bearbeitete, erweiterte und aktualisierte Auflage, S. 143–169). Springer.
- Kelava, A. & Moosbrugger, H. (2020b). Einführung in die Item-Response-Theorie (IRT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3., vollständig neu bearbeitete, erweiterte und aktualisierte Auflage, S. 369–409). Springer. https://doi.org/10.1007/978-3-662-61532-4_16
- Kelle, U. (2014). Mixed Methods. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 153–166). Springer VS.
- Keller, M., Neumann, K. & Fischer, H. E. (2017). The impact of physics teachers' pedagogical content knowledge and motivation on students'

- achievement and interest. *Journal of Research in Science Teaching*, 54(5), 586–614. <https://doi.org/10.1002/tea.21378>
- Kircher, E., Girwidz, R. & Häußler, P. (Hrsg.). (2009). *Physikdidaktik: Theorie und Praxis* (2. Aufl.). Springer. <https://doi.org/10.1007/978-3-642-01602-8>
- Kirschner, S. (2013). *Modellierung und Analyse des Professionswissens von Physiklehrkräften. Studien zum Physik- und Chemielernen: Bd. 161*. Logos Verlag.
- Kirschner, S., Borowski, A., Fischer, H. E., Gess-Newsome, J. & Aufschnaiter, C. von (2016). Developing and evaluating a paper-and-pencil test to assess components of physics teachers' pedagogical content knowledge. *International Journal of Science Education*, 38(8), 1343–1372. <https://doi.org/10.1080/09500693.2016.1190479>
- Kirstein, J. & Nordmeier, V. (2013). Interaktive Realbild-Animationen: Experimentier- und Lernmöglichkeiten mit dem Computer. *Naturwissenschaften im Unterricht: Physik*, 24(137), 37–41.
- Klieme, E. & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik*, 52(6), 876–903. <https://doi.org/10.25656/01:4493>
- Klinzing, H. G. (2002). Wie effektiv ist Microteaching? Ein Überblick über fünfunddreißig Jahre Forschung. *Zeitschrift für Pädagogik*, 48(2), 194–214. <https://doi.org/10.25656/01:3829>
- Koehler, M. J. & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary issues in technology and teacher education*, 9(1), 60–70.
- Koehler, M. J., Mishra, P. & Cain, W. (2013). What is technological pedagogical content knowledge (TPACK)? *Journal of Education*, 193(3), 13–19.
- Koehler, M. J., Shin, T. S. & Mishra, P. (2012). How do we measure TPACK? Let me count the ways. In *Educational technology, teacher knowledge, and classroom impact: A research handbook on frameworks and approaches* (S. 16–31). IGI Global.

- Konrad, U. (2019). Elektrizitätslehre: Geräte und Materialien für den Unterricht in der Sekundarstufe I. *Naturwissenschaften im Unterricht: Physik*, 30(171/172), 20–24.
- Kopcha, T. J., Ottenbreit-Leftwich, A., Jung, J. & Baser, D. (2014). Examining the TPACK framework through the convergent and discriminant validity of two measures. *Computers & Education*, 78, 87–96.
<https://doi.org/10.1016/j.compedu.2014.05.003>
- Korneck, F., Lamprecht, J., Wodzinski, R. & Schecker, H. (2010). *Quereinsteiger in das Lehramt Physik-Lage und Perspektiven der Physiklehrerbildung in Deutschland; eine Studie der Deutschen Physikalischen Gesellschaft*. Deutsche Physikalische Gesellschaft eV.
- Kotzebue, L. von (2022a). Beliefs, Self-reported or Performance-Assessed TPACK: What Can Predict the Quality of Technology-Enhanced Biology Lesson Plans? *Journal of Science Education and Technology*, 31(5), 570–582. <https://doi.org/10.1007/s10956-022-09974-z>
- Kotzebue, L. von (2022b). Two is better than one—Examining biology-specific TPACK and its T-dimensions from two angles. *Journal of research on Technology in Education*, 55(5), 765–782.
<https://doi.org/10.1080/15391523.2022.2030268>
- Kotzebue, L. von, Meier, M., Finger, A., Kremser, E., Huwer, J., Thoms, L.-J., Becker, S., Bruckermann, T. & Thyssen, C. (2021). The Framework DiKoLAN (Digital Competencies for Teaching in Science Education) as Basis for the Self-Assessment Tool DiKoLAN-Grid. *Education Sciences*, 11(12), 775. <https://doi.org/10.3390/educsci11120775>
- Krebs, R. (1997). The Swiss way to score multiple true-false items: theoretical and empirical evidence. In *Advances in medical education* (S. 158–161). Springer.
- Krebs, R. (2004). *Anleitung zur Herstellung von MC-Fragen und MC-Prüfungen für die ärztliche Ausbildung*. Institut für Medizinische Lehre, Abteilung für Assessment und Evaluation. Zugegriffen am 16.11.2024 unter https://www.ectaveo.ch/Mediathek/2015/08/Anleitung_MC-Fragen-Uni-Bern.pdf
- Krebs, R. (2019). *Prüfen mit Multiple Choice: Kompetent planen, entwickeln, durchführen und auswerten* (1. Auflage). Hogrefe.

- Kröger, J. (2019). *Struktur und Entwicklung des Professionswissens angehender Physiklehrkräfte* [Dissertation]. Christian-Albrechts-Universität zu Kiel.
- Kröger, J., Neumann, K. & Petersen, S. (2013). Messung Professioneller Kompetenz im Fach Physik. In S. Bernholt (Hrsg.), *Inquiry-based Learning – Forschendes Lernen. Gesellschaft für Didaktik der Chemie und Physik Jahrestagung in Hannover 2012* (S. 533–535). IPN.
- Kröger, J., Neumann, K. & Petersen, S. (2015). Struktur und Entwicklung des Professionswissens angehender Physiklehrkräfte. In S. Bernholt (Hrsg.), *Heterogenität und Diversität – Vielfalt der Voraussetzungen im naturwissenschaftlichen Unterricht. Gesellschaft für Didaktik der Chemie und Physik Jahrestagung in Bremen 2014* (S. 106–109). IPN.
- Kuckartz, U. (2014). *Mixed Methods: Methodologie, Forschungsdesigns und Analyseverfahren*. Springer VS. <https://doi.org/10.1007/978-3-531-93267-5>
- Kuckartz, U. (2018). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung* (4. Aufl.). Beltz Juventa.
- Kuhn, J., Müller, A., Hirth, M., Hochberg, K., Klein, P. & Molz, A. (2015). Experimentieren mit Smartphone und Tablet-PC: Einsatzmöglichkeiten für den Physikunterricht im Überblick. *Naturwissenschaften im Unterricht: Physik*, 26(145), 4–9.
- Kulgemeyer, C. (2018a). A Framework of Effective Science Explanation Videos Informed by Criteria for Instructional Explanations. *Research in Science Education*, 50, 2441–2462. <https://doi.org/10.1007/s11165-018-9787-7>
- Kulgemeyer, C. (2018b). Wie gut erklären Erklärvideos? *Computer + Unterricht*, 109, 8–11.
- Kulgemeyer, C., Borowski, A., Buschhüter, D., Enkrott, P., Kempin, M., Reinhold, P., Riese, J., Schecker, H., Schröder, J. & Vogelsang, C. (2020). Professional knowledge affects action-related skills: The development of preservice physics teachers' explaining skills during a field experience. *Journal of Research in Science Teaching*, 57(10), 1554–1582. <https://doi.org/10.1002/tea.21632>
- Kulgemeyer, C., Borowski, A., Fischer, H. E., Gramzow, Y., Reinhold, P., Riese, J., Schecker, H., Tomczyszyn, E. & Walzer, M. (2012). ProfiLe-P-

Professionswissen in der Lehramtsausbildung Physik. Vorstellung eines Forschungsverbundes. *PhyDid B - Didaktik der Physik - Beiträge zur DPG-Frühjahrstagung 2012 in Mainz*.

- Kulgemeyer, C. & Riese, J. (2018). From professional knowledge to professional performance: The impact of CK and PCK on teaching quality in explaining situations. *Journal of Research in Science Teaching*, 55(10), 1393–1418. <https://doi.org/10.1002/tea.21457>
- Kultusministerkonferenz (2016). *Bildung in der digitalen Welt: Strategie der Kultusministerkonferenz*. Kultusministerkonferenz (KMK). Zugegriffen am 16.11.2024 unter https://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2018/Digitalstrategie_2017_mit_Weiterbildung.pdf
- Kultusministerkonferenz (2019). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung* [(Beschluss der Kultusministerkonferenz von 16.10.2008 i. d. F. vom 16.05.2019)]. Kultusministerkonferenz (KMK). Zugegriffen am 24.09.2020 unter https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf
- Kultusministerkonferenz (2022). *Standards für die Lehrerbildung: Bildungswissenschaften: (Beschluss der Kultusministerkonferenz vom 16.12.2004 i. d. F. vom 07.10.2022)*. Kultusministerkonferenz (KMK). Zugegriffen am 19.02.2024 unter https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV*. Waxmann.
- Kunter, M. & Klusmann, U. (2010). Kompetenzmessung bei Lehrkräften – Methodische Herausforderungen. *Unterrichtswissenschaft*, 38(1), 68–86.
- Kunter, M., Kunina-Habenicht, O., Baumert, J., Dicke, T., Holzberger, D., Lohse-Bossenz, H., Leutner, D., Schulze-Stocker, F. & Terhart, E. (2017). Bildungswissenschaftliches Wissen und professionelle Kompetenz in der Lehramtsausbildung: Ergebnisse des Projekts BilWiss. In C. Gräsel & K. Trempler (Hrsg.), *Entwicklung von Professionalität pädagogischen Personals* (S. 37–54). Springer VS. https://doi.org/10.1007/978-3-658-07274-2_3

- Lachner, A., Backfisch, I. & Stürmer, K. (2019). A test-based approach of Modeling and Measuring Technological Pedagogical Knowledge. *Computers & Education*, 142, 103645.
<https://doi.org/10.1016/j.compedu.2019.103645>
- Lachner, A., Fabian, A., Franke, U., Preiß, J., Jacob, L., Führer, C., Kuchler, U., Paravicini, W., Randler, C. & Thomas, P. (2021). Fostering pre-service teachers' technological pedagogical content knowledge (TPACK): A quasi-experimental field study. *Computers & Education*, 174, 104304.
<https://doi.org/10.1016/j.compedu.2021.104304>
- Lampe, H.-U., Liebner, F., Urban-Woldron, H. & Tewes, M. (2015). *Innovativer naturwissenschaftlicher Unterricht mit digitalen Werkzeugen: Experimente mit Messwerterfassung in den Fächern Biologie, Chemie, Physik* (1. Aufl.). MNU Themenreihe Bildungsstandards. Seeberger.
- Laumann, D., Wichtrup, P. & Friege, G. (2019). Zwei Schlüssel zur Physik - Reale Experimente und digitale Medien als Schlüssel zu physikalischen Inhalten. *Naturwissenschaften im Unterricht: Physik*, 30(171/172), 4–9.
- Linacre, J. M. (1998). Thurstone thresholds and the Rasch model. *Rasch Measurement Transactions*, 12(2), 634–635. Zugegriffen am 15.07.2024 unter <https://www.rasch.org/rmt/rmt122j.htm>
- Magnusson, S., Krajcik, J. & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Hrsg.), *Examining pedagogical content knowledge* (S. 95–132). Kluwer Academic Publishers.
- Marks, R. (1990). Pedagogical content knowledge: From a mathematical case to a modified conception. *Journal of teacher education*, 41(3), 3–11.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174.
- Max, A.-L. (2022). *Die Entwicklung von TPACK bei angehenden Lehrkräften durch medienbezogene Lernprozesse im Makerspace* [Dissertation]. Pädagogische Hochschule Weingarten.
- Max, A.-L., Lukas, S. & Weitzel, H. (2020). Entwicklung von Aufgaben zur Bewertung des TPACK von Lehramtsstudierenden der Biologie. *Erkenntnisweg Biologiedidaktik*, 19, 57–71.

- Max, A.-L., Lukas, S. & Weitzel, H. (2022). The relationship between self-assessment and performance in learning TPACK: Are self-assessments a good way to support preservice teachers' learning? *Journal of Computer Assisted Learning*, 38(4), 1160–1172.
<https://doi.org/10.1111/jcal.12674>
- Max, A.-L., Weitzel, H. & Lukas, S. (2023). Factors influencing the development of pre-service science teachers' technological pedagogical content knowledge in a pedagogical makerspace. *Frontiers in Education*, 8, 1166018. <https://doi.org/10.3389/educ.2023.1166018>
- Mayer, P. & Girwidz, R. (2019). Physics Teachers' Acceptance of Multimedia Applications—Adaptation of the Technology Acceptance Model to Investigate the Influence of TPACK on Physics Teachers' Acceptance Behavior of Multimedia Applications. *Frontiers in Education*, 4, 73.
<https://doi.org/10.3389/educ.2019.00073>
- Mayer, R. E. (2009). *Multimedia Learning* (Second Edition). Cambridge university press.
- Mayer, R. E. & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, 38(1), 43–52.
https://doi.org/10.1207/S15326985EP3801_6
- Mayring, P. (2000). Qualitative Inhaltsanalyse. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(2).
<https://doi.org/10.17169/fqs-1.2.1089>
- Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12. überarb. Aufl.). Beltz.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749.
- Mikelskis, H. F. (Hrsg.). (2006). *Fachdidaktik. Physik-Didaktik: Praxishandbuch für die Sekundarstufe I und II* (1. Aufl.). Cornelsen Scriptor.
- Mishra, P. & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers college record*, 108(6), 1017–1054.

- Moosbrugger, H. & Brandt, H. (2020). Antwortformate und Itemtypen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3., vollständig neu bearbeitete, erweiterte und aktualisierte Auflage, S. 91–117). Springer. https://doi.org/10.1007/978-3-662-61532-4_5
- Moosbrugger, H., Gäde, J. C., Schermelleh-Engel, K. & Rauch, W. (2020). Klassische Testtheorie (KTT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3., vollständig neu bearbeitete, erweiterte und aktualisierte Auflage, S. 275–304). Springer. https://doi.org/10.1007/978-3-662-61532-4_13
- Moosbrugger, H. & Kelava, A. (2020). Qualitätsanforderungen an Tests und Fragebogen („Gütekriterien“). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3., vollständig neu bearbeitete, erweiterte und aktualisierte Auflage, S. 13–38). Springer. https://doi.org/10.1007/978-3-662-61532-4_2
- Olszewski, J. (2010). *The Impact of Physics Teachers' Pedagogical Content Knowledge on Teacher Actions and Student Outcomes. Studien zum Physik- und Chemielernen: Bd. 109*. Logos Verlag.
- Park, S. & Oliver, J. S. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education*, 38(3), 261–284.
- Puentedura, R. R. (2006). *Transformation, technology, and education* [Blog post]. Zugegriffen am 28.03.2024 unter <http://hippasus.com/resources/tte/>
- Puentedura, R. R. (2012). *Focus: Redefinition* [Blog post]. Zugegriffen am 28.03.2024 unter <http://hippasus.com/blog/archives/68>
- Rädiker, S. & Kuckartz, U. (2019). *Analyse qualitativer Daten mit MAXQDA*. Springer.
- Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu*. Publications Office of the European Union. Zugegriffen am 16.11.2024 unter <https://publications.jrc.ec.europa.eu/repository/handle/JRC107466>
- Rehm, M. & Bölsterli, K. (2014). Entwicklung von Unterrichtsvignetten. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 213–225). Springer.

- Reinhold, P. & Riese, J. (2020). Physik in der Lehrerinnen- und Lehrerbildung: Auf dem Weg zu empirisch fundierten Curricula. In C. Cramer, J. König, M. Rothland & S. Blömeke (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 501–508). Verlag Julius Klinkhardt.
- Riese, J. (2009). *Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkräften. Studien zum Physik- und Chemielernen: Bd. 97*. Logos Verlag.
- Riese, J., Gramzow, Y. & Reinhold, P. (2017). Die Messung fachdidaktischen Wissens bei Anfängern und Fortgeschrittenen im Lehramtsstudiengang Physik. *Zeitschrift für Didaktik der Naturwissenschaften*, 23(1), 99–112.
- Riese, J., Kulgemeyer, C., Zander, S., Borowski, A., Fischer, H. E., Gramzow, Y., Reinhold, P., Schecker, H. & Tomczyszyn, E. (2015). Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik. *Zeitschrift für Pädagogik*(61), 55–79.
- Riese, J. & Reinhold, P. (2008). Entwicklung und Validierung eines Instruments zur Messung professioneller Handlungskompetenz bei (angehenden) Physiklehrkräften. *Lehrerbildung auf dem Prüfstand*, 1(2), 625–640. <https://doi.org/10.25656/01:15868>
- Riese, J. & Reinhold, P. (2012). Die professionelle Kompetenz angehender Physiklehrkräfte in verschiedenen Ausbildungsformen. *Zeitschrift für Erziehungswissenschaft*, 15(1), 111–143.
- Riese, J. & Reinhold, P. (2014). Entwicklung eines Leistungstests für fachdidaktisches Wissen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 257–267). Springer.
- Riese, J., Schröder, J. & Vogelsang, C. (2022a). Die Entwicklung physikdidaktischen Wissens im Längsschnitt. In S. Habig & H. van Vorst (Hrsg.), *Gesellschaft für Didaktik der Chemie und Physik Jahrestagung virtuell 2021*.
- Riese, J., Vogelsang, C., Schröder, J., Borowski, A., Kulgemeyer, C., Reinhold, P. & Schecker, H. (2022b). Entwicklung von Unterrichtsplanungsfähigkeit im Fach Physik: Welchen Einfluss hat Professionswissen? *Zeitschrift für Erziehungswissenschaft*, 25(4), 843–867. <https://doi.org/10.1007/s11618-022-01112-0>

- Rindermann, H. (2006). Was messen internationale Schulleistungsstudien? *Psychologische Rundschau*, 57(2), 69–86.
- Robitzsch, A., Kiefer, T. & Wu, M. (2024). *TAM: Test Analysis Modules: R package version 4.2-21*. Zugegriffen am 15.07.2024 unter <https://doi.org/10.32614/CRAN.package.TAM>
- Ropohl, M., Härtig, H., Kampschulte, L., Lindmeier, A., Ostermann, A. & Schwanewedel, J. (2018). Planungsbereiche für Medieneinsatz im Fachunterricht. *MNU*, 71(3), 148–155.
- Rose, N. (2020). Parameterschätzung und Messgenauigkeit in der Item-Response-Theorie (IRT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3., vollständig neu bearbeitete, erweiterte und aktualisierte Auflage, S. 447–500). Springer. https://doi.org/10.1007/978-3-662-61532-4_19
- Rutten, N., van Joolingen, W. R. & van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education*, 58(1), 136–153.
- Sandmann, A. (2014). Lautes Denken–die Analyse von Denk-, Lern- und Problemlöseprozessen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 179–188). Springer.
- Schaber, M. & Friege, G. (2023). Digitalisierungsbezogene Kompetenzen angehender Physiklehrkräfte - Oder: Wie setzen Lehramtsstudierende (digitale) Medien im Physikunterricht ein? In H. van Vorst (Hrsg.), *Lernen, Lehren und Forschen in einer digital geprägten Welt: Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Aachen 2022*.
- Schaper, N. (2009). Aufgabenfelder und Perspektiven bei der Kompetenzmodellierung und -messung in der Lehrerbildung. *Lehrerbildung auf dem Prüfstand*, 2(1), 166–199. <https://doi.org/10.25656/01:14697>
- Schaper, N. (2015). Validitätsaspekte von Kompetenzmodellen und -tests für hochschulische Kompetenzdomänen. In F. Musekamp & G. Spöttl (Hrsg.), *Kompetenz im Studium und in der Arbeitswelt- Competence in Higher Education and the Working Environment* (S. 21–48). Peter Lang.
- Schecker, H. (2014). Überprüfung der Konsistenz von Itemgruppen mit Cronbachs α . In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung*. Springer.

- Schecker, H., Parchmann, I. & Krüger, D. (2014). Formate und Methoden naturwissenschaftsdidaktischer Forschung. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 1–15). Springer.
- Scherer, R., Tondeur, J. & Siddiq, F. (2017). On the quest for validity: Testing the factor structure and measurement invariance of the technology-dimensions in the Technological, Pedagogical, and Content Knowledge (TPACK) model. *Computers & Education*, *112*, 1–17.
<https://doi.org/10.1016/j.compedu.2017.04.012>
- Schiering, D., Sorge, S. & Neumann, K. (2021). Hilft viel viel? Der Einfluss von Studienstrukturen auf das Professionswissen angehender Physiklehrkräfte. *Zeitschrift für Erziehungswissenschaft*, *24*(3), 545–570.
<https://doi.org/10.1007/s11618-021-01003-w>
- Schiering, D., Sorge, S., Petersen, S. & Neumann, K. (2019). Konstruktion eines qualitativen Niveaumodells im fachdidaktischen Wissen von angehenden Physiklehrkräften. *Zeitschrift für Didaktik der Naturwissenschaften*, *25*(1), 211–229. <https://doi.org/10.1007/s40573-019-00100-y>
- Schmid, M., Brianza, E. & Petko, D. (2021). Self-reported technological pedagogical content knowledge (TPACK) of pre-service teachers in relation to digital technology use in lesson plans. *Computers in Human Behavior*, *115*, 106586. <https://doi.org/10.1016/j.chb.2020.106586>
- Schmid, M., Krannich, M. & Petko, D. (2020). Technological Pedagogical Content Knowledge. Entwicklungen und Implikationen. *Journal für LehrerInnenbildung*, *20*(1), 116–124. https://doi.org/10.35468/jlb-01-2020_10
- Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J. & Shin, T. S. (2009). Technological pedagogical content knowledge (TPACK) the development and validation of an assessment instrument for preservice teachers. *Journal of research on Technology in Education*, *42*(2), 123–149.
- Schmiemann, P. & Lücken, M. (2014). Validität – Misst mein Test, was er soll? In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 107–118). Springer.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353. <https://doi.org/10.1037/1040-3590.8.4.350>

- Schödl, A. (2017). *FALKO-Physik - Fachspezifische Lehrerkompetenzen im Fach Physik: Entwicklung und Validierung eines Testinstruments zur Erfassung des fachspezifischen Professionswissens von Physiklehrkräften. Studien zum Physik- und Chemielernen: Band 236*. Logos Verlag.
- Schödl, A. & Göhring, A. (2017). FALKO-P: Fachspezifische Lehrerkompetenzen im Fach Physik: Entwicklung und Validierung eines Testinstruments zu Erfassung des fachspezifischen Professionswissens von Physiklehrkräften. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff, R. H. Mulder & J. Baumert (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen: Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik: mit neuen Daten aus der COACTIV-Studie* (201-244). Waxmann.
- Schreier, M. (2014). Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der Begrifflichkeiten. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 15(1).
<https://doi.org/10.17169/fqs-15.1.2043>
- Schröder, J., Vogelsang, C. & Riese, J. (2020). Messung der Fähigkeit zur Unterrichtsplanung bei Lehramtsstudierenden. In S. Habig (Hrsg.), *Naturwissenschaftliche Kompetenz in der Gesellschaft von morgen. Gesellschaft für Didaktik der Chemie und Physik Jahrestagung in Wien 2019*.
- Schubatzky, T., Burde, J.-P., Große-Heilmann, R., Haagen-Schützenhöfer, C., Riese, J. & Weiler, D. (2023). Predicting the development of digital media PCK/TPACK: The role of PCK, motivation to use digital media, interest in and previous experience with digital media. *Computers & Education*, 206, 104900. <https://doi.org/10.1016/j.compedu.2023.104900>
- Schubatzky, T., Burde, J.-P., Große-Heilmann, R., Riese, J. & Weiler, D. (2022). Das Gesamtuntersuchungsdesign im Verbundprojekt DiKoLeP. In S. Habig & H. van Vorst (Hrsg.), *Gesellschaft für Didaktik der Chemie und Physik Jahrestagung virtuell 2021*.
- Schwanewedel, J., Ostermann, A. & Weigand, H.-G. (2018). Medien sind gut! Gut für was? Funktionen von Medien im Fachunterricht. In M. Ropohl, A. Lindmeier, H. Härtig, L. Kampschulte, A. Mühling & J. Schwanewedel (Hrsg.), *Medieneinsatz im mathematisch-naturwissenschaftlichen*

- Unterricht: Fachübergreifende Perspektiven auf zentrale Fragestellungen* (14-37). Joachim Herz Stiftung.
- Seifert, A., Hilligus, A. H. & Schaper, N. (2009). Entwicklung und psychometrische Überprüfung eines Messinstruments zur Erfassung pädagogischer Kompetenzen in der universitären Lehrerbildung. *Lehrerbildung auf dem Prüfstand*, 2(1), 82–103.
- Seifert, A. & Schaper, N. (2012). Die Entwicklung von bildungswissenschaftlichem Wissen: Theoretischer Rahmen, Testinstrument, Skalierung und Ergebnisse. In J. König & A. Seifert (Hrsg.), *Lehramtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerausbildung* (S. 183–214). Waxmann.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational researcher*, 15(2), 4–14.
- Sorge, S., Keller, M., Petersen, S. & Neumann, K. (2018). Die Entwicklung des Professionswissens angehender Physiklehrkräfte. In C. Maurer (Hrsg.), *Qualitätvoller Chemie- und Physikunterricht-normative und empirische Dimensionen: Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Regensburg 2017* (S. 114–117). IPN.
- Sorge, S., Kröger, J., Petersen, S. & Neumann, K. (2019). Structure and development of pre-service physics teachers' professional knowledge. *International Journal of Science Education*, 41(7), 862–889.
<https://doi.org/10.1080/09500693.2017.1346326>
- Stamann, C., Janssen, M. & Schreier, M. (2016). Qualitative Inhaltsanalyse – Versuch einer Begriffsbestimmung und Systematisierung. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 17(3).
<https://doi.org/10.17169/fqs-17.3.2581>
- Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (2022). *Digitalisierung im Bildungssystem: Handlungsempfehlungen von der Kita bis zur Hochschule.: Gutachten der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz (SWK)*. Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (SWK). Zugegriffen am 16.11.2024 unter <http://dx.doi.org/10.25656/01:25273>
- Steffensky, M. & Neuhaus, B. J. (2018). Unterrichtsqualität im naturwissenschaftlichen Unterricht. In D. Krüger, I. Parchmann & H. Schecker

- (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 299–313). Springer. https://doi.org/10.1007/978-3-662-56320-5_18
- Steinmetz, T. (2021). *Kumulatives Lehren und Lernen Im Lehramtsstudium Physik: Theorie und Evaluation Eines Lehrkonzepts. Studien zum Physik- und Chemielernen: Bd. 324*. Logos Verlag.
- Sterzing, F., Varnai, A. S. & Reinhold, P. (2019). Erklärvideos im Physikunterricht. *PhyDid B - Didaktik der Physik - Beiträge zur DPG-Frühjahrstagung 2019 in Aachen*, 315-322.
- Stinken-Rösner, L. (2021a). Digitale Medien in der naturwissenschaftlichen Lehrkräftebildung: Integriert statt zusätzlich. *PhyDid B - Didaktik der Physik - Beiträge zur DPG-Frühjahrstagung 2021 virtuell*, 179–185.
- Stinken-Rösner, L. (2021b). Implementation digitaler Medien in die naturwissenschaftliche Lehramtsausbildung. In C. Maurer, K. Rincke & M. Hemmer (Hrsg.), *Fachliche Bildung und digitale Transformation - Fachdidaktische Forschung und Diskurse: Fachtagung der Gesellschaft für Fachdidaktik 2020*, Regensburg.
- Stinken-Rösner, L., Hofer, E., Rodenhauser, A. & Abels, S. (2023). Technology Implementation in Pre-Service Science Teacher Education Based on the Transformative View of TPACK: Effects on Pre-Service Teachers' TPACK, Behavioral Orientations and Actions in Practice. *Education Sciences*, 13, 732. <https://doi.org/10.3390/educsci13070732>
- Strobl, C. (2015). *Das Rasch-Modell: Eine verständliche Einführung für Studium und Praxis* (3. erweiterte Auflage). *Sozialwissenschaftliche Forschungsmethoden: Bd. 2*. Reiner Hampp Verlag.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, 4(4), 295–312.
- Teichrew, A. & Erb, R. (2020). How augmented reality enhances typical classroom experiments: examples from mechanics, electricity and optics. *Physics Education*, 55(6), 65029. <https://doi.org/10.1088/1361-6552/abb5b9>
- Tepner, O., Borowski, A., Dollny, S., Fischer, H. E., Jüttner, M., Kirschner, S., Leutner, D., Neuhaus, B. J., Sandmann, A. & Sumfleth, E. (2012). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 7–28.

- Tepner, O. & Dollny, S. (2014). Entwicklung eines Testverfahrens zur Analyse fachdidaktischen Wissens. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 311–323). Springer.
- Terhart, E. (2002). *Standards für die Lehrerbildung. Eine Expertise für die Kultusministerkonferenz*. Institut für Schulpädagogik und Allgemeine Didaktik, Westphälische Wilhelms-Universität Münster.
- Tondeur, J., van Braak, J., Sang, G., Voogt, J., Fisser, P. & Ottenbreit-Leftwich, A. (2012). Preparing pre-service teachers to integrate technology in education: A synthesis of qualitative evidence. *Computers & Education*, 59(1), 134–144. <https://doi.org/10.1016/j.compedu.2011.10.009>
- Veltum, V. (2023). *Die Lernwirksamkeit bestimmter Seminarelemente zum Erwerb von digitalen Kompetenzen im Fach Physik aus Sicht der Studierenden* [Unveröffentlichte Bachelorarbeit]. Universität Paderborn.
- VERBI GmbH (2022). *MAXQDA 2022 Manual*. Zugegriffen am 16.11.2024 unter <https://www.maxqda.com/de/hilfe-mx22/willkommen>
- Vogelsang, C., Borowski, A., Buschhüter, D., Enkrott, P., Kempin, M., Kulgemeyer, C., Reinhold, P., Riese, J., Schecker, H. & Schröder, J. (2019). Entwicklung von Professionswissen und Unterrichtsperformanz im Lehramtsstudium Physik–Analysen zu valider Testwertinterpretation. *Zeitschrift für Pädagogik*, 65 (4), 473–491.
- Voogt, J., Fisser, P., Pareja Roblin, N., Tondeur, J. & van Braak, J. (2013). Technological pedagogical content knowledge – a review of the literature. *Journal of Computer Assisted Learning*, 29(2), 109–121. <https://doi.org/10.1111/j.1365-2729.2012.00487.x>
- Walpert, D. & Wodzinski, R. (2023). Die Vermittlung digitaler Kompetenzen im Lehr-Lern-Labor-Setting. In H. van Vorst (Hrsg.), *Lernen, Lehren und Forschen in einer digital geprägten Welt: Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Aachen 2022*.
- Wang, W., Schmidt-Crawford, D. & Jin, Y. (2018). Preservice Teachers' TPACK Development: A Review of Literature. *Journal of Digital Learning in Teacher Education*, 34(4), 234–258.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3), 427–450. <https://doi.org/10.1007/BF02294627>

- Weidenmann, B. (2002). Multicodierung und Multimodalität im Lernprozess. In L. J. Issing & P. Klimsa (Hrsg.), *Information und Lernen mit Multimedia und Internet: Lehrbuch für Studium und Praxis* (3., vollst. überarb. Aufl., S. 45–62). Beltz Psychologie-Verlags-Union.
- Weiler, D., Burde, J.-P., Große-Heilmann, R., Lachner, A., Riese, J. & Schubatzky, T. (2022a). Bedarfsanalyse zu digitalen Medien bei Physik-Lehramtsstudierenden. In S. Habig & H. van Vorst (Hrsg.), *Gesellschaft für Didaktik der Chemie und Physik Jahrestagung virtuell 2021*.
- Weiler, D., Burde, J.-P., Große-Heilmann, R., Lachner, A., Riese, J. & Schubatzky, T. (2023). Förderung von digitalisierungsbezogenen Kompetenzen von angehenden Physiklehrkräften mit dem SQD-Modell im Projekt DiKoLeP. In M. Meier, G. Greefrath, M. Hammann, R. Wodzinski & K. Ziepprecht (Hrsg.), *Edition Fachdidaktiken. Lehr-Lern-Labore und Digitalisierung* (S. 47–62). Springer VS. https://doi.org/10.1007/978-3-658-40109-2_4
- Weiler, D., Burde, J.-P., Große-Heilmann, R., Lachner, A., Riese, J. & Schubatzky, T. (2024). Einsatz digitaler Medien: Charakterisierung von Physik-LA-Studierenden. In H. van Vorst (Hrsg.), *Frühe naturwissenschaftliche Bildung: Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Hamburg 2023*.
- Weiler, D., Burde, J.-P., Lachner, A., Große-Heilmann, R., Riese, J. & Schubatzky, T. (2022b). Digitale Medien im Physikunterricht: Entwicklung eines Seminar-konzepts. *PhyDid B - Didaktik der Physik - Beiträge zur DPG-Frühjahrstagung 2022 virtuell*.
- Weinert, F. E. (2001a). Concept of competence: A conceptual clarification. In *Defining and selecting key competencies* (S. 45–65). Hogrefe & Huber.
- Weinert, F. E. (2001b). Vergleichende Leistungsmessung in Schulen - eine umstrittene Selbstverständlichkeit. In *Leistungsmessungen in Schulen* (S. 17–32). Beltz.
- Wieman, C. E., Adams, W. K., Loeblein, P. & Perkins, K. K. (2010). Teaching physics using PhET simulations. *The Physics Teacher*, 48(4), 225–227.
- Wiesner, H., Schecker, H. & Hopf, M. (2011). *Physikdidaktik kompakt*. Aulis-Verlag.

- Wilhelm, T. (Hrsg.). (2020). *Plus Lucis: 01/2020. Videoanalyse von Bewegungen*.
- Willermark, S. (2018). Technological Pedagogical and Content Knowledge: A Review of Empirical Studies Published From 2011 to 2016. *Journal of Educational Computing Research*, 56(3), 315–343.
<https://doi.org/10.1177/0735633117713114>
- Wolf, K. D. (2018). Video statt Lehrkraft? Erklärvideos als didaktisches Element im Unterricht. *Computer + Unterricht*(109), 4–7.
- Wolf, K. D. & Kulgemeyer, C. (2016). Lernen mit Videos? Erklärvideos im Physikunterricht. *Naturwissenschaften im Unterricht: Physik*, 27(152), 36–41.
- Wright, B. & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370. Zugegriffen am 16.07.2024 unter <https://www.rasch.org/rmt/rmt83b.htm>
- Zimmermann, F. (2021). *Entwicklung und Evaluation digitalisierungsbezogener Kompetenzen von angehenden Chemielehrkräften. Studien zum Physik- und Chemielernen: Bd. 322*. Logos Verlag.
- Zimmermann, F. & Melle, I. (2023). Förderung digitalisierungsbezogener Kompetenzen angehender Chemielehrkräfte: Vergleich der Wirkung eines Hochschulseminars im Präsenz- und Online-Format. *CHEMKON*, 30(2), 68–74. <https://doi.org/10.1002/ckon.202100066>
- Zimmermann, F., Melle, I. & Huwer, J. (2021). Developing prospective chemistry teachers' TPACK—A comparison between students of two different universities and expertise levels regarding their TPACK self-efficacy, attitude, and lesson planning competence. *Journal of Chemical Education*, 98(6), 1863–1874.

Anhang

| | | |
|----------|---|------------|
| A | Modellierung des FDW zum Einsatz digitaler Medien | 293 |
| B | Testinstrument zum FDW zum Einsatz digitaler Medien | 299 |
| C | Ergänzende Materialien zum Testinstrument | 315 |
| D | Studien zur Untersuchung der Validität | 329 |
| E | Interviewleitfaden und Interviewmaterial..... | 331 |
| F | Interviewauswertung und Kategoriensystem | 341 |
| G | Kategorisierung zur Identifikation lernförderlicher und lernhinderlicher Seminarelemente | 357 |
| H | Hypothesenbildung zur Gestaltung von Lerngelegenheiten zum Einsatz digitaler Medien | 363 |

A Modellierung des FDW zum Einsatz digitaler Medien

Kodierung: in der Modellierung enthalten, aber Bezug zum Fach Physik notwendig

Allgemein/Grundlagen

Möglichkeiten und Anforderungen bei der Mediennutzung im Physikunterricht

- **Medien als Hilfsmittel/Vermittler**, kognitive und kommunikative Werkzeuge zum Erreichen des Lernziels (Ropohl et al. 2018; Girwidz 2015a, 2020b)
- Ziel des Medieneinsatzes sollte ein effektiver und zielgerichteter Einsatz sein, der sonst **mit anderen Unterrichtsmitteln schwieriger zu erreichen** wäre (Girwidz 2020c)
- **Ersetzen keine Experimente** (Girwidz 2020b) **und keine Lehrkraft** (Hillmayr et al. 2017; Wieman et al. 2010)
- **Funktionen:** Informationsvermittlung, Motivierung, Veranschaulichung, Visualisierung, Differenzierung, zum selbstgesteuerten Arbeiten (Wiesner et al. 2011)
- Als **Erweiterung des Erfahrungskegels**, wenn direkte Erfahrung nicht möglich (Wiesner et al. 2011)
- Umformung und/oder Erweiterung klassischer Unterrichtstechniken nach **SAMR-Modell** von Puentedura (2013): Ersetzung, Augmentation, Modifizierung oder Neudefinition (Girwidz 2020c)
- Gute Möglichkeit für Einbettung in Kontexte, verschiedene Perspektivenbetrachtung zur Förderung von Interesse, kognitiver Flexibilität, anwendbarem Wissen und Aufbau mentaler Modelle (Weidenmann 2002)
- Medien zum Entwickeln neuer Erkenntnistheorien und Beseitigen/Adressieren von Lernproblemen/-schwierigkeiten (Mishra und Koehler 2006)
- physikdidaktisches Potenzial neuer Medien, **typische Schülervorstellungen zu adressieren** und das **Konzeptverständnis zu fördern**

Differenzierung

- Digitale Medien **bieten Potenzial zur individuellen Förderung** (Netzwerk Digitale Bildung und Henkelmann 2018)
- Binnendifferenzierung auf Basis von Schülervorstellungen anhand von digitalen Medien (z.B. Audience Response Systeme)

Anforderungen:

- **Lernhindernis bei überflüssigen, nicht ziel- und aufgabenadäquaten Visualisierungen** (Girwidz 2015b, 2020c; Issing 2002)
- **Interne Codierung entspricht nicht immer externen Repräsentation** (Girwidz 2015b, 2020c; Weidenmann 2002)
- **Kognitive Belastung** (Sweller 1994), Belastungseffekte -> Gestaltungsprinzipien beachten (s.u.)
- **außerdem Vorwissen, Vorerfahrungen, Instruktionsqualität und lernunterstützende (metakognitive) Maßnahmen relevant für Lernförderlichkeit** (Schanze und Girwidz 2018)

Multimedia & lerntheoretische Grundlagen

- **Multimodalität:** mehrere sensorische Systeme/ Sinneskanäle, **Multicodierung:** unterschiedliche Symbolsysteme / Codierungen
- **Multimediaeffekte/ Designprinzipien** (Mayer 2009)
- **Grenzen des Arbeitsgedächtnisses, Cognitive Load** (Sweller 1994): extraneous, intrinsic und germane cognitive load; extraneous & intrinsic load gering halten und Kapazitäten für Lernaktivitäten lassen
- **Belastungseffekte & Maßnahmen** beachten bei Gestaltung
- z.B. **Chunking, grafische Hilfen, gesprochene Test, farbliche Codierung, single-concept, Multimodalität etc.** (Girwidz 2015b, 2020c)
- **aber: Belastung abhängig von Vorwissen u. kognitiven Fähigkeiten** (Girwidz 2015b, 2020c)
- **Mentale Modelle** (Girwidz 2004, 2015b, 2020c; Wieman et al. 2010; Weidenmann 2002)
- **Kognitive Flexibilität** (Spiro et al. 1995; Weidenmann 2002)
- **Multiple Repräsentationen**

Fortsetzung auf nächster Seite

Medienbezogene Kompetenz

Aus fachdidaktischer Sicht:

- Klärung **fachlicher „Bildungsanliegen“** und deren Adressierung (Ropohl et al. 2018)
- Relevanz **fachspezifischer/-typischer Medien, an fachlichen Lernzielen und Denk-/Arbeitsweisen des Fachs** orientierte Gestaltung (Ropohl et al. 2018)
- Digitale Kompetenzen von Lehrkräften **als Ergänzung der bisherigen fachdidaktischen Kompetenzen**, d.h. fachdidaktische Prinzipien bilden eine wichtige Grundlage für den **Einsatz digitaler Medien, der durch didaktische und methodische Analysen begründet sein muss.** (Arbeitsgruppe Digitale Basiskompetenzen: Becker, Sebastian et al. 2020, S. 16)
- **Vor- und Nachteile digitaler Medien für Unterrichtsgestaltung erkennen** (Arbeitsgruppe Digitale Basiskompetenzen: Becker, Sebastian et al. 2020, S. 16)
- **digitale Medien im Fachunterricht professionell und didaktisch sinnvoll nutzen & gemäß dem Bildungs- und Erziehungsauftrag inhaltlich reflektieren** (Kultusministerkonferenz 2019, S. 50)
- **Entwicklungen im Bereich Digitalisierung aus fachlicher und fachdidaktischer Sicht angemessen zu rezipieren** (Kultusministerkonferenz 2019, S. 50)
- **Erkenntnisse in fachdidaktischen Kontexten nutzen und in Weiterentwicklung unterrichtlicher und curricularer Konzepte einbringen** (Kultusministerkonferenz 2019, S. 50)
- **adäquaten Einsatz digitaler Medien planen, durchführen und reflektieren** (Kultusministerkonferenz 2017)
- **identifizieren geeigneter Materialien aus angebotenen Bildungsmedien anhand Qualitätskriterien** (Kultusministerkonferenz 2017)

Inhaltsspezifische Nutzung von Medien

- **Supplantation** nach Salomon (1979): externe mediale Visualisierung kognitiver Prozesse, die Lernende nicht selbständig realisieren können (Girwidz 2015b, 2020c)
- Relevanz für **Erweiterung des Erfahrungskegels**, wenn direkter Zugang zum physikalischen Sachverhalt nicht möglich ist (Wiesner et al. 2011)
- Potenzial neuer Medien, um **Schülervorstellungen zu adressieren und Konzeptverständnis zu fördern**
- Physik und Musik verknüpfen mit Computeranwendungen (Girwidz 2004)

Digitale Messwerterfassung, -aufbereitung & -auswertung

- Automatische **Datenerfassung**, in Echtzeit, **Auswertung** und **Präsentation** der Daten (Girwidz 2015c, 2020c)
- **Digitale Messwerterfassungssysteme (MES)**: große Datenmengen in kurzer Zeit
- digitale Präsentation von Prozessen (z.B. Zeitlupe für Bewegungen, Hochgeschwindigkeitsaufnahmen)
- **Videoanalyse**; Messwerte aus **visuellen Medien** (Stroboskopbilder, Videoaufnahme)
- Einsatz von **Sensoren (Smartphones und Messwertssysteme)**
- **Mikrocontroller** zum MES bauen

Experimentierphasen und digitale Hilfsmittel dazu: (Girwidz 2020a, S. 285)

- digitale Aufbauten: Computer-Interfaces, mobile Endgeräte, Simulationen zur Übung
- digitale Steuerung/Regelung, **systematische Variablenänderung** etc.
- **Digitale Datenaufnahme** (Zeiteffizienz)
- **Digitale Messwertdarstellung** (Diagramme, Zeiteffizienz)
- Digitale Auswertung (**Tabellenkalkulation, Diagramme**)

Fortsetzung auf nächster Seite

- Digitale Modelle realer Experimente: IBE, Simulationen und VR-Experimente (Laumann et al. 2019)
 - **Virtual / Virtual Reality Labs** (realitätsnahe Gestaltung), **Interaktive Bildschirmexperimente IBE** (Fotoserien), **Remote Labs** (ferngesteuertes Realexperiment, echte Live-Daten)
 - VR und remote Labs ermöglichen selbstständiges, gefahrenloses und kostengünstiges Durchführen von Experimenten (Girwidz 2020a, S. 286–287)
 - über **Augmented Reality** können beim Experimentieren zusätzlich Hilfen geboten werden (Girwidz 2020a, S. 286–287)
-
- Umgang mit großen Datensätzen, Kompetenzen **Data Literacy, Digital Literacy** und **Big Data** Messwerterfassung (Arbeitsgruppe Digitale Basiskompetenzen: Becker, Sebastian et al. 2020)
 - Alternativen zur fachwissenschaftlichen **digitalen Messwerterfassung (dMWE) für Schuleinsatz** nennen
 - **Sachgerechte Einsatzszenarien und verbundene Messstrategien** (z.B. **computergestützt, mobile Endgeräte mit Kameras, integrierten oder externen Sensoren, Wärmebildkamera**)
 - Didaktische **Voraussetzungen** für Unterrichtseinsatz von dMWE (z.B. individuell angepasste Instruktionen)
 - **Auswirkungen** auf Unterrichtsverfahren (z.B. Möglichkeit für **forschend-entdeckendes Lernen**)
 - **Einbindung dMWE im Unterricht**, unter Berücksichtigung geeigneter Organisations-/ Sozialformen
 - *Auswirkungen, pädagogische didaktische Voraussetzungen und Vor- und Nachteile durch Einsatz dMWE* Datenaufbereitung (Arbeitsgruppe Digitale Basiskompetenzen: Becker, Sebastian et al. 2020)
 - Werkzeuge für **sachgerechten Einsatz, Einsatzszenarien** in spez. Lehr-Lern-Situationen (Passung **inhaltlich sinnvoller Kontext**)
 - Didaktische Voraussetzungen, Auswirkungen auf Unterrichtsverfahren, Anwendung im Unterricht
 - Fachwissenschaftliche Szenarien mit zugehörigen **Methoden der Datenverarbeitung** z.B. Bestimmung von Kurvenmaxima (Schallpegel, Beschleunigung) sowie **Messunsicherheiten**, Standardfehler und Streuung
- Dokumentation (Arbeitsgruppe Digitale Basiskompetenzen: Becker, Sebastian et al. 2020)
- Digitale Techniken zur **Dokumentation, Datenarchivierung**, Back-up-Erstellung für **bestimmte Lehr-Lern-Situationen (Experimentieren, Literaturrecherche)** nennen & didaktisch begründete **Vorgehensweisen** dazu beschreiben
 - *Elementarisierung fachwissenschaftlicher Darstellungen im Schulkontext*
-
- ### **Nutzung von Smartphones & Tablets im PU**
- Smartphones/Tablets als Experimentiergeräte (Messsensoren, Apps, Videoaufnahme) (Girwidz 2015c, 2020c)
 - Analyse mit mobilen Endgeräten (**Kamera, integrierte und externe Sensoren**) (Arbeitsgruppe Digitale Basiskompetenzen: Becker, Sebastian et al. 2020)
 - z.B. **Videoanalyse**-Apps, **Akustikversuche** mit Apps, Medium für **Augmented Reality** (mehr Selbstständigkeit, Kollaboration in Gruppen) (Bewersdorff 2020, S. 67-70)
 - **Dokumentieren** (Bild, Video) des Versuchs, **Datenaustausch** Smartphone
 - Einfach, flexibel und ortsunabhängig (Girwidz 2015c, 2020c), transportable Messwerterfassung, Integration der Alltagswelt
 - Wissensspeicher, Messwerterfassung, kognitives Werkzeug, Simulationen in einem Gerät (Kuhn et al. 2013; Becker et al. 2019)
 - Lernwirksamkeit einer Tablet-PC-gestützte Videoanalyse in Themenbereichen der Mechanik gegenüber traditionellem Unterricht (Becker et al. 2018)

Fortsetzung auf nächster Seite

Simulationen / Modellbildung

- **Modellierung/Rekonstruktion** eines Phänomens/einer Aktivität; Interaktion durch **Steuerung von Parametern** (Wiesner et al. 2011; Blake und Scanlon 2007; Jong 2011)
- Vier Phasen: Analyse, **Hypothesengenerierung**, Hypothesen testen und Evaluation (Schulmeister 2007)
- Wenn reale Darstellung zu schwierig, kompliziert, teuer, gefährlich, unmöglich; als **Ergänzung**, wenn Experiment nicht möglich
- Hilfreich als „**add-on**“ im Unterricht, **nicht als Ersatz von Realversuchen** (Rutten et al. 2012; Girwitz 2015b, 2020c)
- **Vorteile:** Vorwissen aktivieren, Prüfen eigener Vermutungen, multiple Repräsentationen, interaktive Visualisierungen (Jong 2011), einfache **Variablenkontrolle**, systematisches **Hypothesenprüfen**, Zeitersparnis und Fokussierung auf Interaktion mit SuS, **Umgang mit Schülervorstellungen** (Blake und Scanlon 2007; Rutten et al. 2012)
- Modelle/Modellbildung wichtig für **Erkenntnisgewinnung** (Mikelskis-Seifert 2006, S. 127ff)
- Simulation als Bsp. für **wissenschaftliches Modell** thematisieren (Wieman et al. 2010)
- Beleuchten nur Teilaspekt der Realität

- Geeignet für entdeckendes Lernen, aber mit Orientierungshilfen u. Unterstützung wie Lösungsbeispiele etc. (Jong 2011; Alfieri et al. 2011; Rutten et al. 2012; Girwitz 2015b, 2020c)
- **Problemlösekompetenzen** fördern mit Simulationen (Schulmeister 2007; Goodyear 1992)
- Beim Simulationseinsatz zur Förderung von Problemlösefertigkeiten sollten den Lernenden **Hinweise** zu den Begriffen, den Zusammenhängen im Modell und Strategien zum Hypothesentesten geboten werden (Goodyear 1992)

Weitere Alternativen für das Experimentieren:

- **IBE, Virtual / Virtual Reality Labs** (s.o.)
- Vor- und Nachteile von **gefilmten Experiment, remote controlled Lab** gegenüber traditionellem Experiment

Simulation und Modellierung (Arbeitsgruppe Digitale Basiskompetenzen: Becker, Sebastian et al. 2020)

- Strategien zum Einsatz in Lehr-Lern-Szenarien: Möglichkeit des **Erkenntnisgewinns** (mangels finanzieller, zugänglicher oder sicherer Methoden oder als fachspezifische Arbeitsweise); als **zeitlich optimierte** Form der Datengewinnung, als **interaktive** Methode, als Ansatz für **gezielte, variable Modellkritik**
- Didaktische Voraussetzungen und Auswirkungen auf Unterrichtsverfahren; Anwendung im Unterricht, Berück. OF&SF
- Zugang zu Basiskompetenzen, v.a. **Erkenntnisgewinnung**, ggfs. Kommunikation
- *Vor- und Nachteile, typische Eigenschaften und Grenzen in Lehr-Lernszenarien (hinsichtlich fachlicher Korrektheit, Qualität Repräsentationen, mathematische Modelle (Parameter, Rundungsfehler, Eingabegenauigkeit), notwendiger Vorkenntnisse)*
- + *adäquate Elementarisierung*
- *Bewertung bezüglich Motivation (Usability, Attraktivität, Klarheit)*
- z.B. **Tabellenkalkulation, Geogebra**

Modellbildungssysteme

- Ermöglichen zusätzlich Änderung der Modellannahmen (Wiesner et al. 2011)
- Größere Verarbeitungstiefe, Gedächtnishaftung und selbstbestimmtes Lernen (Girwitz 2015c, 2020c)
- benötigt Vor- und Nachbereitung
- Herausforderung bei Auswahl der Software, Durchführung, Reduktion auf Wesentliches und Evaluation
- z.B. Excel, Newton3

Fortsetzung auf nächster Seite

Erklärvideos im PU

- *Hinweis: grundsätzlich nicht physikspezifisch, aber durchaus relevant für den Ph-Unterricht und die physikdidaktische Forschung aufgrund der komplexen Inhalte in der Physik, für die sich Erklärvideos gerade besonders eignen*
- Produzenten & Rezipienten: Lehrkraft für andere LK oder für SuS, SuS für LK oder für MitSuS (Wolf und Kulgemeyer 2016; Wolf 2018)
- 2 Startbedingungen: **neues Konzept**, zu komplex für Selbsterklärung und **geringes Vorwissen** der Lernenden (Kulgemeyer 2018b)
- 7 Faktoren für Effektivität: Struktur, Adaption, Werkzeuge (Beispiele, Darstellungsformen, Mathematisierung, Sprache), minimale Erklärung, Relevanz, Lernaufgaben, Fokus auf wissenschaftliches Prinzip (Kulgemeyer 2018a)
- **Vertiefende Aufgaben** anschließend (Kulgemeyer 2018c)
- **Minimalistische Erklärungen**, um kognitive Belastung gering zu halten, Regel-Beispiel-Struktur für Wissenserwerb hilfreich, vorgehende Orientierung und abschließende Zusammenfassung (nicht zu viel) sind förderlich (Kulgemeyer 2018c)
- **Fachwissen verbalisieren** und **kommunizieren** fördern

Literaturverzeichnis zur Modellierung

- Alfieri, Louis; Brooks, Patricia J.; Aldrich, Naomi J.; Tenenbaum, Harriet R. (2011): Does discovery-based instruction enhance learning? In: *Journal of educational psychology* 103 (1), S. 1–18. DOI: 10.1037/a0021017.
- Arbeitsgruppe Digitale Basiskompetenzen: Becker, Sebastian; Bruckermann, Till; Finger, Alexander; Huwer, Johannes; Kremser, Erik; Meier, Monique et al. (2020): Orientierungsrahmen Digitale Kompetenzen für das Lehramt in den Naturwissenschaften - DiKoLAN. In: Sebastian Becker, Jenny Meßinger-Koppelt und Christoph Thyssen (Hg.): *Digitale Basiskompetenzen. Orientierungshilfe und Praxisbeispiele für die universitäre Lehramtsausbildung in den Naturwissenschaften*. Hamburg: Joachim Herz Stiftung, S. 14–43.
- Becker, Sebastian; Klein, Pascal; Gößling, Alexander; Kuhn, Jochen (2019): Using mobile devices to augment inquiry-based learning processes with multiple representations. In: *arXiv preprint arXiv:1908.11281*.
- Blake, C.; Scanlon, E. (2007): Reconsidering simulations in science education at a distance: features of effective use. In: *Journal of Computer Assisted Learning* 23 (6), S. 491–502. DOI: 10.1111/j.1365-2729.2007.00239.x.
- Girwitz, Raimund (2004): Lerntheoretische Konzepte für Multimediaanwendungen zur Physik. In: *PhyDid A-Physik und Didaktik in Schule und Hochschule* 1 (3), S. 9–19.
- Girwitz, Raimund (2015a): Medien im Physikunterricht. In: Ernst Kircher, Raimund Girwitz und Peter Häußler (Hg.): *Physikdidaktik. Theorie und Praxis*. 3. Aufl. Heidelberg: Springer, S. 193–245.
- Girwitz, Raimund (2015b): Multimedia unter lerntheoretischen Aspekten. In: Ernst Kircher, Raimund Girwitz und Peter Häußler (Hg.): *Physikdidaktik. Theorie und Praxis*. 3. Aufl. Heidelberg: Springer, S. 843–877.
- Girwitz, Raimund (2015c): Neue Medien und Multimedia. In: Ernst Kircher, Raimund Girwitz und Peter Häußler (Hg.): *Physikdidaktik. Theorie und Praxis*. 3. Aufl. Heidelberg: Springer, S. 401–427.
- Girwitz, Raimund (2020a): Experimente im Physikunterricht. In: Ernst Kircher, Raimund Girwitz und Hans E. Fischer (Hg.): *Physikdidaktik. Grundlagen*. 4. Aufl. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 263–291.
- Girwitz, Raimund (2020b): Medien im Physikunterricht. In: Ernst Kircher, Raimund Girwitz und Hans E. Fischer (Hg.): *Physikdidaktik. Grundlagen*. 4. Aufl. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 293–335.
- Girwitz, Raimund (2020c): Multimedia und digitale Medien im Physikunterricht. In: Ernst Kircher, Raimund Girwitz und Hans E. Fischer (Hg.): *Physikdidaktik. Grundlagen*. 4. Aufl. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 457–527.
- Goodyear, Peter (1992): The provision of tutorial support for learning with computer-based simulations. In: *Computer-based learning environments and problem solving*: Springer, S. 391–409.

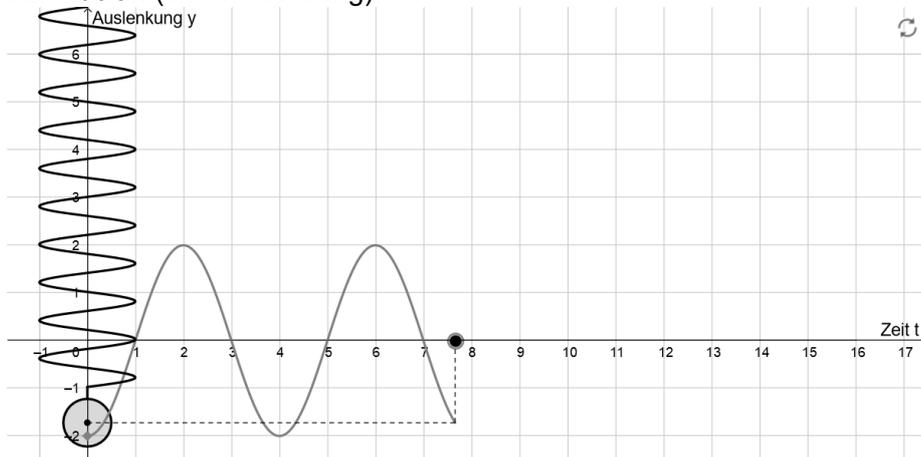
Fortsetzung auf nächster Seite

- Gramzow, Yvonne (2015): Fachdidaktisches Wissen von Lehramtsstudierenden im Fach Physik. Modellierung und Testkonstruktion. Berlin: Logos Verlag.
- Hillmayr, Delia; Reinhold, Frank; Ziernwald, Lisa; Reiss, Kristina (2017): Digitale Medien im mathematisch-naturwissenschaftlichen Unterricht der Sekundarstufe. Einsatzmöglichkeiten, Umsetzung und Wirksamkeit. Münster, New York: Waxmann. Online verfügbar unter <https://www.waxmann.com/?elD=texte&pdf=3766Volltext.pdf&typ=zusatztext>.
- Jong, Ton de (2011): Instruction based on computer simulations. In: *Handbook of research on learning and instruction*, S. 446–466.
- Kuhn, Jochen; Wilhelm, Thomas; Lück, Stephan (2013): Physik mit Smartphones und Tablet-PCs. In: *Physik in unserer Zeit* 44 (1), S. 44–45.
- Kulgemeyer, Christoph (2018a): A Framework of Effective Science Explanation Videos Informed by Criteria for Instructional Explanations. In: *Res Sci Educ*. DOI: 10.1007/s11165-018-9787-7.
- Kulgemeyer, Christoph (2018b): Towards a framework for effective instructional explanations in science teaching. In: *Studies in Science Education* 54 (2), S. 109–139. DOI: 10.1080/03057267.2018.1598054.
- Kulgemeyer, Christoph (2018c): Wie gut erklären Erklärvideos? In: *Computer+ Unterricht* 109, S. 8–11.
- Kultusministerkonferenz (2017): Bildung in der digitalen Welt. Strategie der Kultusministerkonferenz. Online verfügbar unter https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2018/Strategie_Bildung_in_der_digitalen_Welt_idF_vom_07.12.2017.pdf.
- Kultusministerkonferenz (2019): Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung. (Beschluss der Kultusministerkonferenz von 16.10.2008 i. d. F. vom 16.05.2019). Online verfügbar unter https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf, zuletzt geprüft am 24.09.2020.
- Laumann, Daniel; Wichtrup, Philipp; Friege, Gunnar (2019): Zwei Schlüssel zur Physik - Reale Experimente und digitale Medien als Schlüssel zu physikalischen Inhalten. In: *Naturwissenschaften im Unterricht: Physik* 30 (171/172), S. 4–9.
- Mayer, Richard E. (2009): Multimedia Learning. Second Edition. Cambridge: Cambridge university press.
- Netzwerk Digitale Bildung; Henkelmann, Sarah (2018): Lehren und Lernen mit digitalen Werkzeugen. Ideen für einen zeitgemäßen Unterricht, der Neugier und natürlichen Wissensdrang fördert. Hg. v. Netzwerk Digitale Bildung. Online verfügbar unter <https://www.netzwerk-digitale-bildung.de/wp-content/uploads/NDB-Broschu%CC%88re-Lehren-und-Lernen-mit-digitalen-Werkzeugen-DOWNLOAD.pdf>, zuletzt geprüft am 14.07.2020.
- Ropohl, Mathias; Härtig, Hendrik; Kampschulte, Lorenz; Lindmeier, Anke; Ostermann, Anje; Schwanewedel, Julia (2018): Planungsbereiche für Medieneinsatz im Fachunterricht. In: *MNU* 71 (3), S. 148–155.
- Rutten, Nico; van Joolingen, Wouter R.; van der Veen, Jan T. (2012): The learning effects of computer simulations in science education. In: *Computers & Education* 58 (1), S. 136–153.
- Schulmeister, Rolf (2007): Grundlagen hypermedialer Lernsysteme. Theorie - Didaktik - Design. 4., überarbeitete und aktualisierte Auflage. München, Wien: Oldenbourg Verlag.
- Spiro, R.; Feltovich, P. J.; Jacobson, M. J.; Coulson, R. L. (1995): Cognitive Flexibility, Constructivism and Hypertext: Random Access Instruction for Advanced Knowledge Acquisition. In: *Ill-Structured Domains*.
- Sweller, John (1994): Cognitive load theory, learning difficulty, and instructional design. In: *Learning and instruction* 4 (4), S. 295–312.
- Weidenmann, Bernd (2002): Multicodierung und Multimodalität im Lernprozess. In: Ludwig J. Issing und Paul Klimsa (Hg.): Information und Lernen mit Multimedia und Internet. Lehrbuch für Studium und Praxis. 3., vollst. überarb. Aufl. Weinheim: Beltz PVU, S. 45–62.
- Wieman, Carl E.; Adams, Wendy K.; Loeblein, Patricia; Perkins, Katherine K. (2010): Teaching physics using PhET simulations. In: *The Physics Teacher* 48 (4), S. 225–227.
- Wiesner, Hartmut; Schecker, Horst; Hopf, Martin (2011): Physikdidaktik kompakt: Aulis-Verlag.
- Wolf, Karsten. D. (2018): Viedo statt Lehrkraft? Erklärvideos als didaktisches Element im Unterricht. In: *Computer+ Unterricht* (109), S. 4–7.
- Wolf, Karsten. D.; Kulgemeyer, C. (2016): Lernen mit Videos? Erklärvideos im Physikunterricht. In: *Naturwissenschaften im Unterricht: Physik* 27 (152), S. 36–41.

B Testinstrument zum FDW zum Einsatz digitaler Medien

Aufgabe 1

Zur Erkundung einer harmonischen Schwingung am Federpendel nutzt eine Lehrkraft eine **dy-**
namische Animation (siehe Abbildung).



Quelle: <https://www.geogebra.org/m/KtVV2Rcm> (erstellt mit GeoGebra von C. Wolfseher)

Die Schülerinnen und Schüler sollen nach einer kurzen Demonstration eines realen Federpendels mithilfe dieser Animation den zeitlichen Verlauf der Schwingung eines Federpendels beschreiben und zentrale Begriffe wie Ruhelage, Umkehrpunkte, Auslenkung sowie Amplitude und Periodendauer der harmonischen Schwingung erarbeiten.

Sind die folgenden Aspekte in Bezug auf die beschriebene Mediennutzung typischerweise zutreffend?

| | ja | nein | Quelle/ Idee/ Begründung Supplantionskonzept (Girwidz, 2020, S. 474) |
|--|----|------|---|
| Die dynamische Verknüpfung des Federschwingers und der zugehörigen graphischen Darstellung ergibt ein kohärenteres Bild des physikalischen Phänomens. | x | | |
| Eine statische Veranschaulichung der Schwingung und ihres zeitlichen Verlaufs ohne die Animation würde die kognitive Belastung bei der Verknüpfungsleistung vergrößern und weniger Kapazität für inhaltliche Auseinandersetzungen frei lassen. | x | | |
| Die animierte Darstellung zeigt die sachlogische Struktur des Lerngegenstands auf und garantiert damit tiefere kognitive Aktivitäten bei den Lernenden. | | x | Animation prägt erstmal Oberflächenstruktur und Sichtstruktur; (Girwidz 2013: NiU Ph 137, S.5), „garantiert“ zu absolut |
| Die gleichzeitige Darstellung der Schwingung und des Diagramms führt zu einer unnötigen Verwirrung der Schülerinnen und Schüler. | | x | |

Kodierung:

| Anzahl richtiger Kreuze | Erreichte Punktzahl |
|-------------------------|---------------------|
| 4 | 2 |
| 3 | 1 |
| 0, 1, 2 | 0 |

Aufgabe 2

Handelt es sich im Folgenden um Aspekte, die typischerweise als Vorteile einer digitalen Messwerterfassung (z. B. Systeme von Lehrmittelherstellern mit externen Sensoren oder Nutzung der internen Messsensorik mobiler Endgeräte) gegenüber traditionellen Messmethoden im Physikunterricht aufgefasst werden?

| | ja | nein | Quelle/ Idee/ Begründung |
|---|----|------|---|
| Ermöglichung neuer experimenteller Zugänge | x | | |
| Generelle Reduktion des kognitiven Anspruchs | | x | Digitale MWE kann auch kognitiv anspruchsvoll sein, wenn neues Gerät |
| Überzeugendere Messergebnisse für die Lernenden | | x | Transparenz bei analoger Messung teilweise überzeugender (Konrad 2019, NiU Ph 171/172, S. 22) |
| Einfache Umsetzung von Langzeitmessungen | x | | |
| Fokussierung auf die Dateninterpretation | x | | |
| Auslagerung von Routinearbeiten | x | | |

Kodierung:

| Anzahl richtiger Kreuze | Erreichte Punktzahl |
|-------------------------|---------------------|
| 5, 6 | 2 |
| 4 | 1 |
| 0, 1, 2, 3 | 0 |

Aufgabe 3

Sie behandeln in der Sekundarstufe II den Plattenkondensator und wollen im Unterricht Interesse für das Thema wecken. Handelt es sich bei den folgenden Beschreibungen um einen besonders geeigneten Medieneinsatz für diesen didaktischen Einsatzzweck?

| | ja | nein | Quelle/ Idee/ Begründung |
|--|----|------|--|
| Ein Erklärvideo zum Plattenkondensator schauen lassen, in dem der Aufbau und die Funktionsweise des Kondensators erklärt werden. | | x | Zum Interesse wecken sind Aufbau und Funktionsweise nicht so spannend. |
| Eine Simulation zeigen, die die Entstehung des elektrischen Feldes im Kondensator nach Anlegen einer Spannung visualisiert. | | x | Zu innerphysikalisch |
| Ein digitales Messwerterfassungssystem verwenden, um das Potenzial des Themas zur Anwendung moderner Technologien im Physikunterricht zu zeigen. | | x | |
| Ein kurzes Video aus der Tagespresse vorführen, das aufzeigt, inwiefern der Plattenkondensator im Alltag nützlich sein kann. | x | | |
| Ein digitales Lernquiz zum Einstieg in das Thema durchführen, um mögliche Schülervorstellungen zum Kondensator zu erkunden. | | x | Eigene Vorstellungen zu einem (noch unbekanntem) Thema erkunden, ist nicht unbedingt Interesse fördernd. |

Kodierung:

| Anzahl richtiger Kreuze | Erreichte Punktzahl |
|-------------------------|---------------------|
| 4 | 2 |
| 3 | 1 |
| 0, 1 | 0 |

Aufgabe 4

Eine Lehrkraft lässt Schülerinnen und Schüler im Anfangsunterricht zur Elektrizitätslehre eine Simulation zu einfachen Stromkreisen verwenden. Stellen die folgenden Aussagen sinnvolle Begründungen für diesen Medieneinsatz im Physikunterricht dar?

| | ja | nein | Quelle/ Idee/ Begründung |
|--|----|------|---|
| Durch die Verwendung der Simulation als zentrales Medium zur Erarbeitung einfacher Stromkreise kann die Lehrkraft Zeit und Arbeitsaufwand einsparen, um schneller neue Unterrichtsinhalte durchführen zu können. | | x | Simulationen sollen Experimente nicht ersetzen, sondern sinnvoll ergänzen. |
| Das Überprüfen von vorab aufgestellten Vermutungen kann durch eine einfache und gezielte Variablenkontrolle in der Simulation ermöglicht werden. | x | | Im Realexperiment ggfs. weniger einfach möglich bzw. aufwändiger. |
| Eine geeignete Simulation ermöglicht das Fokussieren auf zentrale Aspekte, indem sichergestellt wird, dass diese ohne überlagernde Nebeneffekte erkennbar werden. | x | | Etwas wie „ihr solltet jetzt eigentlich sehen, dass“ wird vermieden. |
| Den Lernenden wird durch die Verwendung der Simulation anstelle des Realexperiments eine Interaktivität geboten, da die Simulation auf Veränderungen der Parameter reagiert. | | x | Die Interaktivität bietet das Realexperiment auch. |
| Die Simulation kann als Vorbereitung auf ein folgendes analoges Realexperiment genutzt werden, um eine erste Vorstellung der physikalischen Zusammenhänge zu bekommen. | | x | Vorbereitung auf ein Realexperiment könnte zwar sinnvoll sein, aber weniger in dieser einfachen Unterrichtssituation und wenn es das analoge Realexperiment wäre, ist es für dieses auch demotivierend. |

Kodierung:

| Anzahl richtiger Kreuze | Erreichte Punktzahl |
|-------------------------|---------------------|
| 4, 5 | 2 |
| 3 | 1 |
| 0, 1, 2 | 0 |

Aufgabe 5

Stellen die folgenden Beispiele geeignete Einsatzszenarien für mobile Endgeräte zum Experimentieren im Physikunterricht dar?

| | ja | nein | Quelle/ Idee/ Begründung |
|---|----|------|--|
| Versuche zur Einführung von Kern- und Halbschatten mithilfe des Lichtsensors | | x | Visuelles Phänomen spannender als Zahlenwerte des Lichtsensors. |
| Bestimmung der Periodendauer einer harmonischen Schwingung am Federpendel mithilfe des Beschleunigungssensors | x | | z. B. mit phyphox oder auch Hochberg, Kuhn & Müller (2015): NiU 145, S. 18 |
| Bestimmung der Schallgeschwindigkeit mit Mikrofon, Lautsprecher, einer Oszilloskop-App und einer Metronom-App | x | | z. B. mit drei Smartphones oder 1 Smartphone und 2 elektr. Metronome (Hirth et al. 2015: NiU 145, S. 12) |
| Visualisierung der Magnetlinien eines Stabmagneten durch die Auswertung und graphische Aufbereitung von Messwerten des Magnetometers | | x | Eisenspäne wären viel schneller und anschaulicher. |
| Untersuchung der Druckveränderung in einem Gefrierbeutel in Abhängigkeit der auf den Beutel wirkenden Kraft mithilfe des Luftdrucksensors | x | | Beispiel von phyphox; Druck in der Tüte |

Kodierung:

| Anzahl richtiger Kreuze | Erreichte Punktzahl |
|-------------------------|---------------------|
| 4, 5 | 2 |
| 3 | 1 |
| 0, 1, 2 | 0 |

Aufgabe 6

Analysieren Sie den Ausschnitt des Erklärvideos zur Einführung der Beschleunigung in der Sekundarstufe I. (Video einsehbar unter: <https://rwth-aachen.sciebo.de/s/Y3CiHfXkudW9wq1>)
Treffen die folgenden Qualitätsaspekte auf das gezeigte Video zu?

| | ja | nein | Quelle/ Idee/ Begründung |
|---|----|------|---|
| Die Erklärung fokussiert die wesentlichen Aspekte des physikalischen Konzepts der Beschleunigung. | | x | Betrachtet vor allem Alltagsvorstellung und missachtet den Richtungsaspekt. |
| Die Alltagsvorstellungen der Lernenden werden aufgegriffen und zur physikalischen Vorstellung erweitert. | | x | Alltagsvorstellungen aufgreifen ok, aber das Erweitern auf physikalisch korrekte Vorstellung passiert hier nicht. |
| Es werden Beispiele aus einem den Lernenden bekannten Phänomenbereich zur Erklärung verwendet. | x | | (Kulgemeyer, 2018) Auto, freier Fall |
| Gesprochene Erläuterungen und die zugehörigen Darstellungen sind aufeinander abgestimmt (Kontinuität). | | x | zeitlich nicht abgestimmt; Begrifflich: Diagramm zeigt Höhe; gesprochen wird immer von Weg/Strecke (sowieso ungünstig) |
| Das Video verwendet verschiedene Darstellungsformen, welche zum besseren Verständnis des Erklärten beitragen. | x | | |
| Das Video vermeidet redundante Informationen, die eine erhöhte kognitive Beanspruchung verursachen könnten. | | x | Erklärung eines einfachen Bildes und zusätzlich der identische Text gesprochen und geschrieben; |
| Das Video vermeidet Nebeninformationen und Exkurse, die für das zu erklärende Prinzip weniger relevant sind. | | x | unnötiger Exkurs zum Stroboskop |

Kodierung:

| Anzahl richtiger Kreuze | Erreichte Punktzahl |
|-------------------------|---------------------|
| 6, 7 | 2 |
| 4, 5 | 1 |
| 0, 1, 2, 3 | 0 |

Aufgabe 7

Im Folgenden sind Situationen im Physikunterricht aufgeführt, in denen auf die Heterogenität einer Lerngruppe eingegangen werden soll. Zeigen digitale Medien in der jeweiligen Situation dafür ein besonderes Potenzial gegenüber nicht-digitalen Medien?

| | ja | nein | Quelle/ Idee/ Begründung |
|---|----|------|---|
| Vorbereitende Hausaufgabe zur individuellen Erarbeitung erster Grundlagen eines bevorstehenden Unterrichtsinhalts | | x | Könnte ein Erklärvideo sein, aber mit einem Schulbuchtext kann man Grundlagen auch im individuellen Tempo erarbeiten. Den Aspekt der Multimodalität, den ein Video gegenüber dem Schulbuch leisten würde, ist für die Heterogenität weniger relevant. |
| Reduktion mathematischer Anforderungen bei der Auswertung von Experimenten | x | | mit dMWE-Systemen, die math. Schritte (z. B. Regression) übernehmen können (MNU-Heft: Lampe et al. 2015, S. 8) |
| Bereitstellung passender Übungsangebote nach individuell diagnostiziertem Lernstand | x | | Intelligente Lernprogramme, die den SuS je nach Beantwortung individuelle Rückmeldung & entsprechende weiterführende/wiederholende Aufgaben stellen; Mehrwert, weil eine Lehrkraft das so individuell kaum leisten kann. |
| Adressierung unterschiedlicher Interessenslagen von Schülerinnen und Schülern | | x | Spezifische Interessensförderung eher mit Kontexten als über digitale Medien. |

Kodierung:

| Anzahl richtiger Kreuze | Erreichte Punktzahl |
|-------------------------|---------------------|
| 4 | 2 |
| 3 | 1 |
| 0, 1, 2 | 0 |

Aufgabe 8

Handelt es sich im Folgenden um Aspekte, die generell für einen Einsatz von Simulationen im Physikunterricht sprechen?

| | ja | nein | Quelle/ Idee/ Begründung |
|--|----|------|--|
| Simulationen bieten eine präzise Abbildung der Wirklichkeit. | | x | Simulationen stellen oft nur einen Teilaspekt der Realität dar. |
| Simulationen können nicht direkt sichtbare physikalische Sachverhalte visualisieren. | x | | |
| Simulationen können als Beispiel eines wissenschaftlichen Modells thematisiert werden. | x | | |
| Simulationen ermöglichen das Abdecken von Lernzielen durch die Strukturierung des Lernprozesses. | | x | Simulationen benennen selbst noch keine Lernziele (Schulmeister 2007, S. 349). |
| Simulationen bereichern den Lernprozess durch fachbezogene Übungsaufgaben an. | | x | Simulationen geben i.d.R. keinen methodischen Rahmen; Übungen muss LK geben (Girwidz 2020, Schulmeister 2007, S. 349). |
| Simulationen unterstützen die Modellbildung durch das Ausblenden weniger relevanter Aspekte. | x | | |

Kodierung:

| Anzahl richtiger Kreuze | Erreichte Punktzahl |
|-------------------------|---------------------|
| 5, 6 | 2 |
| 4 | 1 |
| 0, 1, 2, 3 | 0 |

Aufgabe 9

Eine Lehrkraft möchte in einer leistungsstarken 10. Klasse ein Erklärvideo einsetzen, um Stöße einzuführen. Handelt es sich im Folgenden um Aspekte, die ein Erklärvideo im gegebenen Fall aufweisen sollte?

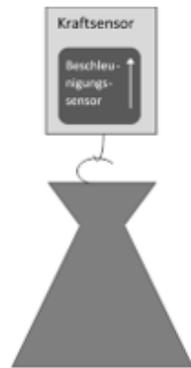
| | ja | nein | Quelle/ Idee/ Begründung |
|---|----|------|--|
| Das Video begleitet die vorkommenden mathematischen Formeln zu den Stößen durch gesprochene Erläuterungen. | x | | Adaptionswerkzeug Mathematisierung (Kulgemeyer 2018) |
| Das Video erklärt die verschiedenen Stoßarten sehr detailliert und enthält viele Zusammenfassungen. | | x | Minimalistische Erklärungen, um kognitive Belastung gering zu halten (Kulgemeyer 2018). |
| Das Video erklärt ein weiterführendes Beispiel für Stöße (z. B. das Rückstoßprinzip beim Raketenantrieb). | | x | Minimalistische Erklärung, keine Exkurse; „erklärt“ nicht „zeigt auf“, sodass es schon ein Exkurs wäre und kein passendes Anwendungsbeispiel zur Einführung. |
| Neue fachsprachliche Wendungen (z. B. inelastisch) werden über Alltagssprache oder bekannte Fachsprache eingeführt. | x | | (Kulgemeyer, 2018) |
| Das Video bietet die präsentierten Informationen sowohl gesprochen als auch in geschriebener Form an. | | x | Zu viele redundante Informationen sind insbesondere für leistungsstarke SuS kognitiv belastend, expertise reversal effect (kalyuga et al.). |
| Das Video zeigt nach der Erklärung der Stöße ein Anwendungsbeispiel aus dem Alltag auf. | x | | Für Wissenserwerb ist Regel-Beispiel-Struktur förderlicher (Kulgemeyer, 2018). |
| Das Video hat im Internet eine sehr gute Bewertung bekommen (z. B. mit Likes auf YouTube). | | x | Likes und Aufrufe hängen nicht mit der Erklärqualität des Videos zusammen (Kulgemeyer & Peters, 2016). |

Kodierung:

| Anzahl richtiger Kreuze | Erreichte Punktzahl |
|-------------------------|---------------------|
| 6, 7 | 2 |
| 4, 5 | 1 |
| 0, 1, 2, 3 | 0 |

Aufgabe 10

Eine Lehrkraft führt in einer Unterrichtsstunde zum Grundgesetz der Mechanik $F = m \cdot a$ ein Experiment durch, in welchem ein Massestück wiederholt angehoben und damit beschleunigt wird. Sie verwendet dazu ein digitales Messwerterfassungssystem mit einem Beschleunigungs- und einem Kraftsensor, um beim dynamischen Vorgang die beschleunigende Kraft und die erzielte Beschleunigung direkt zu messen. Die Messdaten können jeweils als zeitlicher Verlauf sowie in Abhängigkeit zueinander grafisch aufgetragen werden.



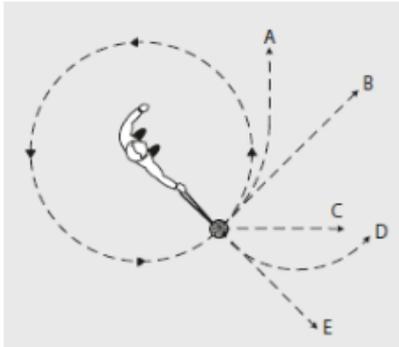
Handelt es sich im Folgenden um typische Zielstellungen (didaktische Funktionen) für die digitale Messwerterfassung im Physikunterricht, unter denen dieses Vorgehen sinnvoll ist?

| | ja | nein | Quelle/ Idee/ Begründung |
|---|----|------|---|
| Transparenz des Messvorgangs für die Lernenden | | x | Der Messvorgang ist weniger transparent über die Sensoren, als es bei einem Kraftmesser der Fall wäre. |
| Auslagerung von Routinearbeiten zur Fokussierung auf die Dateninterpretation | x | | |
| Abdeckung des Motivationsaspekts durch die Nutzung digitaler Messtechnik | | x | Digitale Technik gibt einen Motivationsbonus, aber Motivation sollte auch durch die Fragestellung oder den Inhalt des Experimentes adressiert werden. |
| Zeiteffiziente Untersuchung des physikalischen Zusammenhangs | x | | |
| Verständnisförderung durch die simultane Visualisierung der Messdaten in verschiedenen Darstellungsformen | x | | |

Kodierung:

| Anzahl richtiger Kreuze | Erreichte Punktzahl |
|-------------------------|---------------------|
| 4, 5 | 2 |
| 3 | 1 |
| 0, 1, 2 | 0 |

Aufgabe 11



Quelle: Schecker & Wilhelm, 2018, S. 69

Ein Schüler äußert folgende Aussage zu dieser Abbildung: Wenn der Ball losgelassen wird, verlässt er die Kreisbahn wie in Bahnkurve A, weil er die Kreisbewegung noch eingepreßt hat und nicht direkt verlassen will.

Bei diesem Schüler liegt demnach die Schülervorstellung vor, dass Körper sich ihre Bewegungsform einprägen und diese erst allmählich ausklingt, als hätten die Körper eine „innere Kraft“ gespeichert, die erst langsam aufgebraucht wird.

Sie wollen auf diese Schülervorstellung reagieren: Handelt es sich im Folgenden um geeignete Vorgehensweisen, um die Schülervorstellung mithilfe eines Medieneinsatzes zu adressieren?

| | ja | nein | Quelle/ Idee/ Begründung |
|---|----|------|--|
| Mithilfe einer Videoanalyse eine Kreisbewegung wie im Beispiel untersuchen und auswerten, um dem Schüler die tangentiale Bewegung nach dem Loslassen zu verdeutlichen. | x | | Durch Betrachtung von oben erfolgt 2D-Projektion wie auch in der Abbildung. |
| Ein mobiles Endgerät (in einer gesicherten Verpackung) als zu umherschwingenden Gegenstand verwenden, die Beschleunigung über Fernzugriff aufnehmen und anschließend auswerten. | | x | Medieneinsatz verkompliziert eher; entweder man muss die Bewegung aus der Beschleunigung rekonstruieren oder die Kraft. Außerdem misst man die Beschleunigung immer in mehreren Komponenten. |
| In einer Simulation zur Kreisbewegung die Beschleunigung zum Kreismittelpunkt abrupt entfernen und die tangentiale Bewegung des Körpers beobachten. | | x | S. könnte argumentieren, dass Simulation nicht die Realität wiedergibt, sondern eine Idealisierung. Und: der Einsatz wäre Digitalität um seiner selbst willen. |
| Mit einem Messwerterfassungssystem und Kraftsensoren am Seil zum Mittelpunkt zeigen, dass beim Loslassen das Wegfallen der Zentripetalkraft ohne Verzögerung erfolgt und keine innere Kraft gespeichert wird. | | x | Versuch, die Vorstellung der gespeicherten inneren Kraft zu widerlegen, aber kein überzeugender Konzeptwechsel, weil S behaupten kann, die Sensoren können keine innere Kraft messen |
| Eine ferngesteuerte (günstige) Kamera mit tangentialer Blickrichtung umherschleudern, um die geradlinige Bewegung nach dem Loslassen aus der Perspektive des bewegten Objekts nachvollziehen zu können. | x | | Mehrwert: Perspektive des bewegten Gegenstandes durch Medium erfahrbar machen. |

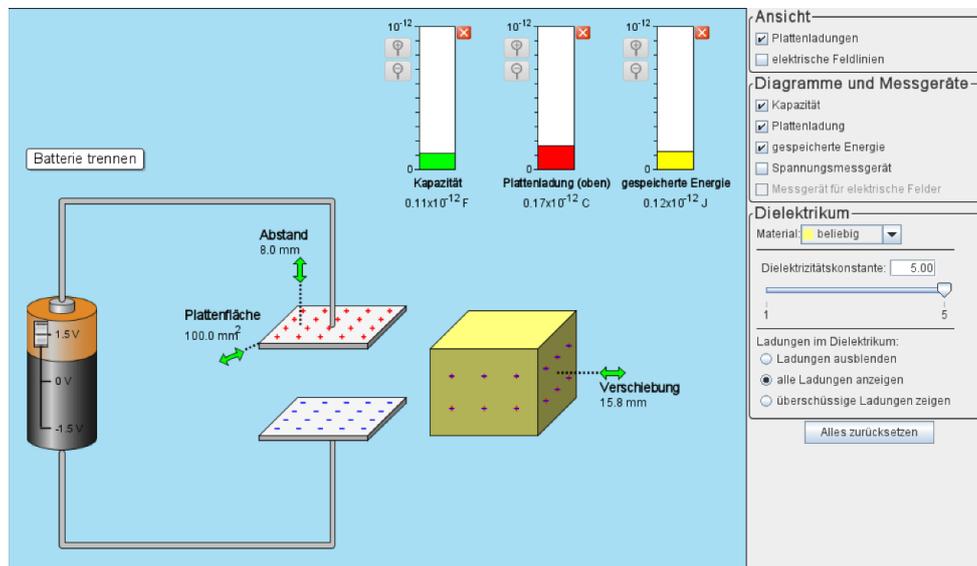
Fortsetzung Aufgabe 11

Kodierung:

| Anzahl richtiger Kreuze | Erreichte Punktzahl |
|--------------------------------|----------------------------|
| 4, 5 | 2 |
| 3 | 1 |
| 0, 1, 2 | 0 |

Aufgabe 12

In einer Unterrichtsstunde zum Plattenkondensator setzt eine Lehrkraft eine Simulation zum entdeckenden Lernen ein, mit der die Schülerinnen und Schüler Einflussgrößen für die Kapazität eines Plattenkondensators erkunden können (siehe Abbildung).



Quelle: Screenshot von <https://phet.colorado.edu/sims/cheerpj/capacitor-lab/latest/capacitor-lab.html?simulation=capacitor-lab&locale=de> (13.03.21), Simulation Kondensatorlabor von PhET Interactive Simulations, University of Colorado Boulder unter der Lizenz CC-BY 4.0

Leistet die ausgewählte Simulation unter den folgenden Gesichtspunkten einen sinnvollen Beitrag zum entdeckenden Lernen?

| | ja | nein | Quelle/ Idee/ Begründung |
|--|----|------|---|
| Möglichkeit zum Überprüfen zuvor aufgestellter Vermutungen | x | | |
| Nutzung verschiedener Repräsentationsformen zur Förderung des konzeptionellen Verständnisses | x | | z. B. durch zusätzliche Graphen zu den Größen |
| Leicht zugängliche Darstellung zur Fokussierung auf relevante Größen | x | | (Richtberg & Girwidz, 2013) |
| Angebot unmittelbar einleuchtender Aufgaben durch die klare Struktur | | x | Simulationen geben keine Aufgabe direkt vor. |
| Visualisierung unsichtbarer Sachverhalte zur Förderung des konzeptionellen Verständnisses | x | | |
| Realistische Abbildung des Realexperiments zur Förderung der Motivation | | x | Komplexreduzierte Darstellung sinnvoll für Fokus auf wesentliche Aspekte; Abbildung von nicht Sichtbarem. |
| Förderung der Kompetenz, relevante Informationen auszuwählen zu können | x | | |

Fortsetzung Aufgabe 12

Kodierung:

| Anzahl richtiger Kreuze | Erreichte Punktzahl |
|-------------------------|---------------------|
| 6, 7 | 2 |
| 4, 5 | 1 |
| 0, 1, 2, 3 | 0 |

Aufgabe 13

Eine Lehrkraft bittet die Klasse zum Einstieg in eine Unterrichtsstunde zur Schwingung eines Fadenpendels, ein in die Schulcloud hochgeladenes Erklärvideo mit Smartphone und Headset anzusehen, um die Lerngruppe zu motivieren. Das Video gibt einen ersten Überblick über die Zusammenhänge zwischen der Periodendauer eines Fadenpendels und ihren Einflussgrößen. Anschließend werden in der Unterrichtsstunde die Einflussgrößen der Periodendauer in einem Schülerexperiment untersucht.

Beurteilen Sie den Einsatz des Erklärvideos in dieser Unterrichtsstunde. Sind die folgenden Aussagen für die beschriebene Situation zutreffend?

| | ja | nein | Quelle/ Idee/ Begründung |
|---|----|------|--|
| Der Einsatz des Erklärvideos erhöht die Motivation und steigert die Lernbereitschaft während der gesamten Unterrichtsstunde. | | x | Motivation beim Experiment umso geringer durch das Vorwegnehmen des Zusammenhangs. |
| Das technische Potential des Erklärvideos wird gut genutzt, da die Schülerinnen und Schüler das Video mit individuellem Tempo (zurückspulen, pausieren) ansehen können. | x | | Potential: zurückspulen, pausieren, individuelles Tempo |
| Das Erklärvideo wird sinnvoll in den Lernprozess eingebettet, indem der zentrale Inhalt der Stunde vorbereitet wird. | | x | Der zentrale Inhalt der Stunde wird eher vorweggenommen. |
| Das Video kann die Abhängigkeit und Unabhängigkeit verschiedener physikalischer Größen auf die Periodendauer des Fadenpendels überzeugend darstellen. | | x | |
| Die erkenntnisgewinnende Funktion des Experiments geht durch die Nutzung des Erklärvideos zum Einstieg verloren. | x | | |

Kodierung:

| Anzahl richtiger Kreuze | Erreichte Punktzahl |
|-------------------------|---------------------|
| 4, 5 | 2 |
| 3 | 1 |
| 0, 1, 2 | 0 |

Aufgabe 14

Handelt es sich im Folgenden um Aspekte, die für den Einsatz einer digitalen Videoanalyse im Physikunterricht sprechen?

| | ja | nein | Quelle/ Idee/ Begründung |
|---|----|------|---|
| Die Nutzung eigener Endgeräte bei einer mobilen Videoanalyse bewirkt eine Fokussierung auf den physikalischen Inhalt. | | x | Gefahr der Ablenkung durch eigene Geräte (Heusler & Laumann, 2020: NiU Ph 179, S. 13). |
| Eine Videoanalyse bietet die Gelegenheit, die Medienkompetenz der Lernenden hinsichtlich verschiedener Aspekte (z. B. Erstellung und Aufbereiten geeigneter Videos) zu fördern. | x | | Nordmeier (2002), NiU Ph 69, S. 27 |
| Im Vergleich zu Experimenten mit klassischen Messverfahren ist eine Videoanalyse grundsätzlich zeitökonomischer. | | x | Kinematik an der Luftkissenbahn wäre z. B. schneller. |
| Durch das Aufnehmen oder Auswählen realer Bewegungsvorgänge aus dem Alltag ist eine Einbindung in lebensnahe und interessante Kontexte einfach umsetzbar. | x | | Nordmeier (2002), NiU Ph 69, S. 27 |
| Die Nutzung verschiedener Repräsentationsformen innerhalb des Verfahrens unterstützt das konzeptionelle Verständnis der Lernenden. | x | | Positive Lernwirkung bezüglich unterschiedlicher Repräsentationsformen und konzeptionellen Verständnis (Becker et al. 2019, S. 2-3) |
| Eine Videoanalyse stellt ein sinnvolles Medium zur Motivation beim Unterrichtseinstieg dar. | | x | |

Kodierung:

| Anzahl richtiger Kreuze | Erreichte Punktzahl |
|-------------------------|---------------------|
| 5, 6 | 2 |
| 4 | 1 |
| 0, 1, 2, 3 | 0 |

C Ergänzende Materialien zum Testinstrument

Zuordnung der Aufgaben in Kategorien und Kognitive Anforderungen

| Aufgabe | Kategorie | Kognitive Anforderung |
|----------------|------------------|------------------------------|
| 1 | Grundlagen | Analysieren |
| 2 | dMWE | Reproduzieren |
| 3 | Grundlagen | Anwenden |
| 4 | Simulationen | Anwenden |
| 5 | dMWE | Anwenden |
| 6 | Erklärvideo | Analysieren |
| 7 | Grundlagen | Analysieren |
| 8 | Simulationen | Reproduzieren |
| 9 | Erklärvideo | Anwenden |
| 10 | dMWE | Analysieren |
| 11 | Grundlagen | Anwenden |
| 12 | Simulationen | Analysieren |
| 13 | Erklärvideo | Analysieren |
| 14 | dMWE | Reproduzieren |

Bewertungsregeln für die Mehrfachwahlaufgaben

die Einträge in der Tabelle geben die erreichte Punktzahl an; Schema in Anlehnung an K' (Kprim) nach Krebs (2004, 2019)

| | | Anzahl an Antwortalternativen in der Aufgabe | | | |
|--------------------|---|--|---|---|---|
| | | 4 | 5 | 6 | 7 |
| Richtige Antworten | 0 | 0 | 0 | 0 | 0 |
| | 1 | 0 | 0 | 0 | 0 |
| | 2 | 0 | 0 | 0 | 0 |
| | 3 | 1 | 1 | 0 | 0 |
| | 4 | 2 | 2 | 1 | 1 |
| | 5 | / | 2 | 2 | 1 |
| | 6 | / | / | 2 | 2 |
| | 7 | / | / | / | 2 |

verwendete Literaturquellen in den Erläuterungen zum Testinstrument

- Becker, S., Klein, P., Gößling, A. & Kuhn, J. (2019). Förderung von Konzeptverständnis und Repräsentationskompetenz durch Tablet-PC-gestützte Videoanalyse. *Zeitschrift für Didaktik der Naturwissenschaften*, 25(1), 1-24.
- Girwidz, R. (2020). Multimedia und digitale Medien im Physikunterricht. In *Physikdidaktik| Grundlagen*. Springer Spektrum, Berlin, Heidelberg, S. 457-527.
- Gramzow, Y. (2015). *Fachdidaktisches Wissen von Lehramtsstudierenden im Fach Physik. Modellierung und Testkonstruktion*. In Niedderer, H., Fischler, H. & Sumfleth, E. (Hrsg.). Studien zum Physik- und Chemielernen, 181. Berlin: Logos Verlag.
- Hochberg, K., Kuhn, J. & Müller, A. (2018). Using Smartphones as Experimental Tools—Effects on Interest, Curiosity, and Learning in Physics Education. *J Sci Educ Technol* 27, 385–403.
- Kulgemeyer, C. (2020). Erklären im Physikunterricht. In *Physikdidaktik| Grundlagen*. Springer Spektrum, Berlin, Heidelberg. S. 403-426.
- Kulgemeyer, C. (2018). Wie gut erklären Erklärvideos? Ein Bewertungsleitfaden. *Computer + Unterricht*, 109, 9-11.
- Lampe, H. U., Liebner, F., Urban-Woldron, H., & Tewes, M. (2015). Innovativer naturwissenschaftlicher Unterricht mit digitalen Werkzeugen. Experimente mit Messerwerterfassung in den Fächern Biologie, Chemie, Physik. *MNU Themenreihe Bildungsstandards*. Neuss: Verlag Klaus Seeberger.
- Rutten, N., Van Joolingen, W. R., & Van Der Veen, J. T. (2014). The learning effects of computer simulations in science education. *Computers & Education*, 58 (1), 136-153.

- Schecker, H., & Wilhelm, T. (2018). Schülervorstellungen in der Mechanik. In *Schülervorstellungen und Physikunterricht*. Springer Spektrum, Berlin, Heidelberg, S. 63-88.
- Schulmeister, Rolf (2007): *Grundlagen hypermedialer Lernsysteme. Theorie - Didaktik - Design*. 4., überarbeitete und aktualisierte Auflage. München, Wien: Oldenbourg Verlag.
- Wieman, C. E., Adams, W. K., Loeblein, P., & Perkins, K. K. (2010). Teaching physics using PhET simulations. *The Physics Teacher*, 48(4), 225-227.

aus unterrichtspraktischen Zeitschriften:

NiU Physik 69 (2002): Neue Medien

- Nordmeier, V. (2002). Videoanalyse von Bewegungen mit dem Computer. *Unterricht Physik*, 13 (69), 27-30.

PdN Physik in der Schule 7/58 (2009): E-Learning

- Kirstein, J. & Nordmeier, V. (2009). E-Learning-Szenario zur Einbeziehung außerschulischer Lernorte. *Praxis der Naturwissenschaften. Physik in der Schule. E-Learning*, (7/58), 16-18.

NiU Physik 137 (2013): Animationen und Simulationen

- Girwitz, R. (2013). Lernen mit bewegten Bildern. Animationen und dynamische Visualisierungen im Physikunterricht. *Unterricht Physik*, 24 (137), 4-9.

- Richtberg, S. & Girwitz, R. (2013). Entdeckendes Lernen mit dem Computer. Möglichkeiten und Anforderungen von Animationen und Simulationen. *Unterricht Physik*, 24 (137), 14-17.

NiU Physik 145 (2015): Experimentieren mit Smartphones und Tablets

- Hirth, M. Gröber, S. Kuhn, J. & Müller, A. (2015). Bestimmung der Schallgeschwindigkeit mit der Differenzmethode. *Unterricht Physik*, 26 (145), 12-14.

- Hochberg, K., Kuhn, J. & Müller, A. (2015). Untersuchung des Federpendels. *Unterricht Physik*, 26 (145), 18-20.

NiU Physik 171/172 (2019): Schlüsselexperimente – real und digital

- Konrad, U. (2019). Elektrizitätslehre. Geräte und Materialien für den Unterricht in der Sekundarstufe I. *Unterricht Physik*, 30 (171/172), 20-24.

- Laumann, D. (2019a). Akustik. Ausgewählte Geräte, Materialien und Medien für den Unterricht. *Unterricht Physik*, 30 (171/172), 62-64.

- Laumann, D. (2019b). Wärmelehre. Ausgewählte Geräte, Materialien und Medien für den Unterricht. *Unterricht Physik*, 30 (171/172), 44-46.

- Weiermann, M. & Wichtrup, P. (2019). Sehen, was man hört. Reale Visualisierung und Darstellung von Schallschwingungen mit dem Smartphone. *Unterricht Physik*, 30 (171/172), 65-68.

NiU Physik 179 (2020): Digitale Bildung

- Heusler, S. & Laumann, D. (2020). Smartphone, Tablet und Notebook: Was eignet sich wofür? Ein Überblick zu allgemeinen und fachbezogenen Aspekten. *Unterricht Physik*, 31 (179), 12-13.

entfernte Aufgaben aus dem Testinstrument

Aufgabe 3 (bis zur Pilotierung; danach als geschlossene Aufgabe zum ersten didaktischen Einsatzzweck)

Im Folgenden sind verschiedene didaktische Einsatzzwecke für eine Unterrichtsstunde zum Plattenkondensator in der Sekundarstufe II aufgeführt. Wählen Sie für jeden didaktischen Einsatzzweck ein passendes Medium aus der Liste aus, dessen Einsatz sich für den jeweiligen Zweck am besten eignet. Beschreiben Sie danach kurz, wie ein entsprechender Medieneinsatz mit diesem Medium für den jeweiligen Zweck aussehen könnte.

| Didaktischer Einsatzzweck | Medien (Auswahl nach Liste, Musterlösung in fett) | Kurze Beschreibung des Medieneinsatzes (Musterlösung in <i>kursiv</i>) |
|--|--|---|
| Interesse für das Thema wecken | <input type="checkbox"/> Übungsprogramm mit adaptiver Zuweisung von Aufgaben <input type="checkbox"/> Digitales Lernquiz <input type="checkbox"/> Video aus der Tagespresse <input type="checkbox"/> Digitale Messwerverfassung <input type="checkbox"/> Digitale Videoanalyse <input type="checkbox"/> Simulation <input type="checkbox"/> Erklärvideo | <i>Ein kurzes Video, welches die Relevanz und den Nutzen von Kondensatoren aufzeigt.</i> |
| Aufbau und Funktion des Plattenkondensators erlernen | <input type="checkbox"/> Übungsprogramm mit adaptiver Zuweisung von Aufgaben <input type="checkbox"/> Digitales Lernquiz <input type="checkbox"/> Video aus der Tagespresse <input type="checkbox"/> Digitale Messwerverfassung <input type="checkbox"/> Digitale Videoanalyse <input type="checkbox"/> Simulation <input type="checkbox"/> Erklärvideo | <i>Ein Erklärvideo zur Bauweise eines Plattenkondensators.</i> |
| Zusammenhang zwischen der Kapazität eines Plattenkondensators und ihren Einflussgrößen untersuchen | <input type="checkbox"/> Übungsprogramm mit adaptiver Zuweisung von Aufgaben <input type="checkbox"/> Digitales Lernquiz <input type="checkbox"/> Video aus der Tagespresse <input type="checkbox"/> Digitale Messwerverfassung <input type="checkbox"/> Digitale Videoanalyse <input type="checkbox"/> Simulation <input type="checkbox"/> Erklärvideo | <i>Eine Simulation, in der Abstand und Fläche der Platten und Isolator zwischen den Platten variiert werden können.</i> |
| Lade- und Entladekurve eines Plattenkondensators für verschiedene Gesamtwiderstände untersuchen | <input type="checkbox"/> Übungsprogramm mit adaptiver Zuweisung von Aufgaben <input type="checkbox"/> Digitales Lernquiz <input type="checkbox"/> Video aus der Tagespresse <input type="checkbox"/> Digitale Messwerverfassung <input type="checkbox"/> Digitale Videoanalyse <input type="checkbox"/> Simulation <input type="checkbox"/> Erklärvideo | <i>Ein computergestütztes Messwerverfassungssystem zur Aufnahme und Analyse des zeitlichen Spannungsverlaufs.</i> |
| Vorwissen zum Thema erkunden | <input type="checkbox"/> Übungsprogramm mit adaptiver Zuweisung von Aufgaben <input type="checkbox"/> Digitales Lernquiz <input type="checkbox"/> Video aus der Tagespresse <input type="checkbox"/> Digitale Messwerverfassung <input type="checkbox"/> Digitale Videoanalyse <input type="checkbox"/> Simulation <input type="checkbox"/> Erklärvideo | <i>Ein digitales Lernquiz zur Eingangsdiagnose des Vorwissens.</i> |

Kodierung:

Pro Aufgabenteil (didaktischer Einsatzzweck) erhält man 0,5 Punkte, falls man entweder

- das Medium gemäß Musterlösung direkt richtig gewählt hat
- eine angemessene Beschreibung/Begründung gibt, auch wenn die Auswahl nicht mit der Musterlösung übereinstimmt.

Erwartungshorizont für die offene Abfrage zu Aufgabe 3 in der Pilotierung:

| Erwartungshorizont korrekt | inkorrekt (allgemein: es wird kein konkreter Einsatz beschrieben; nur Schlagworte) |
|--|--|
| <i>a) Interesse für das Thema wecken mit Video aus der Tagespresse</i> | |
| <p>Ein kurzes Video, welches die Relevanz und den Nutzen von Kondensatoren aufzeigt</p> <p><i>Beispiele:</i></p> <p>„Video, welches den Schülern anschaulich zeigt, was mit einem Kondensator möglich ist.“</p> <p>„Video vorführen und Relevanz für den Alltag hervorheben.“</p> <p>„Zeigt, inwiefern der Plattenkondensator im Alltag nützlich sein kann (Alltagsbezug).“</p> | <p>kein Bezug zur Lebenswelt/zum Alltag, um Interesse zu wecken</p> <p><i>Beispiele:</i></p> <p>„Einstieg. Video zeigen und SuS nach Meinung fragen.“</p> <p>„Betrachtung und darauffolgende Besprechung des Videos.“</p> <p>Es wird kein Einsatz beschrieben, nur Schlagworte „Alltagsbezug“; „Anknüpfung an Alltag“</p> |
| <i>b) Aufbau und Funktion eines Plattenkondensators mit Erklärvideo</i> | |
| <p>Ein Erklärvideo zur Bauweise eines Plattenkondensators ansehen lassen;</p> <p>Inhalt des Videos sollte schlagwortartig genannt werden, wie z. B. Aufbau, Bauweise, Funktion, Fachbegriffe, Skizze/Abbildung etc.</p> <p><i>Beispiele:</i></p> <p>„Ohne Erklärung vorab wird ein kurzes Erklärvideo gezeigt, anschließend wird an der Tafel gemeinsam der Aufbau rekonstruiert und Begrifflichkeiten geklärt, die Funktionen können mündlich gesammelt werden.“</p> <p>„Notizen machen/eine Abbildung eines Plattenkondensators auf Grundlage des Videos beschriften.“</p> <p>„SuS sollen sich Video anschauen und wichtigste Aspekte zusammenfassen oder durch AB-Inhalte sichern.“ (<i>Grenzfall</i>)</p> | <p>kein Bezug zu Inhalten des Videos erkennbar (Aufbau, Bauweise, Funktion, Fachbegriffe, Skizze/Abbildung etc.)</p> <p><i>Beispiele:</i></p> <p>„Erklärvideo anschauen, evtl. mit Pausen und Fragen.“</p> <p>keine Beschreibung des Einsatzes „Häufig gute Videos vorhanden, die das grundlegende Wissen fördern können.“ „Zurückspulen möglich, kann Einzelteile präsentieren.“</p> <p>nur Wiederholung des Einsatzzweckes „Video, welches die Funktion und den Aufbau verständlich aufzeigt.“ (<i>Grenzfall</i>)</p> |

| | |
|--|---|
| <p>c) <i>Zusammenhang zwischen der Kapazität eines Plattenkondensators und ihren Einflussgrößen untersuchen mit Simulation</i></p> | |
| <p>Mit einer Simulation jeweils Abstand und Fläche der Platten und Isolator/Dielektrikum zwischen den Platten variieren; Nennung der Variationsmöglichkeit durch die Simulation relevant; Nennung physikalischer Größen;</p> <p><i>Beispiele:</i> „Hier können Plattenaufbau und -abstand, Spannungsquelle und Dielektrikum variiert werden.“ „Mithilfe einer Simulation können die Schüler auswählen, mit welchen Bauteilen und Materialien sie die Kapazität beeinflussen möchten und so Zusammenhänge herausfinden.“ (Grenzfall)</p> | <p>Einsatz nicht bzw. zu ungenau beschrieben <i>Beispiele:</i> „Ermöglicht eigenständiges Arbeiten der SuS ohne auf Material angewiesen zu sein.“ „Ermöglicht mehrere Dinge gleichzeitig und Schritt für Schritt zu betrachten“ Variationsmöglichkeiten werden nicht beschrieben bzw. zu ungenau (z. B. keine Größen benannt) „Erkunden der Zusammenhänge“ „Messvariablen leicht zu variieren“; „Einfache Variation unterschiedlicher Variablen“; „Die SuS können frei Einflussgrößen variieren und Zusammenhänge erkennen“ (Grenzfälle)</p> |
| <p>d) <i>Lade- und Entladekurve eines Plattenkondensators für verschiedene Gesamtwiderstände untersuchen mit Digitale Messwerterfassung</i></p> | |
| <p>Aufnahme und Analyse des zeitlichen Spannungsverlaufs als Graph mit einem computergestützten MWE-System Bezug zur grafischen Auftragung relevant; ggfs. Option mehrerer Datenreihen zum Vergleich verschiedener R; ggfs. Option schnell viele Daten messe (und aufzeichnen) zu können</p> <p><i>Beispiele:</i> „Messung und Gleichzeitige Aufzeichnung durch Computerprogramm mit anschließender Erstellung eines Diagramms.“ „Als Demoexperiment oder Schülerexperiment können recht schnell t-I oder t-U Kurven aufgenommen werden für verschiedene Widerstände. Die Diagramme können dann verglichen und ausgewertet werden.“</p> | <p>kein Bezug zur grafischen Auftragung und Analysemöglichkeiten von dMWE-Systemen <i>Beispiele:</i> „Messwerte digital erfassen und analysieren“ „Selbst experimentieren, aber mit verlässlichen Messdaten und ohne sich wiederholendes ablesen und notieren.“ kein Einsatz beschrieben oder zu ungenau „Graphen genauer und schneller als handgezeichnete“ „Lade und Entladekurve sichtbar machen“ Unpassend bzw. Bezug zur digitalen MWE unklar „Messwerte grafisch auftragen (dabei auftragen lernen), eigene Messwerte anschaulicher als fremde, bereitgestellte“</p> |
| <p>e) <i>Vorwissen zum Thema erkunden mit Digitalem Lernquiz</i></p> | |
| <p>Eingangsd Diagnose des Vorwissens mit einem digitalen Lernquiz <i>Beispiele:</i> „Prüfen, was die SuS schon wissen. Am Anfang der Lernreihe.“ „Quiz in dem Schülervorstellungen geprüft werden“ „Quiz, welches das Vorwissen abfragt, je nachdem als kleinen Wettbewerb (z. B. Kahoot)“ „verschiedene Fragen (auch mit Freitextfeldern), um Motivation, Kontexte etc. im weiteren Unterrichtsverlauf zu nutzen.“</p> | <p>Diagnoseaspekt nicht deutlich genug; zu ungenau <i>Beispiele:</i> „Multiple choice bogen bearbeiten lassen“ „Lernquiz durchführen“ „Vielleicht wie so eine Art Wer wird Millionär? Zwei Schüler vorne und der Rest der Klasse kann gegebenenfalls als Publikums oder Telefonjoker dienen.“ keine Beschreibung des Einsatzes bzw. Quizes; zu ungenau „Interaktivität“ reine Wiederholung des Einsatzzweckes</p> |

Aufgabe 4 (bis zur curricularen Validierung)

Eine Lehrkraft hat in einer 8. Klasse die Lichtausbreitung und Lichtreflexion thematisiert. Nachfolgend setzt sie zum Einstieg in das Thema Lichtbrechung an der Grenzfläche zweier Medien eine Simulation ein, in der die Schülerinnen und Schüler die Brechungswinkel für verschiedene Medien qualitativ und quantitativ erkunden können. In der Erarbeitungsphase zeigen sich einige Schwierigkeiten bei den Lernenden. Es ergibt sich folgendes Gespräch in der Klasse:

Schüler A: Herr Meier, wir kommen mit der Simulation nicht zurecht.

Schüler B: Wir auch nicht!

Weitere Schülerinnen und Schüler signalisieren Zustimmung.

Lehrkraft: Ihr habt also alle Probleme mit der Simulation? Komisch, ich dachte, die Simulation würde euch Spaß machen.

Schüler B: Nein, mir macht das überhaupt keinen Spaß. Ich verstehe ja gar nicht, was die Simulation genau zeigt.

Schüler A: Ich auch nicht wirklich. Und ich kann mir das mit dieser Brechung auch gar nicht vorstellen. Das ist doch total unrealistisch.

Schüler C: Ich habe auch nicht kapiert, was das mit den verschiedenen Winkeln da auf sich hat. Ich dachte, Licht verläuft immer geradlinig, solange es nicht reflektiert wird.

Schüler D: Das dachte ich auch. Und ich habe auch diese Formel in der Simulation nicht

Analysieren Sie die Situation. Handelt es sich im Folgenden um Möglichkeiten für die auftretenden Probleme bei diesem Medieneinsatz?

| | ja | nein | Quelle/ Idee/ Begründung |
|--|----|------|--|
| Vorausgehende qualitative Realexperimente zum Einstieg in das Phänomen wären für die Lernenden überzeugender. | x | | Auch spannender: z. B. Strohhalm im Wasserglas, Münze in der Tasse |
| Die Lernenden haben noch nicht das nötige Vorwissen, um in der Stunde die Lichtbrechung an der Grenzfläche zweier Medien zu thematisieren. | | x | Vorwissen ist eig. gegeben (Reflexion, Lichtausbreitung). |
| Der Motivationsaspekt durch eine direkte Erfahrung des natürlichen Phänomens fällt aufgrund der Nutzung der Simulation weg. | x | | |
| Schülerinnen und Schüler verbringen heute so viel Zeit mit digitalen Medien, dass sie Realexperimente gegenüber Simulationen bevorzugen. | | x | |

Kodierung:

| Anzahl richtiger Kreuze | Erreichte Punktzahl |
|-------------------------|---------------------|
| 4 | 2 |
| 3 | 1 |
| 0, 1 | 0 |

Aufgabe 5 (bis zur Pilotierung)

In einer Unterrichtsstunde zur Akustik setzen die Schülerinnen und Schüler ihre mobilen Endgeräte und eine Experimentier-App zur Messwertaufnahme und -analyse bei Schall-experimenten ein. Stellen die folgenden Aussagen sinnvolle Argumente für diesen Medi-einsatz im Physikunterricht dar?

| | ja | nein | Quelle/ Idee/ Begründung |
|--|----|------|--|
| Durch die Verwendung von mobilen Endgeräten wird das Interesse der Lernenden beim Experimentieren gefördert. | x | | z. B. Hochberg, Kuhn & Müller (2018) ggfs. aber zu pauschal? |
| Die Nutzung bekannter Geräte bewirkt eine Fokussierung auf den physikalischen Inhalt. | | x | Gefahr der Ablenkung durch eigene Geräte (Heusler & Laumann, 2020: NiU Ph 179, S. 13). |
| Die Datenaufnahme und die anschließende Analyse können in ein und demselben Gerät erfolgen. | x | | Weniger cognitive load oder Zeitaufwand durch Transferieren der Daten. |
| Es braucht keine Einarbeitungszeit zur Durchführung und Analyse der Experimente, weil die Schülerinnen und Schüler der heutigen Generation mit digitalen Medien vertraut sind. | | x | SuS kennen zwar das Gerät, beschäftigen sich in ihrer Freizeit sicherlich nicht mit der App; gerade das Analysieren der Schwingungsbilder ist nicht trivial. |
| Durch die Nutzung mobiler Endgeräte werden keine zeitlichen Kapazitäten für die Förderung des Umgangs mit Experimentiermaterialien benötigt. | | x | Angemessener Umgang mit Experimentier-materialien hat aber trotzdem Relevanz? z. B. ein Oszilloskop zu kennen und nutzen zu können. |
| Die Kombination der akustischen Wahrnehmung und der graphischen Darstellung verschiedener Schallereignisse bietet ein multimodales Lernangebot. | x | | Girwidz 2020, S. 481-482 |

Kodierung:

| Anzahl richtiger Kreuze | Erreichte Punktzahl |
|-------------------------|---------------------|
| 5, 6 | 2 |
| 4 | 1 |
| 0, 1, 2, 3 | 0 |

Aufgabe 13 (bis zur Pilotierung)

Eine Lehrkraft setzt zur Einführung des Wechselwirkungsprinzips ein Erklärvideo ein. Handelt es sich bei den folgenden Aspekten um sinnvolle Argumente, die für diesen Medieneinsatz zur Erklärung des Wechselwirkungsprinzips sprechen?

| | ja | nein | Quelle/ Idee/ Begründung |
|---|----|------|---|
| Ein Video ermöglicht die gleichzeitige Darstellung von vielen verschiedenen Visualisierungen und Animationen, wodurch die Lernenden kognitiv entlastet werden. | | x | Gedanke dahinter: Viele, viele Bilder zum Einstieg, aber kaum sinnvolle Erklärung. Zu viele Animationen erhöhen kognitive Belastung. |
| Im Video können zentrale Aspekte des Wechselwirkungsprinzips wie die Richtungen und Angriffspunkte der Kräfte fachgerecht visualisiert werden. | x | | Sinnvoll hinsichtlich typischer Fehlvorstellungen (gleicher Angriffspunkt, Verwechslung Kräftegleichgewicht). |
| Bei dem Thema handelt es sich um ein neues Prinzip, zu welchem die Lernenden noch kein direkt das Konzept betreffende Vorwissen haben. | x | | (Kulgemeyer, 2018) |
| Da das Prinzip komplex ist und häufig Fehlvorstellungen dazu auftreten, bietet sich eine Erklärung des Prinzips eher an als eine eigenständige Erarbeitung durch die Lernenden. | x | | (Kulgemeyer, 2020, S. 417-418) |
| Im Video kann das Wechselwirkungsprinzip am Beispiel von realen Gegenständen erläutert werden und somit eine mangelnde Ausstattung der Physiksammlung kompensiert werden. | | x | Könnte generell ein Vorteil von Erklärvideos sein, im Fall WW-Prinzip braucht man aber keine Materialien der Physiksammlung, man kann SuS nehmen. |
| Durch die dynamische Visualisierung im Video kann die Gleichzeitigkeit der auftretenden Kräfte bei der Wechselwirkung verdeutlicht werden. | x | | Video erlaubt dynamische Visualisierung: sinnvoll, um Gleichzeitigkeit zu zeigen; entgegen der Vorstellung ERST Kraft, DANN Gegenkraft. |

Kodierung:

| Anzahl richtiger Kreuze | Erreichte Punktzahl |
|-------------------------|---------------------|
| 5, 6 | 2 |
| 4 | 1 |
| 0, 1, 2, 3 | 0 |

Abfragen zur Demographie im Rahmen der Prä-Post-Erhebungen

(exemplarischer Ausschnitt der Druckfahne zum eingesetzten Fragebogen in *Sosci Survey* im Prätest im Wintersemester 2022/23 in Aachen)



DiKoLeP → AC-W22-MZP1

10.04.2025, 10:31

Seite 01

1. Allgemeine Teilnehmerinformationen über die Untersuchung

Liebe Studierende,

wir möchten dazu beitragen, angehende Physiklehrkräfte auf die zunehmende Bedeutung der Digitalisierung und den digitalen Medien in Schule und Unterricht vorzubereiten und sie mit den Potentialen und der lernwirksamen Nutzung digitaler Medien im Physikunterricht vertraut zu machen. Das Projekt *DiKoLeP – digitale Kompetenzen von Lehramtsstudierenden im Fach Physik* der Universitäten in Aachen, Graz, Innsbruck, Paderborn und Tübingen widmet sich daher dem **Angebot und der Evaluation fachdidaktischer Lerngelegenheiten zum Einsatz digitaler Medien im Physikunterricht** in der universitären Lehramtsausbildung.

Die im Rahmen dieser Befragung erfassten Daten werden vertraulich behandelt und anonymisiert. Dazu werden die individuellen Angaben zunächst in pseudonymisierter Form über einen von Ihnen generierten Code gespeichert und ausschließlich zur Evaluation verwendet. Die Beantwortung des Fragebogens birgt keine Risiken für Sie als Teilnehmende und hat zu keinem Zeitpunkt Einfluss auf Ihre Studien- oder Prüfungsleistungen. Wir bitten Sie daher, den Fragebogen eigenständig und ehrlich auszufüllen.

Zu Beginn der Befragung werden einige demografische Angaben zu Ihrer Person, Ihrem Studium und Ihren Lehrerfahrungen erhoben. Anschließend folgt der zweite Teil mit Testaufgaben zum fachdidaktischen Wissen in Physik, zum Einsatz digitaler Medien im Physikunterricht. Im dritten Teil werden Motivation und Einstellungen zum Einsatz digitaler Medien abgefragt. Bitte beachten Sie, dass Sie während der Bearbeitung der Testaufgaben nicht mehr zu einer vorherigen Testaufgabe zurückgehen können!

Am Schluss der Umfrage erfolgt die Abfrage Ihrer E-Mail-Adresse, um Ihrem*r Dozierenden die vollständige Teilnahme an der Befragung rückzumelden. **Ihre E-Mail-Adresse wird dabei getrennt gespeichert und nicht mit Ihren Antworten im Test in Verbindung gebracht.** Bitte beachten Sie, dass die Rückmeldung der vollständigen Bearbeitung über die Angabe der E-Mailadresse nur möglich ist, wenn Sie die Befragung bis zum Ende bearbeiten, nicht also bei einem frühzeitigen Abbruch.

Ihre Antworten werden automatisch gespeichert. Zum Speichern und Beenden können Sie das Browser-Fenster am Ende der Umfrage einfach schließen.

Vielen Dank für Ihre Teilnahme!

Sollten weitere Fragen aufkommen, wenden Sie sich bitte an die verantwortliche Ansprechperson während der Evaluation:

Rike Große-Heilmann, rike.grosse.heilmann@uni-paderborn.de.

Einwilligungserklärung

Ich habe die „Allgemeine Teilnehmerinformationen über die Untersuchung“ zu dieser Studie gelesen und verstanden. Eine Kopie hiervon wurde mir ausgehändigt. Außerdem hatte ich entsprechend Gelegenheit, Fragen (z. B. zu Inhalt, Ziel, Verlauf und Risiken der Studie) zu stellen.

Ich erkläre mich damit einverstanden, an dieser Studie teilzunehmen. Meine Teilnahme erfolgt freiwillig. Ich bin darüber informiert, dass ich die Möglichkeit habe, meine Teilnahme an dieser Studie jederzeit und ohne Angabe von Gründen abzubrechen, ohne dass mir daraus Nachteile entstehen. Außerdem bin ich informiert, dass ich die Möglichkeit habe, die individuellen Daten unter Angabe meines Pseudonyms bis zum Zeitpunkt der vollständigen Anonymisierung zurückzuziehen.

Ich versichere mit meiner Bestätigung, dass ich mindestens 18 Jahre alt bin.

Klausel zum Datenschutz:

Ich erkläre, dass ich mit der im Rahmen dieser Studie erfolgenden Aufzeichnung von Studiendaten und ihrer Verwendung in anonymisierter Form zu wissenschaftlichen Zwecken einverstanden bin.

- Ich stimme der Einwilligungserklärung zu
- Ich lehne die Einwilligungserklärung ab

2. Code Erstellung**Wie erstellen Sie Ihr persönliches Codewort?**

Um Ihre Daten richtig zuordnen zu können, ohne die Geheimhaltung zu verletzen, benötigen wir ein Kenn- oder Codewort. Das Codewort ist so aufgebaut, dass niemand von Ihrem Codewort auf Ihre Person rückschließen kann, auch wir nicht. Sie selbst können Ihr Codewort aber jederzeit rekonstruieren, wenn Sie danach gefragt werden und es vergessen haben sollten. Sie müssen sich hierzu nur an die folgende Regel halten

Dies sind die Bestandteile Ihres Codeworts:

1. Die beiden letzten Buchstaben des Geburtsnamens Ihrer Mutter
2. Die Anzahl der Buchstaben des (ersten) Vornamens Ihrer Mutter
3. Die beiden letzten Buchstaben des (ersten) Vornamens Ihres Vaters
4. Ihr eigener Geburtstag (nur der Tag, nicht Monat und/oder Jahr).

Hinweis:

Bitte schreiben Sie alle Zahlen zweistellig, d.h. wenn nötig mit führender Null.

Bei mehreren oder zusammengesetzten Vornamen berücksichtigen Sie bitte nur den ersten.

Wenn Sie den jeweiligen Namen nicht kennen, schreiben Sie statt der Buchstaben XX bzw. für die Zahl 00.

Beispiel (fiktiv)

- | | | |
|----|----------------------------------|---------------------------------------|
| 1. | Name der Mutter: | Elke-Hannelore Müller geb. Mayerhofer |
| 2. | Name des Vaters: | Wolf-Rüdiger Müller |
| 3. | Ihr Geburtstag: | 09.11.1987 |
| 4. | Daraus ergibt sich als Codewort: | ER04LF09 |

Das Vorgehen ist also:

Die beiden letzten Buchstaben des Geburtsnamens Ihrer Mutter:

Die Anzahl der Buchstaben des (ersten) Vornamens Ihrer Mutter:

Die beiden letzten Buchstaben des (ersten) Vornamens Ihres Vaters:

Ihr eigener Geburtstag (nur der Tag):

3. Bitte geben Sie hier ihren persönlich erstellten Code ein.

4. Bitte geben Sie Ihr Geschlecht an.

Mit welchem Geschlecht identifizieren Sie sich?

[Bitte auswählen] ▾

5. Bitte geben Sie Ihr Alter an.

Alter:

6. An welcher Hochschule / Universität studieren Sie?

Hochschule / Standort: [Hochschule] ▾

7. angestrebter Abschluss bzw. Ausbildungsphase

Bitte geben Sie an, welchen Abschluss sie als Nächstes erwerben bzw. in welcher Ausbildungsphase Sie sich derzeit befinden.

- Bachelor Lehramt Europalehramt Sonstiges
- Master Lehramt

8. Schulabschluss

Bitte geben Sie an, welchen Schulabschluss Sie erworben haben.

Schulabschluss [Bitte auswählen] ▾

9. Abschlussnote

Bitte geben Sie an, mit welcher Note Sie die Mathematik-Matura absolviert haben:

Mathematik-Maturanote:

10. Studienfächer

Bitte geben Sie Ihr zweites Studienfach (neben Physik) im Lehramtsstudium an.

Zweifach

11. Fachsemester

Bitte geben Sie an, in welchem Fachsemester Sie derzeit in Physik studieren.

Wenn Sie z.B.

- 6 Semester im Bachelor studiert haben und

- im 2. Semester im Master sind,

sind Sie im 8. Fachsemester.

Fachsemester:

Fachsemester
Zweifach:

12. Praxiserfahrung

Bitte geben Sie an, ob Sie bereits ein Praxissemester / Praktikum absolviert haben.

| | 3 wöchiges Praktikum | Praxis- semester | Sonstiges Praktikum | kein Praktikum |
|--|--------------------------|--------------------------|--------------------------|--------------------------|
| An welchem Praktikum haben Sie schon teilgenommen? | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

13. Praxiserfahrung

Bitte geben Sie an, ob Sie bereits ein PPS absolviert haben.

| | PPS 1 | PPS 2 | PPS 3 | PPS 4 | kein PPS |
|---|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Bitte geben Sie an, welche Praktika Sie schon in Ihrem Lehramtsstudium Physik absolviert haben. | <input type="checkbox"/> |

14. Medieneinsatz im Praxissemester

Bitte geben Sie an, ob Sie während des Praxissemesters / PPS / Schulpraktikums den Einsatz digitaler Medien kennengelernt haben.

| Beurteilen Sie bitte die folgenden Aussagen zum Einsatz von digitalen Medien bei Ihren Praxiserfahrungen. | nie | in manchen Stunden | bei der Hälfte der Stunden | im Großteil aller Stunden | in allen Stunden | keine Praxis- erfahrung |
|---|-----------------------|--------------------------|----------------------------------|------------------------------------|-----------------------|-------------------------------|
| Wie oft haben Sie beobachtet, dass in Unterrichtsstunden digitale Medien eingesetzt wurden? | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Wie oft haben Sie in Ihren gehaltenen Unterrichtsstunden digitale Medien eingesetzt? | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

15. Abschlussnote

Bitte geben Sie an, mit welcher Note Sie das Abitur abgeschlossen haben.

Abiturnote:

16. Abschlussnote

Welche Abschlussnote haben Sie in Ihrer letzten Ausbildung erreicht?

Abschlussnote

17. Besuchte Lehrveranstaltungen zu digitalen Medien

Welche der folgenden Lehrveranstaltungen haben Sie besucht?

- Medienbildung und Schule – Vorlesung
- Medienbildung und Schule – Seminar neue Medien
- Umgang mit Heterogenität im Physikunterricht
- Lehren und Lernen mit digitalen Medien I
- Fachdidaktik Mechanik und Wärme
- Lehren und Lernen mit digitalen Medien
- Lehren und Lernen mit digitalen Medien im naturwissenschaftlichen Unterricht

D Studien zur Untersuchung der Validität

Hinweise und Interviewfragen in der Think-Aloud-Studie (kognitive Validierung)

Think-Aloud-Studie - Wichtige Hinweise und Informationen

1. Die Studie stellt die Verständlichkeit des neuen Testinstruments in den Vordergrund und nicht deine „Leistung“.
Deine Aussagen helfen uns bei dem weiteren Gestaltungsprozess des Testinstruments.
 2. Spreche alles deutlich aus, was du denkst und gerade tust. Dazu gehören nicht nur deine Lösungen der Aufgaben, sondern auch deine Absichten, Empfindungen, Gedankengänge und Probleme.
 3. Denke daran laut und deutlich zu sprechen.
 4. Wenn du die Aufgabenstellung liest, lies sie bitte laut vor und erwähne, dass es sich um eine Aufgabenstellung handelt und um welche. Beschreibe Bilder in einer Aufgabe bitte kurz und erläutere kurz, wie du das Bild wahrnimmst.

Du könntest deine Sätze zum Beispiel wie folgt formulieren:
„... zuerst lese ich die Aufgabenstellung vor...dabei fällt mir auf...“
„... jetzt überlege ich gerade...“
„...die Aufgabe verstehe ich nicht, weil...“
„...die Bedeutung der Zeichnung ist mir ein Rätsel...“
 5. Eine konkrete Zeitvorgabe gibt es nicht. Wenn deine Konzentration nachlässt, sag Bescheid und wir machen eine Pause oder hören ganz auf.
 6. Falls Du längere Zeit nicht redest, wird der Versuchsleiter dich bitten, weiter Deine Gefühle, Absichten und Handlungen zu erläutern. Der Versuchsleiter gibt jedoch keine fachlichen Hilfestellungen.
 7. Gibt es von deiner Seite noch Fragen?
- Danke für die Teilnahme und viel Spaß!

Die nachfolgenden Interviewfragen wurden im Rahmen der Masterarbeit von Conrads (2022) entwickelt.

Interviewleitfaden:

- Wie würden Sie Ihren Kommilitonen den durchgeführten Test beschreiben?
- Bewerten Sie Ihren Eindruck über die sprachliche Schwierigkeit des Tests
- Bewerten Sie Ihren Eindruck über die inhaltliche Schwierigkeit des Tests und Ihr persönliches Gelingen.
- Haben Sie noch weitere Anmerkungen zu dem durchgeführten Fragebogen?

Anschreiben zur Expert:innenbefragung hinsichtlich der curricularen Passung der Testaufgaben (erste Seite der Befragung)

Befragung zur curricularen Validierung eines Testinstruments zum Fachdidaktischen Wissen zum Einsatz digitaler Medien im Physikunterricht

Liebe Teilnehmende,

die folgende Befragung dient dazu, dass Sie als dozierende Physikdidaktiker*innen die vorliegenden Testaufgaben zum Fachdidaktischen Wissen zum Einsatz digitaler Medien dahingehend einschätzen, inwiefern das notwendige Wissen zur Bearbeitung der Aufgabe in der physikdidaktischen Ausbildung Ihres Hochschulstandorts erworben werden kann.

Auf den folgenden Seiten werden Ihnen die 15 Aufgaben des Testinstruments zur Verfügung gestellt. **Unterhalb** jeder Testaufgabe wird Ihnen eine Frage zur beschriebenen Einschätzung gestellt, die Sie durch Ankreuzen einer 6-stufigen Skala beantworten. Optional haben Sie die Möglichkeit eine Begründung oder Anmerkung zu Ihrer Einschätzung zu geben, falls Sie dies für nötig oder hilfreich erachten. Die Testaufgaben selbst brauchen Sie nicht beantworten, es genügt lediglich die Einschätzung für jede Testaufgabe.

Vielen Dank für Ihre Teilnahme an der Befragung und der Unterstützung meines Projekts!

Rike Große-Heilmann

E Interviewleitfaden und Interviewmaterial

Einverständniserklärung zur Interviewteilnahme und Nutzung der Daten

Da die Interviews digital per Videokonferenz (Zoom) stattfanden, wurde die Erklärung vorab per E-Mail an die Teilnehmenden gesendet. Zu Beginn des Interviews haben die Teilnehmenden ihr Einverständnis auf der jeweiligen Audioaufnahme bestätigt.

Einverständniserklärung – Teilnahme an einer Interviewstudie zum Seminar

Ich erkläre mich damit einverstanden, dass mein pseudonymisierter Code, der im Rahmen der schriftlichen Befragung erhoben wurde, für die individuelle Ausgestaltung des Interviews genutzt wird. Die Rückführung auf meine Person ist nur der Interviewenden bekannt und wird zu keinem anderen Zweck weiterverwendet.

Ich erkläre mich damit einverstanden, dass meine Audioaufnahme, die im Rahmen der Interviewstudie zum Seminar erhoben und transkribiert wird, als Transkript für wissenschaftliche Zwecke anonymisiert weiterverwendet werden darf. Die erhobenen Daten dienen dabei ausschließlich dem wissenschaftlichen Erkenntnisgewinn.

Interviewleitfaden (Rohform)

(für jedes Interview wurde der Leitfaden individuell angepasst, um Antworten in der Kurzumfrage (Hauptphase A) sowie individuelle Testantworten (Hauptphase B) einzubinden)

Vorab:

- angenehme und offene Gesprächsatmosphäre schaffen
- ermutigen, eigene Einschätzungen, Aussagen oder Antworten gerne immer direkt zu begründen; dann brauche ich weniger nachzufragen
- Einverständnis zur Audioaufzeichnung klären

Einstieg:

Erfahrung bei der Fragebogenbearbeitung (1 Frage, mögliche Nachfragen in kursiv)

Um den Lernzuwachs im Seminar reflektieren zu können, haben Sie an einer Befragung teilgenommen. Heute möchte ich mit Ihnen über den Einsatz digitaler Medien im Physikunterricht sprechen, wozu Sie entsprechende Fragen einmal vor und einmal nach der Lehrveranstaltung beantwortet haben.

1. Wie erging es Ihnen beim ersten Durchführen des Fragebogens? Wie erging es Ihnen beim zweiten Mal? Erzählen Sie mal.
 - a. *Wie sicher waren Sie sich bei der Beantwortung?*
 - b. *bei Unterschieden: Woran könnte das gelegen haben?*
 - c. *Sie haben eben gesagt, dass Sie auch an Inhalte des Seminars gedacht haben. Nennen Sie dazu ein Beispiel.*

Hauptteil:

A: Einschätzungen zum Seminar (4 Fragen, mögliche Nachfragen in kursiv)

2. Was ist Ihnen am Seminar besonders in Erinnerung geblieben? Erzählen Sie mal.
3. Das Seminar war aufgeteilt in einen theoretischen Block mit einigen Input-Sitzungen und einen anschließenden praktischeren Block. Welchen Block fanden Sie eher lernförderlich, welchen weniger lernförderlich? Begründen Sie Ihre Antwort. *Seminarplan vorlegen (Material A)*
4. Wir wollen nun nacheinander auf verschiedene Aspekte aus dem Seminar eingehen. Sie haben vorab eine Tabelle ausgefüllt und bewertet, inwiefern Sie einzelne Inhalte oder Aspekte aus dem Seminar beim Erwerb von Kompetenzen zum Einsatz digitaler Medien unterstützt haben. *Tabelle vorlegen (B)*

****Bei Aspekten, an die sich nicht erinnert wurde****

Sie haben sich an manche Dinge nicht mehr erinnern können. Schätzen Sie bitte noch diese Aspekte nachträglich ein. *Stimuli vorlegen (C)*

****Bei Aspekten, die als eher nicht/ nicht/ gar nicht unterstützend eingeschätzt wurden****

Sie haben den Aspekt „X“ als eher nicht/nicht/gar nicht unterstützend eingeschätzt. Begründen Sie Ihre Einschätzung.

- a. *Was hätte für Sie anders sein müssen, damit diese Aktivität oder dieser Aspekt hilfreicher gewesen wäre?*

****Bei Aspekten, die als unterstützend/ sehr unterstützend eingeschätzt wurden****

Sie haben den Aspekt „X“ als unterstützend/sehr unterstützend eingeschätzt. Begründen Sie Ihre Einschätzung.

5. (nur in Aachen): Sie haben vor dem Interview angegeben, welche Medien Sie im praktischen Teil des Seminars verwendet haben. Warum haben Sie diese Medien für Ihren Lernzirkel ausgewählt? *Tabelle vorlegen (B)*

Fortsetzung auf nächster Seite

5. (nur in Graz und Tübingen): Sie haben vor dem Interview angegeben, welche Medien Sie in der praktischen Auseinandersetzung vertieft haben. Warum haben Sie diese Medien ausgewählt?
- Hatten Sie vorher bereits Erfahrungen mit diesen Medien gemacht?*
 - Vergleichen Sie die beiden Praktischen Auseinandersetzungen. Welche ist Ihnen leichter gefallen? Welche fanden Sie besser? Begründen Sie.*

B: Individuelle Vergleiche der Testantworten in Vor- und Nachtest

(Frage 6 mehrmals für verschiedene Aufgaben stellen => Anzahl der Fragen und Nachfragen hängt von den individuellen Testergebnissen ab)

6. Wir wollen nun über ein paar Ihrer Antworten im Fragebogen zu den unterschiedlichen Zeitpunkten sprechen, d.h. vor und nach dem Seminar.
Bei Aufgabe X1 hat sich Ihre Einschätzung vom ersten Zeitpunkt zum zweiten verändert. Können Sie erläutern, wieso Sie anders geantwortet haben? *Aufgabe + Testantwort vorlegen*
- Aufgabe X1 adressiert das Thema „Y“. Inwiefern hat die Seminarveranstaltung zu diesem Thema Ihre Einschätzung geändert?*
 - Inwiefern hat die Seminarveranstaltung zum Thema „Y“ Ihre Einschätzung geändert?*
6. Bei Aufgabe X2 hat sich Ihre Einschätzung vom ersten Zeitpunkt zum zweiten verändert. Können Sie erläutern, wieso Sie anders geantwortet haben? *Aufgabe + Testantwort vorlegen*
- Aufgabe X2 adressiert das Thema „Y“. Inwiefern hat die Seminarveranstaltung zu diesem Thema Ihre Einschätzung geändert?*
 - Inwiefern hat die Seminarveranstaltung zum Thema „Y“ Ihre Einschätzung geändert?*
- ...
7. Sie haben in den Aufgaben zum Bereich „Z“ die stärkste Verbesserung erreicht. Können Sie erklären, woran das liegen könnte?
- Inwiefern hat die Seminarveranstaltung zum Thema „Z“ Ihre Einschätzung geändert?*

C: Berücksichtigung individueller oder standortspezifischer Besonderheiten (1 Frage, mögliche Nachfragen in kursiv)

8. Wir haben nun viel über das Seminar und Ihre Antworten gesprochen. Gibt es auch andere Situationen, die Ihnen bei der abschließenden Befragung nach dem Seminar geholfen haben? Nennen Sie ein Beispiel.
- Haben Sie neben dem untersuchten Seminar weitere Erfahrungen mit dem Einsatz digitaler Medien im Unterricht in den Bildungswissenschaften gemacht?*
 - Haben Sie neben dem untersuchten Seminar weitere Erfahrungen mit dem Einsatz digitaler Medien im Unterricht in der Fachdidaktik ihres Zweifachs gemacht?*
 - Haben Sie neben dem untersuchten Seminar weitere Erfahrungen mit praktischen Erprobungen gemacht?*
 - Inwiefern hat Sie das in der Bearbeitung der Aufgaben beeinflusst?*
 - Haben Sie über die Universität hinaus Erfahrungen gemacht, die Ihnen bei der abschließenden Befragung geholfen haben?*

Ausstieg:

Wünsche der Studierenden für ein Seminkonzept / Sonstiges (2 Fragen, mögliche Nachfragen in kursiv)

9. Beschreiben Sie, welche Aspekte Sie für ein Seminar zum Einsatz digitaler Medien im Physikunterricht für die Zukunft wünschen würden?
- Was hätte Ihnen geholfen, um die Befragung nach dem Seminar sicherer beantworten zu können?*
 - Haben Sie Verbesserungsideen für die Gestaltung des Seminars?*
10. Gibt es noch etwas, dass Sie zu diesem Thema noch sagen möchten?

Interviewmaterial A: Seminarplan

Seminarplan Aachen

| Seminarplan Lehr-Lern-Seminar | | |
|-------------------------------|--|---------|
| Sitzung | Inhalt | |
| 1 | Orga, Einführungsfragebogen, OER | |
| 2 | Grundlagen zu digitalen Medien | Theorie |
| 3 | Simulationen, IBE und AR-Experimente | |
| 4 | Digitale Messwerterfassung | |
| 5 | Schülervorstellungen und Videoanalyse | |
| 6 | Erklärvideos | |
| 7 | Vorstellung Analyse Erklärvideo + Arbeitsblatterstellung | |
| 8 | Vorstellung Lernzirkelstationen | |
| 9 | Vorstellung Arbeitsblätter | |
| 10 | Feedback der SuS zur Lerneinheit | |
| 11 | Probelauf Lernzirkelstationen | |
| 12 | <i>Erprobung Lernzirkel in der Schule</i> | |
| 13 | Reflexion der Erprobung | |
| 14 | Abschlussfragebogen | |

Seminarplan Graz und Tübingen

Seminarplan

| Inhalt | |
|--|---------|
| 1 Grundlagen zu digitalen Medien | Theorie |
| 2 Erklärvideos | |
| 3 Videoanalyse | |
| 4 Digitale Messwerterfassung | |
| 5 IBE, AR und VR-Experimente | |
| 6 Simulationen und Animationen, Modellbildung | |
| 7 Selbstständige Vertiefung mit Aufgaben | Praxis |
| 8 Planung von Unterricht | |
| 9 Vorführung exemplarischer Unterrichtssequenzen | |
| 10 Reflektion über Unterrichtssequenzen | |

Phasen:

- a) Theoretische Einführung
- b) Ausprobieren der Medien
- c) Einsatzbeispiele
- d) Studienergebnisse

Interviewmaterial B: Tabelle zur Einschätzung der empfunden Lernförderlichkeit

(hier exemplarisch für das Seminar in Aachen im Wintersemester 2022/23)

| Schätzen Sie ein, inwiefern Sie diese Inhalte/Tätigkeiten im Seminar beim Erwerb von Kompetenzen zum Einsatz digitaler Medien unterstützt haben. <i>Allgemein, einflührend oder übergreifend</i> | Dieser Inhalt / diese Tätigkeit hat mich beim Kompetenzerwerb... | | | | | | |
|---|--|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|------------------------------|
| | sehr unterstützt | unterstützt | eher unterstützt | eher nicht unterstützt | nicht unterstützt | gar nicht unterstützt | Ich kann mich nicht erinnern |
| Kennzeichen von Multimedia (Multimodalität, Multicodierung, Interaktivität) erkunden | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Designprinzipien zum Multimedialen Lernen und der Cognitive Load Theory | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Empirische Befunde zum Einsatz digitaler Medien im Unterricht | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Beispiele aus dem Physikunterricht zu Ebenen des SAMR-Modells | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Abgrenzung von digitalen Experimentierformaten (Simulation, IBE, AR-Experiment) | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| <i>Simulationen</i> | | | | | | | |
| Möglichkeiten von Simulationen und Einsatzszenarien für den Physikunterricht | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Empirische Befunde zum Einsatz von Simulationen im Unterricht | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Ausprobieren / Vorstellung einer Simulation im Seminar & Diskussion | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| <i>Digitale Messwertfassung</i> | | | | | | | |
| Vergleich zweier Unterrichtsszenarien mit analoger bzw. digitaler Datenaufnahme | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| pädagogischer und fachdidaktischer Mehrwert bei der Nutzung digitaler MWE | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| mögliche Einschränkungen oder Anforderungen bei der Nutzung dMWE | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Möglichkeiten der Messsensorik von Mobilgeräten und deren Einsatz im PU | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Ausprobieren einer dMWE (Smartphone-Experiment oder Wärmebildkamera) | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Beispiele zur Videoanalyse im Physikunterricht | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Ausprobieren einer mobilen Videoanalyse | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| <i>Erklärvideos</i> | | | | | | | |
| Betrachtung eines Erklärvideos und Diskussion hinsichtlich der fachlichen Richtigkeit | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Modell zum Erklärvorgang und Anforderungen für effektive Erklärungen | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Übertragbarkeit der Elemente guter Erklärungen auf Erklärvideos | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Bewertung eines Erklärvideos anhand 6 Kategorien von Gütekriterien | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| didaktische Einsatzmöglichkeiten von Erklärvideos im Unterricht | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Empirische Befunde zu Erklärvideos in Physik | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Analyse eines Erklärvideos aus dem Internet | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Fortsetzung auf nächster Seite

Gab es Tätigkeiten oder Inhalte im Seminar, die den Erwerb von Kompetenzen zum Einsatz digitaler Medien für Sie erschwert haben? Wenn ja, welche?

Abfragen zum praktischen Teil des Seminars

Welche (digitalen) Medien hast du mit deiner Gruppe in eurer Lernereinheit eingesetzt?

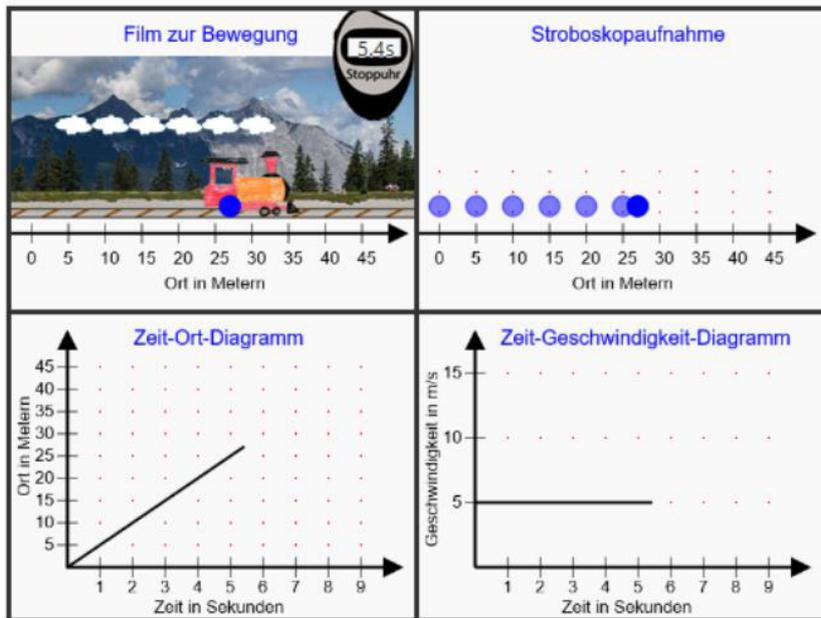
Welche (digitalen) Medien hast du an der von dir betreuten Station eingesetzt?

| Inwiefern hat Sie diese Tätigkeit beim Erwerb von Kompetenzen zum Einsatz digitaler Medien unterstützt? | Dieser Inhalt / diese Tätigkeit hat mich beim Kompetenzerwerb ... | | | | | | |
|---|---|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|------------------------------|
| | sehr unterstützt | unterstützt | eher unterstützt | eher nicht unterstützt | nicht unterstützt | gar nicht unterstützt | Ich kann mich nicht erinnern |
| Auseinandersetzung mit digitalen Medien für die Planung der eigenen Station | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Vorstellung der eigenen Station/ Materialien im Seminar und Feedback dazu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Probedurchlauf der Station im Seminar | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Betreuung der Station und Einsatz des Mediums in der Erprobung an der Schule | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Reflexion der Erprobung anhand der Selbsteinschätzung | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Reflexion der Erprobung anhand des Schüler-Feedbacks und der Beobachtung | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Interviewmaterial C: Stimuli zu Seminarinhalten

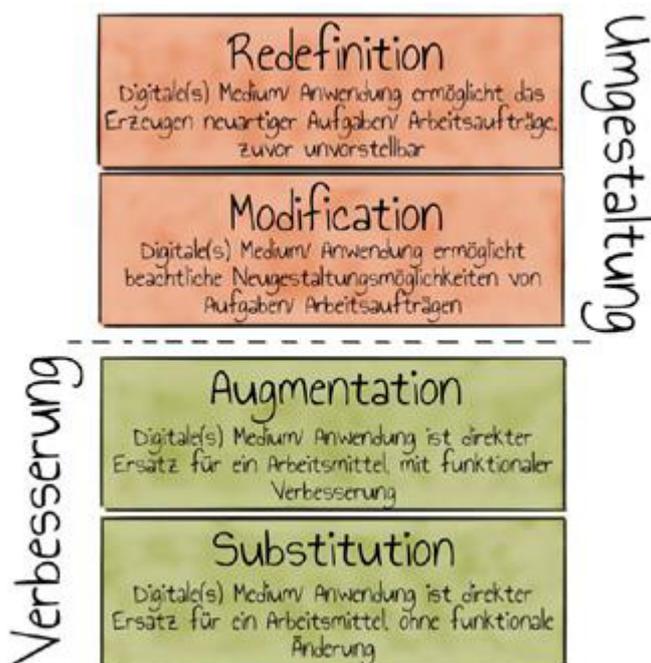
z. B. Abbildungen oder Inhalte aus Seminarfolien, Arbeitsaufträge etc., die nach Bedarf eingesetzt wurden, um Erinnerung an bestimmte Seminarelemente hervorzurufen.

Beispiel zur Multicodierung



(https://www.didaktik.physik.uni-muenchen.de/multimedia/lernen_mit_multimedia/multimod_multicod_interak/multicodierung/index.html; letzter Zugriff: 29.11.2021)

SAMR-Modell

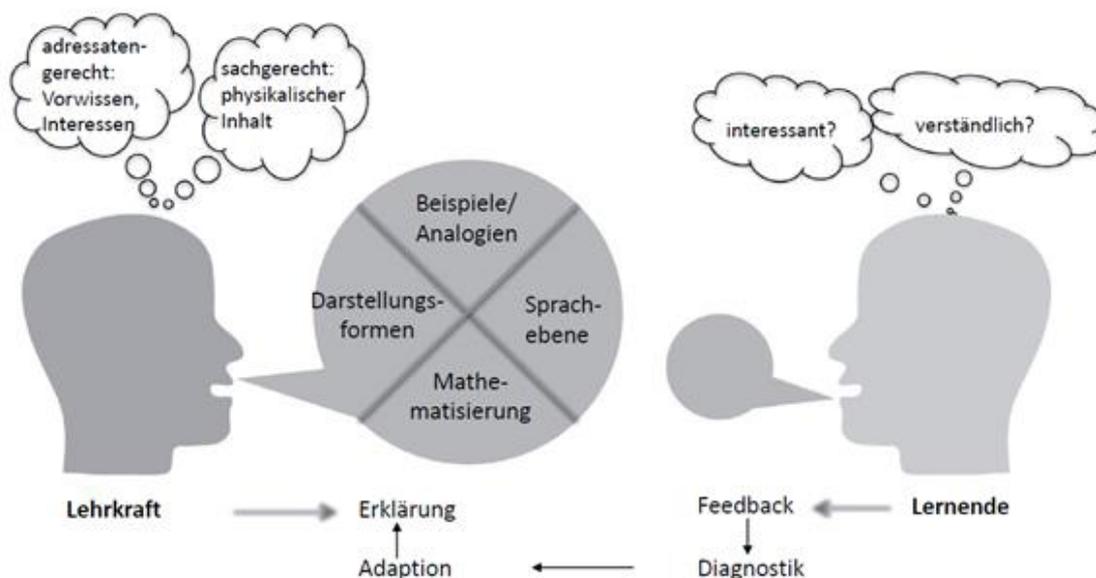


SAMR-Modell nach Puentedura (2006, 2012) von Toolbox Lehrerbildung CC BY SA 4.0

Aufgabe zur Abgrenzung von digitalen Experimentierformaten

| Kriterium | Film/Video | Simulation | IBE | AR-Exp. |
|-----------------------|------------|------------|-----|---------|
| Repräsentation | | | | |
| Künstliches Bild | | x | | x |
| Reales Bild | x | | x | x |
| Datenquelle | | | | |
| Mathematisches Modell | | x | | X |
| Reales Experiment | x | | x | X |
| Struktur | | | | |
| Linear | x | | | |
| Verzweigt | | x | x | x |
| Interaktion | | | | |
| Indirekt | x | | | |
| Direkt | | x | x | x |

Modell zum Erklärvorgang



(nach Kulgemeyer & Schecker, 2013; Kulgemeyer & Peters, 2016)

Didaktische Einsatzmöglichkeiten von Erklärvideos im Unterricht

| | | Produzenten | |
|-------------|--------------------------|---|--|
| | | Lehrkräfte | Schülerinnen und Schüler |
| Rezipienten | Lehrkräfte | <ul style="list-style-type: none"> • „Peer-to-Peer“: didaktische Lehrkräfte-Weiterbildung durch Kolleginnen und Kollegen | <ul style="list-style-type: none"> • Pädagogische Diagnostik • Leistungsbewertung |
| | Schülerinnen und Schüler | <ul style="list-style-type: none"> • SuS lernen von Erklärprofis; beliebiges Wiederholen • Vertiefung oder Wiederholung • im Flipped Classroom-Konzept | <ul style="list-style-type: none"> • SuS lernen von ihren Mit-SuS (Fachwissen) • Lernen beim Erklären (Kommunikationskompetenz) • Alternative Erklärungen, Förderung leistungsstarker oder Unterstützung leistungsschwacher SuS |

(nach Wolf & Kulgemeyer, 2016)

F Interviewauswertung und Kategoriensystem

Zusammenfassung der Vorgehensweise zur Auswertung der Interviews

| | |
|-------------------------------------|--|
| Ausgangsfrage | <i>FF3: Welche Elemente des gemeinsamen Kerns bzw. der standortspezifischen Teile des Lehrkonzepts sind besonders lernförderlich oder eher lernhinderlich und stehen mit Verbesserung oder Verschlechterung im gemessenen FDW zum Einsatz digitaler Medien in Verbindung?</i> |
| Ziel der Untersuchung | <ul style="list-style-type: none">• als wirksam empfundene Lerngelegenheiten im Seminar erkunden• Ursachen für Veränderungen in den Testantworten zum FDW zum Einsatz digitaler Medien beim Vor- und Nachtest untersuchen• Zusammenhang von der Veränderung im Antwortverhalten und der wahrgenommenen Lernwirksamkeit verschiedener Seminarelemente untersuchen• lernförderliche und lernhinderliche Elemente des Lehrkonzepts identifizieren |
| Datenmaterial | 19 Interviewtranskripte aus Interviews mit Teilnehmenden des untersuchten Seminars sowie der schriftlichen Prä-Post-Befragung im Seminar; zusätzliches Begleitmaterial je Interview: beantwortete Kurzumfrage und im Interview gezeigte Veränderungen in einer (individuellen) Auswahl an Testaufgaben |
| Inhaltsanalytische Einheiten | <ul style="list-style-type: none">• Eine Kodiereinheit (Textstelle, die kodiert wird) ist eine Sinneinheit, aber mindestens ein ganzer Satz. In Hauptphase B wird immer die gesamte Textstelle, in der die Befragungsperson über eine veränderte Antwort spricht, kodiert.• Die Kontexteinheit bildet ein Interviewtranskript mitsamt dem zugehörigen Begleitmaterial (Kurzumfrage, Testantworten).• Die Analyseeinheit bildet das gesamte Interviewmaterial entlang der Fragen des Interview-Leitfadens. |

| | |
|--|---|
| Kodierregeln | <ol style="list-style-type: none"> 1. Kodierungen werden nur im Interviewtranskript vorgenommen. Das Begleitmaterial dient zum „Nachschlagen“ von im Interview angesprochenen Aspekten (z. B. Antwortalternativen im Test, gegebene Antwort in der Kurzumfrage). 2. Es werden in der Regel Sinneinheiten kodiert (vgl. Kodiereinheiten), welche auch mehrere Sätze enthalten können. 3. Es können einer Textstelle mehrere Kategorien zugewiesen werden, da in einer Sinneinheit mehrere Aspekte angesprochen werden können, die jeweils anderen Sub- oder Hauptkategorien zugehören. Die Subkategorien <i>Verbesserung</i> und <i>Verschlechterung</i> schließen sich gegenseitig aus. |
| Hauptkategorien und mögliche Subkategorien (deduktiv) | <ul style="list-style-type: none"> • Testbewertung <ul style="list-style-type: none"> ○ Verbesserung; Verschlechterung • Änderungen in Testantworten <ul style="list-style-type: none"> ○ konkrete Seminarerfahrung; konkrete Erfahrung (nicht im Seminar) • Seminar als Lerngelegenheit <ul style="list-style-type: none"> ○ theoretischer Teil; praktischer Teil; gemeinsamer Kern; standortspezifisch Aachen; standortspezifisch Graz/Tübingen • Lernförderlichkeit • Andere Lerngelegenheiten <p>→ induktive Bildung von Subkategorien (v. a. für Hauptkategorie <i>Änderungen in Testantworten</i>) zur thematischen Strukturierung</p> <p>→ Überschneidungen der Kodierungen in <i>Testbewertung</i> und <i>Änderungen in Testantworten</i> zur Identifikation förderlicher und hinderlicher Seminarelemente</p> |

Übersicht zur Beschreibung und Anwendung der Hauptkategorien

| Hauptkategorie | Art der Bildung | Inhaltliche Beschreibung | Anwendung der Kategorie |
|-----------------------------|-----------------|---|---|
| Testbewertung | deduktiv | Aussagen, die die Veränderungen von einzelnen Antworten im Prätest zu Antworten im Posttest aufzeigen. | <ul style="list-style-type: none"> - wird nur im individuellen Teil des Interviews (Hauptphase B) kodiert - dient zur Triangulation mit den quantitativen Testergebnissen |
| Änderungen in Testantworten | deduktiv | Aussagen, die Veränderungen von Antworten im Prätest zu Antworten im Posttest begründen (oder beschreiben). | <ul style="list-style-type: none"> - hauptsächlich in Hauptphase B des Interviews, wenn konkret die veränderten Antworten der Probanden angesprochen werden - falls an anderer Stelle ein Bezug zu einer Änderung in den Testantworten genannt wird - wird nicht kodiert, wenn es nur um allgemeine Aspekte zum Test geht, die keinen Vorher-Nachher-Vergleich herstellen (→ <i>Testinstrument</i>) |
| Seminar als Lerngelegenheit | deduktiv | Aussagen, die einen konkreten oder indirekten Bezug zum untersuchten Seminar "Lehr-Lern-Gelegenheit im Physikunterricht" (Aachen) bzw. "Moderne Medien im Physikunterricht" (Graz, Tübingen) aufzeigen. | <ul style="list-style-type: none"> - wird v. a. in Hauptphase A (Einschätzungen zum Seminar) und im Ausstieg (Wünsche für ein Seminarkonzept) kodiert - Stellen, in denen das Seminar konkret angesprochen wird (z. B. bei einer Begründung für eine Antwortänderung) - wenn indirekt auf Erfahrungen oder Inhalte im Seminar Bezug genommen wird, ohne das Seminar oder den Inhalt konkret zu nennen - bei Verbesserungsvorschlägen, die konkret an das Seminar gerichtet sind (z. B. im Ausstieg) |

| | | | |
|--|----------|---|--|
| Subjektiv wahrgenommene Lernförderlichkeit | deduktiv | Aussagen, die in Bezug auf das Seminar eine Bewertung der (wahrgenommenen) Lernförderlichkeit vornehmen (aus Sicht der Studierenden). | <ul style="list-style-type: none"> - wird v. a. in Hauptphase A (Einschätzungen zum Seminar) kodiert - es muss offensichtlich um das untersuchte Seminar gehen sowie eine Einschätzung gegeben sein, inwiefern ein bestimmtes Element förderlich oder nicht förderlich war - im Ausstieg (Wünsche für ein Seminarkonzept) nur anwenden, wenn ein Bezug zum Ist-Zustand hergestellt wird |
| Andere Lerngelegenheiten | deduktiv | Aussagen, die einen konkreten oder indirekten Bezug zu anderen Lerngelegenheiten neben dem untersuchten Seminar aufzeigen. | <ul style="list-style-type: none"> - wird v. a. in Hauptphase C (Berücksichtigung individueller oder standortspezifischer Besonderheiten) kodiert - weiterhin an jeder Stelle, in der konkret eine andere Lehrveranstaltung oder eine Lernerfahrung angesprochen wird (z. B. bei einer Begründung für eine Antwortänderung). |
| Zukünftiger Unterrichtsalltag/ Lehrberuf | induktiv | Aussagen, die einen konkreten Bezug zum zukünftigen Unterrichten als Lehrkraft aufzeigen. | <ul style="list-style-type: none"> - wird kodiert, wenn auf das eigene zukünftige Unterrichtshandeln Bezug genommen wird oder Konsequenzen diesbezüglich gezogen werden |
| Testinstrument/ Testaufgaben | induktiv | Aussagen, die einen direkten Bezug zur Beantwortung des Tests oder den Testaufgaben zum FDW digitale Medien aufzeigen (im allgemeinen Sinne). | <ul style="list-style-type: none"> - wird kodiert, wenn allgemeine Erfahrungen bei der Testbeantwortung geäußert werden, die nicht den Vergleich zwischen Antworten im Prä- und Posttest betreffen - weiterhin, wenn Aussagen in Bezug zur Güte des Tests stehen (ungünstige Formulierung, Missverständnisse, sprachliche Schwierigkeit, Raten etc.) |

Interratering zur Hauptkategorie *Änderungen in Testantworten*

Ergebnistabelle zur Berechnung der Interrater-Reliabilität

| | | Person 1 | | |
|----------|---|----------|--------|-----|
| | | 1 | 0 | |
| Person 2 | 1 | a = 522 | b = 52 | 574 |
| | 0 | c = 48 | 0 | 48 |
| | | 570 | 52 | 622 |

a: Anzahl der Übereinstimmungen

b: nur die von Person 2 kodierten Segmente

c: nur die von Person 1 kodierten Segmente

Berechnung der Interrater-Reliabilität nach Brennan und Prediger (1981):

$$P(\text{observed}) = P_o = a / (a + b + c) = 0.84$$

$$P(\text{chance}) = P_c = \text{Anzahl der Codes} / (\text{Anzahl der Codes} + 1)^2 = 13 / (13 + 1)^2 = 0.07$$

(bei ungleicher Anzahl an Codes pro Segment oder bei Auswertung eines Codes allein)

$$\kappa_n = (P_o - P_c) / (1 - P_c) = 0.83$$

Kategoriensystem zu den Subkategorien der fokussierten Hauptkategorien *Testbewertung und Änderungen in Testantworten*

(Kategoriendefinitionen gemäß des Schemas nach Kuckartz, 2018; Tabelle 7-1)

Subkategorien der Hauptkategorie „Testbewertung“

| Subkategorie | Verbesserung |
|--------------------------|--|
| Oberkategorie | Testbewertung |
| Inhaltliche Beschreibung | Alle Aussagen, die die Veränderungen von einzelnen Antworten im Prätest zu Antworten im Posttest aufzeigen, in denen die Veränderung der Antwort im Test einer Verbesserung entspricht. |
| Anwendung der Kategorie | Die Kategorie wird nur in Hauptphase B des Interviews angewendet, in dem konkret die veränderten Antworten der Probanden angesprochen werden. Dient zur Triangulation mit den quantitativen Testergebnissen, um zu identifizieren, wann die Veränderung der Antwort zu einer Verbesserung im Test geführt hat. |
| Abgrenzung | Die Kategorie wird nicht kodiert, wenn die veränderte Antwort im Test einer Verschlechterung entspricht, dann <i>Verschlechterung</i> kodieren. |
| Subkategorie | Verschlechterung |
| Oberkategorie | Testbewertung |
| Inhaltliche Beschreibung | Alle Aussagen, die die Veränderungen von einzelnen Antworten im Prätest zu Antworten im Posttest aufzeigen, in denen die Veränderung der Antwort im Test einer Verschlechterung entspricht. |
| Anwendung der Kategorie | Die Kategorie wird nur in Hauptphase B des Interviews angewendet, in dem konkret die veränderten Antworten der Probanden angesprochen werden. Dient zur Triangulation mit den quantitativen Testergebnissen, um zu identifizieren, wann die Veränderung der Antwort zu einer Verschlechterung im Test geführt hat. |
| Abgrenzung | Die Kategorie wird nicht kodiert, wenn die veränderte Antwort im Test einer Verbesserung entspricht, dann <i>Verbesserung</i> kodieren. |

Subkategorien der Hauptkategorie „Änderungen in Testantworten“

| Subkategorie | Konkrete Seminarerfahrung |
|---------------------------|--|
| Oberkategorie | Änderungen in Testantworten |
| Inhaltliche Beschreibung | Aussagen zu Veränderungen von Prä- zu Posttest-Antworten, bei denen direkt Bezug zum untersuchten Seminar (Lehr-Lern-Seminar oder Seminar "Moderne Medien im Physikunterricht) genommen wird. |
| Anwendung der Kategorie | Wenn konkret ein Aspekt, ein Inhalt oder eine Erfahrung aus dem untersuchten Seminar als Begründung für veränderte Antworten genannt wird. Dies kann sich auf die Inhalte aus dem Theorieteil sowie auch auf die Erfahrungen im praktischen Teil und bei der Unterrichtserprobung beziehen. |
| Beispiele für Anwendungen | <i>"Deswegen habe ich jetzt hier nein angekreuzt, beim anderen ja und vor allen Dingen dadurch das ich jetzt, da wir jetzt auch Simulationen eingesetzt haben im Unterricht, das hat quasi so den Eindruck nochmal verändert, was man im Unterricht machen kann, was nicht und wie lange Dinge dauern und ja, Dinge dauern länger als ich eigentlich angenommen habe."</i> (A21a02, Pos. 62) |
| Abgrenzung | Die Kategorie wird nicht kodiert, ... wenn kein konkreter Bezug zu einer Lerngelegenheit deutlich wird und höchstens ein indirekter Bezug vermutet werden kann, der zu einer inhaltlich begründeten Veränderung der Antwort geführt hat, dann <i>Anwendung medienbezogenen Wissens/Seminarbezug vermutbar</i> kodieren. ... wenn von der Befragungsperson nur eine Vermutung für die Veränderung geäußert wird, die das Seminar generell nennt und keine konkrete Erfahrung. |

| Subkategorie | Konkrete Erfahrung (nicht im Seminar) |
|---------------------------|--|
| Oberkategorie | Änderungen in Testantworten |
| Inhaltliche Beschreibung | Aussagen zu Veränderungen von Prä- zu Posttest-Antworten, bei denen direkt Bezug zu einer Lehrveranstaltung oder Lerngelegenheit genommen wird, die nicht das untersuchte Seminar darstellt. |
| Anwendung der Kategorie | Wenn konkret ein Aspekt, ein Inhalt oder eine Erfahrung aus einer anderen Lehrveranstaltung oder Lerngelegenheit als Begründung für verändertes Antwortverhalten genannt wird. |
| Beispiele für Anwendungen | <i>"Bei der Zwei bei der ersten Antwort hier ist es definitiv so, dass ich irgendwann (...) dieses eine Interview zu den Techniksachen (...) gemacht //Ok ja// und hatte dazu eh musste ich dann halt ein bisschen planen und da war ein Video von der Tagespresse dabei, das war unglaublich langweilig und eh das hat hier meine Entscheidung definitiv beeinflusst muss ich sagen" (A21a02, Pos. 46)</i> |
| Abgrenzung | Die Kategorie wird nicht kodiert, ... wenn konkrete Inhalte oder Aspekte aus dem untersuchten Seminar als Begründung aufgeführt werden, dann <i>konkrete Seminarerfahrung</i> kodieren. ... wenn kein konkreter Bezug zu einer Lerngelegenheit deutlich wird und höchstens ein indirekter Bezug vermutet werden kann, der zu einer inhaltlich begründeten Veränderung der Antwort geführt hat, dann <i>Anwendung medienbezogenen Wissens</i> kodieren. |

| Subkategorie | Anwendung medienbezogenen Wissens |
|---------------------------|--|
| Oberkategorie | Änderungen in Testantworten |
| Inhaltliche Beschreibung | Aussagen zu Veränderungen von Prä- zu Posttest-Antworten, die unter Berücksichtigung von Wissen zum Einsatz digitaler Medien reflektiert begründet werden, ohne aber konkret auf eine Erfahrung im untersuchten Seminar oder einer anderen Lehrveranstaltung einzugehen. |
| Anwendung der Kategorie | Wenn kein konkreter Aspekt aus dem untersuchten Seminar (oder anderen Lehrveranstaltung) als Begründung für das Verändern der Antwort genannt wird, aber dennoch eine reflektierte Entscheidungsänderung beschrieben wird für die Wissen zum Einsatz digitaler Medien herangezogen wurde. Es kann dabei auch ein indirekter Seminarbezug naheliegend sein (kein konkreter Inhalt), muss es aber nicht. |
| Beispiele für Anwendungen | "(...) ich habe es jetzt beim zweiten Mal und auch hier kann ich mich daran erinnern, weil ich (..) überlegt habe, ob ich nein oder ja ankreuzen soll, ehm das ist das technische Potential des Erklärvideos insofern, als dass die Schüler das Video- eh das ist halt einfach das technische Potential, nicht das didaktische Potential. Ich glaube beim ersten Mal habe ich das nicht getrennt und beim zweiten Mal schon. (...)" (A21a02, Pos. 50) |
| Abgrenzung | <p>Die Kategorie wird nicht kodiert,</p> <p>... wenn konkrete Inhalte oder Aspekte aus dem untersuchten Seminar als Begründung aufgeführt werden, dann <i>konkrete Seminarerfahrung</i> kodieren.</p> <p>... wenn konkrete Inhalte oder Aspekte aus einer anderen Lehrveranstaltung als Begründung aufgeführt werden, dann <i>konkrete Erfahrung (nicht im Seminar)</i> kodieren.</p> <p>... wenn die Veränderung in der Antwort inhaltlich, aber nicht unter Nutzung von medienbezogenen Wissen begründet wird (z. B. durch rein fachdidaktisches Wissen), dann <i>Anwendung nicht-medienbezogenen Wissens</i> kodieren.</p> <p>... wenn die Veränderung in der Antwort primär nicht inhaltlich oder ohne Berücksichtigung von Wissen zum Einsatz digitaler Medien begründet wird (z. B. sprachlich, Interpretation der Fragestellung), dann <i>Uminterpretation der Fragestellung, Unentschlossenheit oder Sonstiges</i> kodieren.</p> |
| Subkategorien | <ul style="list-style-type: none"> • Seminarbezug vermutbar (wird kodiert, sobald ein genereller oder indirekter Bezug zum Seminar oder einem Seminarinhalt naheliegt bzw. zu vermuten ist, dass die reflektierte Begründung auf einem Seminarinhalt beruht) • kein Seminarbezug vermutbar (wird kodiert, wenn „Seminarbezug vermutbar“ nicht kodiert wird) |

| Subkategorie | Anwendung nicht-medienbezogenen Wissens |
|---------------------------|---|
| Oberkategorie | Änderungen in Testantworten |
| Inhaltliche Beschreibung | Aussagen zu Veränderungen von Prä- zu Posttest-Antworten, in denen nicht-medienbezogenes Wissen zur Begründung einer Antwort oder Veränderung herangezogen wird (sondern z. B. rein fachdidaktisches Wissen). |
| Anwendung der Kategorie | Wenn das Verändern mit einer inhaltlich reflektierten Entscheidungsänderung beschrieben wird, für die primär Wissen herangezogen wird, das keinen Bezug zum Einsatz digitaler Medien hat (z. B. „rein“ fachdidaktisches Wissen ohne Medienbezug). |
| Beispiele für Anwendungen | <i>„(...) Okay, also auf das letzte kann ich sofort eingehen, das mit dem Lernquiz: Also da habe ich einfach mich zurück in meine Schülerrolle versetzt und ich hätte keinen Plan gehabt, was ein Kondensator ist, und da bin ich aber sicher nicht der einzige, von dem her denke ich mal, ist da ein Lernquiz zum Einstieg wahrscheinlich nicht so praktisch.“</i> (G22b09, Pos. 97) |
| Abgrenzung | Die Kategorie wird nicht kodiert, ... wenn die Veränderung in der Antwort inhaltlich, aber unter Nutzung von medienbezogenen Wissen begründet wird, dann <i>Anwendung medienbezogenen Wissens</i> kodieren. ... wenn die Veränderung in der Antwort primär nicht inhaltlich begründet wird (z. B. sprachlich, Interpretation der Fragestellung), dann <i>Uminterpretation der Fragestellung, Unentschlossenheit</i> oder <i>Sonstiges</i> kodieren. |

| Subkategorie | Uminterpretation der Fragestellung |
|---------------------------|---|
| Oberkategorie | Änderungen in Testantworten |
| Inhaltliche Beschreibung | Aussagen zu Veränderungen von Prä- zu Posttest-Antworten, die hauptsächlich auf ein verändertes Verständnis der Fragestellung einer Aufgabe oder Formulierung einer bestimmten Antwortoption zurückzuführen sind. |
| Anwendung der Kategorie | Wenn die Begründung zur Änderung einer Antwort eher in der sprachlichen Auffassung und dem Verständnis der Fragestellung oder der Antwortoption liegt und nicht überwiegend inhaltlich durch Wissen zum Einsatz digitaler Medien im Unterricht begründet wird. |
| Beispiele für Anwendungen | <i>„So jetzt die zweite Antwort muss ich grad nochmal gucken (...) genau, weil du da oben geschrieben hast ‚besonders geeignet‘. Ob das jetzt besonders geeignet ist, weiß ich halt nicht, deswegen habe ich erstmal nein angekreuzt. Habe ich anscheinend vorher anders eingeschätzt, aber das war jetzt vielleicht nicht aus dem Seminar (...)“ (A22a05, Pos. 69)</i> |
| Abgrenzung | Die Kategorie wird nicht kodiert, ... wenn die Veränderung in der Antwort reflektiert inhaltlich begründet wird und dabei Wissen zum Einsatz digitaler Medien herangezogen wird, dann <i>Anwendung medienbezogenes Wissen</i> kodieren. Die Kategorie wird nicht kodiert, wenn die Veränderung ausschließlich auf eine Unentschlossenheit oder Raten zurückzuführen ist, dann <i>Unentschlossenheit</i> kodieren. |

| Subkategorie | Unentschlossenheit |
|---------------------------|---|
| Oberkategorie | Änderungen in Testantworten |
| Inhaltliche Beschreibung | Aussagen zu Veränderungen von Prä- zu Posttest-Antworten, die auf Unentschlossenheit, Zufall oder Raten der Antwort zurückzuführen sind. |
| Anwendung der Kategorie | Wenn die Änderung einer Antwort nicht inhaltlich begründet wird, sondern explizit Unentschlossenheit oder Raten als Grund für die unterschiedlichen Entscheidungen in Prä- und Posttest genannt werden. |
| Beispiele für Anwendungen | <p>„deswegen finde ich es schwierig hier ja oder nein anzukreuzen und ich glaube das ist einfach purer Zufall das ich beim einen nein beim anderen ja angekreuzt habe.“ (A21a02, Pos. 48)</p> <p>„das war ja die Frage, wo ich mir einfach beide Mal nicht sicher war- also es ist- ich habe es mir einfach nicht vorstellen können, ob das jetzt Sinn macht oder nicht. Ehm das war wieder so eine Frage, wenn es da einen Mittelpunkt gegeben hätte, hätte ich wahrscheinlich beide Male den angekreuzt.“ (G21b02, Pos. 98)</p> |
| Abgrenzung | Die Kategorie wird nicht kodiert, ... wenn die Veränderung in der Antwort zwar ebenfalls nicht inhaltlich begründet wird, aber eher auf das (sprachliche) Verständnis der Frage oder Antwortoption zurückzuführen ist, dann wird <i>Uminterpretation der Fragestellung</i> kodiert. |
| Subkategorie | Keine Änderung begründet |
| Oberkategorie | Änderungen in Testantworten |
| Inhaltliche Beschreibung | Aussagen zu veränderten Testantworten, in denen eigentlich keine Begründung der Veränderung gegeben wird. |
| Anwendung der Kategorie | Wenn eher eine aktuelle Einschätzung der Antwortoption gegeben wird, aber kein Bezug zu der veränderten Einschätzung zwischen beiden Messzeitpunkten oder wenn z. B. nur inhaltlich begründet wird, warum im Posttest ja oder nein gewählt wurde, aber nicht, wieso die Einschätzung sich verändert hat. |
| Beispiele für Anwendungen | „Gesprochene Erläuterungen und die zugehörigen Darstellungen sind aufeinander abgestimmt ja eher nicht, ne? Also das passt finde ich nicht immer genau zusammen auch generell finde ich das Video nicht gut einfach.“ (G21b01, Pos. 139) |
| Subkategorien | <ul style="list-style-type: none"> • inhaltliche Begründung eines Messzeitpunkts (s. unten) • keine inhaltliche Begründung eines Messzeitpunkts (wird kodiert, wenn „inhaltliche Begründung eines Messzeitpunkts“ nicht kodiert wird) |

| Subkategorie | inhaltliche Begründung eines Messzeitpunkts |
|---------------------------|--|
| Oberkategorie | Änderungen in Testantworten\keine Änderung begründet |
| Inhaltliche Beschreibung | Aussagen zu Veränderungen von Prä- zu Posttest-Antworten, in denen eine inhaltliche Begründung von einer der beiden Testantworten gegeben wird (aber nicht zur Veränderung) |
| Anwendung der Kategorie | Wenn nur eine Antwort eines Messzeitpunkts begründet wird, aber nicht, wieso die Einschätzung sich verändert hat. Dabei kein ein indirekter Bezug zum Seminar naheliegen. |
| Beispiele für Anwendungen | <i>"Es kommt auf die Aufbereitung an, zum Plattenkondensator, weil man jetzt einfach den Plattenkondensator nur in der Selbsterarbeitungsphase macht. So habe ich beim ersten Mal gedacht, dann wird's wahrscheinlich schwierig mit den Vermutungen. Nach dem Seminar habe ich mir gedacht, ja, wenn es kleine Einführungen gibt, dann kann man das definitiv gut überprüfen, welche Vermutungen man da aufgestellt hat."</i> (G22b06, Pos. 93) |
| Abgrenzung | Die Kategorie wird nicht kodiert, ... wenn die Begründung nicht inhaltlich ist, sondern sprachlich, dann nur <i>keine Änderung begründet</i> oder <i>Uminterpretation der Fragestellung</i> (falls Änderung begründet wird) kodieren. |
| Subkategorien | <ul style="list-style-type: none"> • Seminarbezug vermutbar (wird kodiert, sobald ein genereller oder indirekter Bezug zum Seminar oder einem Seminarinhalt naheliegt bzw. zu vermuten ist, dass die reflektierte Begründung auf einem Seminarinhalt beruht) • kein Seminarbezug vermutbar (wird kodiert, wenn „Seminarbezug vermutbar“ nicht kodiert wird) |
| Subkategorie | keine inhaltliche Begründung eines Messzeitpunkts |
| Oberkategorie | Änderungen in Testantworten\keine Änderung begründet |
| Anwendung der Kategorie | Wird bei Textstellen der Subkategorie „keine Änderung begründet“ genau dann kodiert, wenn „inhaltliche Begründung eines Messzeitpunkts“ nicht kodiert wird. |

| Subkategorie | Nicht nutzbare Textstellen |
|--------------------------|--|
| Oberkategorie | Änderungen in Testantworten |
| Inhaltliche Beschreibung | Reste-Kategorie: alle Textstellen der Hauptkategorie <i>Änderungen in Testantworten</i> , in denen keine der anderen Subkategorien passend ist und die für die weitere Analyse inhaltlich nicht nutzbar sind. |
| Anwendung der Kategorie | <p>z. B. bei fehlender Erinnerung (an Veränderung, Prätest etc.)</p> <p>z. B. wenn Proband:in einen Fehler in der Beantwortung äußert oder dass er/sie nicht mehr der Einschätzung wie im Posttest ist, sodass nun eig. keine Veränderung vorliegt</p> <p>z. B. wenn der Eindruck entsteht, dass Proband:in die Fragestellung nicht verstanden hat und die Begründung daher nicht wirklich nutzbar ist</p> |
| Subkategorien | <ul style="list-style-type: none"> • fehlende Erinnerung (wird kodiert, wenn der/die Proband:in äußert oder andeutet, sich nicht an den Grund für die Veränderung der Testantwort oder eine Beantwortung im Test zu erinnern) • Sonstiges (wird kodiert, wenn „fehlende Erinnerung“ nicht kodiert wird) |

Verteilung der kodierten Segmente zu den Subkategorien aus der Hauptkategorie *Änderungen in Testantworten*

| Subkategorien aus <i>Änderungen in Testantworten</i> | | kod. Segmente (von 314) |
|---|---|------------------------------------|
| 1 | konkrete Seminarerfahrung | 107 |
| 2 | konkrete Erfahrung (nicht im Seminar) | 12 |
| 3 | Anwendung medienbezogenen Wissens | 62 |
| 3.1 | Seminarbezug vermutbar | 45 |
| 3.2 | kein Seminarbezug vermutbar (Dummy zu 3.1) | 17 |
| 4 | Anwendung nicht-medienbezogenen Wissens | 7 |
| 5 | Uminterpretation der Fragestellung | 29 |
| 6 | Unentschlossenheit | 22 |
| 7 | keine Änderung begründet | 71 |
| 7.1 | inhaltliche Begründung eines Messzeitpunkts | 47 |
| 7.1.1 | Seminarbezug vermutbar | 16 |
| 7.1.2 | kein Seminarbezug vermutbar (Dummy zu 7.1.1) | 31 |
| 7.2 | keine inhaltliche Begründung eines Messzeitpunkts (Dummy zu 7.1) | 24 |
| 8 | Nicht nutzbare Textstellen | 49 |
| 8.1 | keine Erinnerung | 15 |
| 8.2 | Sonstiges (Dummy zu 8.1) | 34 |

Teilweise wurden zur Übersichtlichkeit Unterkategorien als Dummy-Kategorien ergänzt, damit alle kodierten Segmente aus der Kategorienebene in der darunterliegenden Ebene kodiert werden. Z. B. wurde für alle Textstellen der Kategorie 7 (keine Änderung begründet) die Kategorie 7.2 (keine inhaltliche Begründung eines Messzeitpunkts) immer genau dann kodiert, wenn 7.1 (inhaltliche Begründung eines Messzeitpunkts) nicht kodiert wurde. Die Dummy-Kategorien 7.1.1, 7.2 und 8.1 wurden nach dem Interrating ergänzt, dies stellte jedoch nur eine strukturelle und keine inhaltliche Anpassung des Kategoriensystems dar, sodass von derselben Güte hinsichtlich der Beurteiler:innenübereinstimmung ausgegangen werden kann.

G Kategorisierung zur Identifikation lernförderlicher und lernhinderlicher Seminarelemente

Kategorien zu Indizien für lernförderliche Seminarelemente

| Oberkategorie | Einsatz / Erfahrung im Seminar | | |
|--------------------------|---|------------------------|---------------|
| Unterkategorien | Ausprobieren | Praxis eigener Einsatz | Praxis andere |
| Inhaltliche Beschreibung | <p>Veränderte Antwort wird mit einer Erfahrung zum Einsatz eines digitalen Mediums im Rahmen des Seminars begründet. Das kann im Theorieteil im Sinne eines <i>Kennenlernens</i> oder <i>Ausprobierens</i> von Medien sein oder im praktischen Seminarteil (<i>eigene praktische Auseinandersetzung/Erprobung</i> oder die der <i>anderen Gruppen</i>).</p> <p>Bei Unsicherheit der Zuordnung in eine Unterkategorie (weil z. B. nicht erkennbar, ob Erfahrung beim Ausprobieren eines Mediums oder in der praktischen Auseinandersetzung/Erprobung erfolgte) wird nur die Oberkategorie kodiert.</p> | | |

| Oberkategorie | Designprinzipien Multimedialernen / Cognitive Load Theory | |
|--------------------------|---|---------------------------------------|
| Unterkategorien | Anwendung auf Erklärvideos | Allgemeiner Bezug zu Designprinzipien |
| Inhaltliche Beschreibung | <p>Veränderte Antwort wird mit den Designprinzipien zum Multimedialernen bzw. der Cognitive Load Theory begründet. Dabei kann es sein, dass die Prinzipien in einer Testaufgabe <i>auf Erklärvideos angewendet</i> werden. Wird in der Begründung der Probanden von der Sitzung zu Erklärvideos oder den Gütekriterien zu Erklärvideos gesprochen, wird ebenfalls die Unterkategorie "<i>Seminarinhalt zu einem Medium im Theorieteil - Erklärvideos</i>" vergeben.</p> | |

| Oberkategorie | Seminarinhalt zu einem Medium im Theorieteil | | |
|--------------------------|---|--------------|--------------|
| Unterkategorien | Digitale Messwerterfassung | Simulationen | Erklärvideos |
| Inhaltliche Beschreibung | <p>Veränderte Antwort wird konkret mit einem theoretischen Seminarinhalt (z. B. Gestaltungsmerkmale, Gütekriterien, typische Vorteile oder Möglichkeiten) zu einem bestimmten Medium begründet (d. h. in der Regel aus dem Theorieteil). Dies kann sich auf die Inhalte zur <i>digitalen Messwerterfassung</i> (inkl. Smartphones oder Videoanalyse), <i>Simulationen</i> (inkl. IBE) oder <i>Erklärvideos</i> beziehen.</p> <p>Wird nur kodiert, wenn der Seminarinhalt oder die zugehörige Seminarsitzung in der Antwort/Begründung genannt wird (nicht, wenn der Bezug zu einem Seminarinhalt nur vermutet wird).</p> <p>Wird die Veränderung weniger durch einen theoretischen Inhalt zum Medium, sondern durch das Ausprobieren des Mediums oder Beispielen zum Medium erklärt (z. B. Smartphone-Experiment, Beispiele von Simulationen), wird "<i>Einsatz/Erfahrung im Seminar - Ausprobieren</i>" vergeben.</p> | | |

| Oberkategorie | Unkonkreter Bezug oder Allgemeines | | |
|--------------------------|--|------------------------|-------------------------|
| Unterkategorien | Kritischere Sichtweise | Akzeptanz für Vorteile | Diskussionen im Seminar |
| Inhaltliche Beschreibung | <p>Begründung der veränderten Antwort legt eine kritischere Ansicht bzgl. eines Medieneinsatzes, Beispiels oder Ähnlichem nahe (diese kritischere Ansicht kann z. T. konkret benannt werden, aber auch indirekt aus der Begründung zu lesen sein).</p> <p>Bei vorher eher kritischer Haltung gegenüber einem Medium (oder fehlendem Wissen zu Möglichkeiten des Mediums), wird nach dem Seminar z. B. ein typischer Vorteil des Mediums eher akzeptiert bzw. eingesehen.</p> <p>Veränderte Antwort wird mit Diskussionen (z. B. auch mit Kommilitonen) im Rahmen des Seminars begründet.</p> | | |

Kategorien zu Indizien für lernhinderliche Seminarelemente

(oder für typische Fehler/Probleme in der Argumentation der Studierenden)

| Oberkategorie | konkrete Inhalte im Seminar (falsch angewendet, verallgemeinert etc.) | | | | |
|--------------------------|---|------|--------------|--------------|--------|
| Unterkategorien | CLT/MML | dMWE | Simulationen | Erklärvideos | Praxis |
| Inhaltliche Beschreibung | <p>In der Begründung zur veränderten Antwort wird konkret ein Seminarinhalt oder eine Erfahrung genannt, die im Seminar gelernt wurden. In der Regel wurden daraus falsche Schlüsse gezogen und eine Antwort zum negativen verändert. Die Seminarinhalte können aus dem Theorieteil sein, z. B. <i>Cognitive Load Theory/ Multimedialernen (CLT/MML)</i>, Inputs zu <i>digitaler Messwerterfassung (dMWE; inkl. Smartphones und Videoanalyse)</i>, <i>Simulationen</i> oder <i>Erklärvideos</i>.</p> <p>Oder es handelt sich um <i>spezielle Erfahrungen aus dem praktischen Seminaranteil</i>. Es lässt sich nicht bei jeder veränderten Antwort ein Inhalt konkret identifizieren, sodass die Kategorie nicht immer vergeben wird.</p> | | | | |

Um zu kategorisieren, auf welche Weise und durch welche Ursache die Teilnehmenden falsche Schlüsse in ihrer Argumentation ziehen, wurden die folgenden Kategorien genutzt:

| Oberkategorie | Eigenschaft des Mediums in einem bestimmten Kontext verallgemeinert / übergeneralisiert | | |
|-------------------------------|---|--|---|
| Unterkategorien | Distraktor nicht kritisch genug bewertet | positive Erfahrung verallgemeinert zu typischem Vorteil | vorher skeptische Einstellung |
| Inhaltliche Beschreibung | <p>Bezieht sich auf veränderte Antworten, in denen im Posttest ein Distraktor (falsche Antwortoption) nicht mehr erkannt wurde und fälschlicherweise mit "ja" beantwortet wird.</p> <p>Veränderte Antwort lässt darauf schließen, dass diese Antwortoption (Distraktor) nicht ausreichend kritisch reflektiert wurde. Tritt häufig mit der Unterkategorie "positive Erfahrung verallgemeinert" auf.</p> | <p>Veränderte Antwort wird mit einer positiven Einzelerfahrung oder einem einzelnen Aspekt zu einem Medium begründet, welche/r übergeneralisiert und auch auf andere Bereiche oder Situationen verallgemeinert wird.</p> <p>Wird häufig mit anderen Unterkategorien ("<i>Distraktor nicht kritisch bewertet</i>" oder "<i>vorher skeptische Einstellung</i>") kodiert.</p> | <p>Begründung zur veränderten Antwort zeigt (u. a.), dass Proband/in vorher eine skeptische Einstellung gegenüber einem Medium hatte (welche sich nun verändert hat, z. B. aufgrund einer positiven Einzelerfahrung).</p> |
| Unterkategorien (Fortsetzung) | „negative“ Erfahrung verallgemeinert | unpassende Verallgemeinerung | |
| Inhaltliche Beschreibung | <p>Veränderte Antwort wird mit einer (eher) negativen Erfahrung / Situation begründet, welche verallgemeinert bzw. übergeneralisiert wurde. D. h. demzufolge wird i. d. R. ein Attraktor (richtige Antwortoption) im Posttest fälschlicherweise mit "Nein" angekreuzt.</p> | <p>Wird kodiert, wenn eine unpassende Verallgemeinerung vorliegt (z. B. Eigenschaft zu einem Medium auf Situation mit anderem Medium), die sich nicht in die beiden Unterkategorien positive Erfahrung bzw. negative Erfahrung verallgemeinert einordnen lässt (z. B., wenn der Bezug nicht ausreichend konkret ist).</p> | |

| Oberkategorie Fehlvverständnis zu einem Medium | | |
|---|--|--|
| Unterkategorien | Vermischung von Medien | Fehlvverständnis Medieneigen-schaft |
| Inhaltliche Beschreibung | In der Begründung wird deutlich, dass der/die Proband/in bestimmte Medien miteinander verwechselt bzw. vermischt oder einzelne Eigen-schaften zu einem Medium falsch wiedergibt bzw. anwendet. | |
| Oberkategorie Erster Kontakt mit Medium | | |
| Unterkategorien | Erwartungen (zu Vortei-len) nicht erfüllt | Fehlendes ‚technisches‘ Wissen |
| Inhaltliche Beschreibung | Veränderte Antwort zeigt, dass ein Vorteil zu einem Medium vor dem Seminar erwartet wurde; dann je-doch keine Erfahrungen diesbezüglich gemacht wurden, d. h. die Erwar-tungen nicht erfüllt wor-den sind. | Veränderte Antwort zeigt, dass vor dem Seminar (Prätest) kein Wissen darüber vorlag, was mit einem bestimmten Medium möglich ist (eher technischer Aspekt), im Seminar wurden die technischen Möglich-keiten dann kennengelernt. D. h. jedoch, dass eine verneinte Antwor-toption eher auf dieses fehlende techni-sche Wissen schließen lässt als auf eine fachdidaktisch begründete Ablehnung. |
| Oberkategorie Sonstiges | | |
| Unterkategorien | Fehlvverständnis Frage | Besseres Beispiel im Seminar als im Test |
| Inhaltliche Beschreibung | In der Begründung zur veränderten Antwort scheint es, als wäre ein Teil der Frage falsch verstan-den oder übersehen worden. | Einschätzung wurde verändert, weil das im Test präsentierte Me-dium aus Sicht des/der Proban-den/in kein so gutes Beispiel ist wie welche aus dem Seminar. |

H Hypothesenbildung zur Gestaltung von Lerngelegenheiten zum Einsatz digitaler Medien

Hypothesen / Folgerungen für die Seminargestaltung

| | Indizien aus | Beispiele aus Interviews |
|--|---|---|
| 1. fehlendes kritisches Reflektieren von Medienbeispielen/-einsätzen | | |
| digitale Medien werden als eine Art "Allheilmittel" (?) gesehen | z.B. digitales Medien "übernimmt" Aufgaben der Lehrkraft (Strukturierung des Lernprozesses, Übungsaufgaben anbieten etc.); | A21a02:72; A22a06:90+92; T22c04:100; T22c05: 94; |
| Reflexion über Grenzen zum Einsatz digitaler Medien fehlt | z.B. aufgrund positiver Einzelerfahrung weniger kritisch bei All-Aussagen | A22a08:42; G21b02:88; |
| nur positive Beispiele zu Medien/Medieneinsätzen erwecken den Eindruck, dass Einsatz digitaler Medien immer gut für den Unterricht ist | z.B. positive Beispiele führen zu weniger kritischer Haltung gegenüber Medieneinsätzen; digitalen Medien wird mehr zugesprochen | A21a02:72; A22a06:90; A22a06:92; G21b01:144 |
| Einzelerfahrungen werden unpassend verallgemeinert | z.B. Vorteil in einer Situation wird auf anderen Kontext übertragen (anderes didaktisches Ziel, anderes Medium, andere Lerngruppe etc.) | A22a04:88; A22a06:52 |
| | z.B. einmaliges Smartphone-Experiment gelingt nicht direkt | T22c04:110; T22c05:127; |

=> Folgerung 1: kritisches Reflektieren von Medienbeispielen/-einsätzen in der Lehrveranstaltung gezielt ermöglichen

Positivismus/über-Enthusiasmus vermeiden;

auch negativ-Beispiele einbinden und gemeinsam diskutieren

für angemessene Verallgemeinerung sensibilisieren (nicht aus 1x-Erfahrung, eher aus empirischen Befunden => anekdotische vs. empirische Evidenz)

Reflektion gemeinsam diskutieren

Fortsetzung auf nächster Seite

2. Praxiserfahrung und -erprobungen scheinen hilfreich (häufig mit Verbesserungen verbunden)

| | | |
|---|---|--|
| Ausprobieren der einzelnen Medien in kurzen Phasen | Vorteile und Möglichkeiten werden kennengelernt oder verdeutlicht | G22b04:117; T22c04:84; A22a04:58; G21b02:82 |
| Implementation digitaler Medien in Unterrichtssequenzen (Vorbereitung und Durchführung) | hilft zum realistischeren Einschätzen, entgegen Positivismus; Anforderungen und Grenzen werden erkannt | A21a02:62; A22a04:45; A22a04:55; G22b03:76 |
| praktische Erfahrungen aus Lernendensicht (weniger häufig als Lehrerperspektive) | Schüler-Rolle kann auch helfen | G21b01:121; T22c04:92 |
| ABER: angemessene Reflexion wichtig | Praxiserfahrungen hängen auch mit Verschlechterungen zusammen; z.T. fehlt kritische Reflexion der Erfahrung | T22c04:91; T22c05:94 |

=> Folgerung 2: eigenes Ausprobieren und praktische Implementierung der Medien ermöglichen (in Verbindung mit anschließender Reflexion)

3. mangelnde Verknüpfung des Wissens zu Medien und z.T. Vermischung

| | | |
|---|--|---|
| Verständnis zu digitalen Medien (aus dem Seminar) | Vermischung einzelner Medien | A21a02:68; G21b02:95; T22c02:92; T22c04:91 |
| | Versuche, Wissen zu verknüpfen/Wissen zu einem Medium auf andere zu übertragen, gelingen nicht | A22a04:88-90; A22a08:54 |
| | Aufgaben, die Vergleiche/übergreifendes Wissen fordern, fallen schlechter aus (A.11, A.7, A.3) | |

=> Folgerung 3: Entwicklung übergreifenden Wissens zu Medien fördern ohne individuelle Gestaltungsmerkmale der einzelnen Medien zu vernachlässigen

Fortsetzung auf nächster Seite

4. Vorerfahrungen und -einstellungen der Seminarteilnehmenden beeinflussen die Wahrnehmung und Reflexion der Medienbeispiele im Seminar

| | | |
|--|--|--|
| unterschiedliche Typen hinsichtlich der Einstellung zum Einsatz digitaler Medien oder zu bestimmten Medien | Skeptiker aufgrund eher negativen Erfahrungen in eigener Schulzeit => Bild positiv gewandelt | z.B. A22a04:51+58 , A22a06:58+90 +92; G21b02:88 |
| | unreflektierte, positive Einstellung vorher => kritischere Sichtweise nachher | z.B. T22c05:96+12 6; A21b03:57 |

=> Folgerung 4: Vorerfahrungen und -einstellungen der Seminarteilnehmenden adressieren

Vorerfahrungen im Seminar diskutieren und bewusst machen

Gelegenheit geben, Erfahrungen im Seminar mit Bezug zu Vorerfahrungen zu reflektieren

Reflexion gemeinsam diskutieren, um extrem-Ansichten zu reduzieren

Bisher erschienene Bände der Reihe
Studien zum Physik- und Chemielernen

ISSN 1614-8967

Vollständige Übersicht auf unserer Website



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

Aktuelle Bände

- 310 Wolfgang Becker (2021): Auswirkungen unterschiedlicher experimenteller Repräsentationen auf den Kenntnisstand bei Grundschulkindern
ISBN 978-3-8325-5255-8 50.00 EUR
- 311 Marvin Rost (2021): Modelle als Mittel der Erkenntnisgewinnung im Chemieunterricht der Sekundarstufe I. Entwicklung und quantitative Dimensionalitätsanalyse eines Testinstruments aus epistemologischer Perspektive
ISBN 978-3-8325-5256-5 44.00 EUR (open access)
- 312 Christina Kobl (2021): Förderung und Erfassung der Reflexionskompetenz im Fach Chemie
ISBN 978-3-8325-5259-6 41.00 EUR (open access)
- 313 Ann-Kathrin Beretz (2021): Diagnostische Prozesse von Studierenden des Lehramts. eine Videostudie in den Fächern Physik und Mathematik
ISBN 978-3-8325-5288-6 45.00 EUR (open access)
- 314 Judith Breuer (2021): Implementierung fachdidaktischer Innovationen durch das Angebot materialgestützter Unterrichtskonzeptionen. Fallanalysen zum Nutzungsverhalten von Lehrkräften am Beispiel des Münchener Lehrgangs zur Quantenmechanik
ISBN 978-3-8325-5293-0 50.50 EUR (open access)
- 315 Michaela Oettle (2021): Modellierung des Fachwissens von Lehrkräften in der Teilchenphysik. Eine Delphi-Studie
ISBN 978-3-8325-5305-0 57.50 EUR (open access)
- 316 Volker Brüggemann (2021): Entwicklung und Pilotierung eines adaptiven Multistage-Tests zur Kompetenzerfassung im Bereich naturwissenschaftlichen Denkens
ISBN 978-3-8325-5331-9 40.00 EUR (open access)
- 317 Stefan Müller (2021): Die Vorläufigkeit und soziokulturelle Eingebundenheit naturwissenschaftlicher Erkenntnisse. Kritische Reflexion, empirische Befunde und fachdidaktische Konsequenzen für die Chemielehrer*innenbildung
ISBN 978-3-8325-5343-2 63.00 EUR

- 318 Laurence Müller (2021): Alltagsentscheidungen für den Chemieunterricht erkennen und Entscheidungsprozesse explorativ begleiten
ISBN 978-3-8325-5379-1 59.00 EUR
- 319 Lars Ehlert (2021): Entwicklung und Evaluation einer Lehrkräftefortbildung zur Planung von selbstgesteuerten Experimenten
ISBN 978-3-8325-5393-7 41.50 EUR (open access)
- 320 Florian Seiler (2021): Entwicklung und Evaluation eines Seminarkonzepts zur Förderung der experimentellen Planungskompetenz von Lehramtsstudierenden im Fach Chemie
ISBN 978-3-8325-5397-5 47.50 EUR (open access)
- 321 Nadine Boele (2021): Entwicklung eines Messinstruments zur Erfassung der professionellen Unterrichtswahrnehmung von (angehenden) Chemielehrkräften hinsichtlich der Lernunterstützung
ISBN 978-3-8325-5402-6 46.50 EUR
- 322 Franziska Zimmermann (2022): Entwicklung und Evaluation digitalisierungsbezogener Kompetenzen von angehenden Chemielehrkräften
ISBN 978-3-8325-5410-1 49.50 EUR
- 323 Lars-Frederik Weiß (2021): Der Flipped Classroom in der Physik-Lehre. Empirische Untersuchungen in Schule und Hochschule
ISBN 978-3-8325-5418-7 51.00 EUR
- 324 Tilmann Steinmetz (2021): Kumulatives Lehren und Lernen im Lehramtsstudium Physik. Theorie und Evaluation eines Lehrkonzepts
ISBN 978-3-8325-5421-7 51.50 EUR
- 325 Kübra Nur Celik (2022): Entwicklung von chemischem Fachwissen in der Sekundarstufe I. Validierung einer Learning Progression für die Basiskonzepte „Struktur der Materie“, „Chemische Reaktion“ und „Energie“ im Kompetenzbereich „Umgang mit Fachwissen“
ISBN 978-3-8325-5431-6 55.00 EUR
- 326 Matthias Ungermann (2022): Förderung des Verständnisses von Nature of Science und der experimentellen Kompetenz im Schüler*innen-Labor Physik in Abgrenzung zum Regelunterricht
ISBN 978-3-8325-5442-2 55.50 EUR
- 327 Christoph Hoyer (2022): Multimedial unterstütztes Experimentieren im webbasierten Labor zur Messung, Visualisierung und Analyse des Feldes eines Permanentmagneten
ISBN 978-3-8325-5453-8 45.00 EUR
- 328 Tobias Schüttler (2022): Schülerlabore als interessefördernde authentische Lernorte für den naturwissenschaftlichen Unterricht nutzen
ISBN 978-3-8325-5454-5 50.50 EUR
- 329 Christopher Kurth (2022): Die Kompetenz von Studierenden, Schülerschwierigkeiten beim eigenständigen Experimentieren zu diagnostizieren
ISBN 978-3-8325-5457-6 58.50 EUR

- 330 Dagmar Michna (2022): Inklusiver Anfangsunterricht Chemie. Entwicklung und Evaluation einer Unterrichtseinheit zur Einführung der chemischen Reaktion
ISBN 978-3-8325-5463-7 49.50 EUR
- 331 Marco Seiter (2022): Die Bedeutung der Elementarisierung für den Erfolg von Mechanikunterricht in der Sekundarstufe I
ISBN 978-3-8325-5471-2 66.00 EUR
- 332 Jörn Hägele (2022): Kompetenzaufbau zum experimentbezogenen Denken und Arbeiten. Videobasierte Analysen zu Aktivitäten und Vorstellungen von Schülerinnen und Schülern der gymnasialen Oberstufe bei der Bearbeitung von fachmethodischer Instruktion
ISBN 978-3-8325-5476-7 56.50 EUR (open access)
- 333 Erik Heine (2022): Wissenschaftliche Kontroversen im Physikunterricht. Explorationsstudie zum Umgang von Physiklehrkräften und Physiklehramtsstudierenden mit einer wissenschaftlichen Kontroverse am Beispiel der Masse in der Speziellen Relativitätstheorie
ISBN 978-3-8325-5478-1 48.50 EUR (open access)
- 334 Simon Goertz (2022): Module und Lernzirkel der Plattform FLexKom zur Förderung experimenteller Kompetenzen in der Schulpraxis. Verlauf und Ergebnisse einer Design-Based Research Studie
ISBN 978-3-8325-5494-1 66.50 EUR
- 335 Christina Toschka (2022): Lernen mit Modellexperimenten. Empirische Untersuchung der Wahrnehmung und des Denkens in Analogien beim Umgang mit Modellexperimenten
ISBN 978-3-8325-5495-8 50.00 EUR (open access)
- 336 Alina Behrendt (2022): Chemiebezogene Kompetenzen in der Übergangsphase zwischen dem Sachunterricht der Primarstufe und dem Chemieunterricht der Sekundarstufe I
ISBN 978-3-8325-5498-9 40.50 EUR (open access)
- 337 Manuel Daiber (2022): Entwicklung eines Lehrkonzepts für eine elementare Quantenmechanik. Formuliert mit In-Out Symbolen
ISBN 978-3-8325-5507-8 48.50 EUR
- 338 Felix Pawlak (2022): Das Gemeinsame Experimentieren (an-)leiten. Eine qualitative Studie zum chemiespezifischen Classroom-Management
ISBN 978-3-8325-5508-5 46.50 EUR
- 339 Liza Dopatka (2022): Konzeption und Evaluation eines kontextstrukturierten Unterrichtskonzeptes für den Anfangs-Elektrizitätslehreunterricht
ISBN 978-3-8325-5514-6 69.50 EUR
- 340 Arne Bewersdorff (2022): Untersuchung der Effektivität zweier Fortbildungsformate zum Experimentieren mit dem Fokus auf das Unterrichtshandeln
ISBN 978-3-8325-5522-1 39.00 EUR (open access)

- 341 Thomas Christoph Münster (2022): Wie diagnostizieren Studierende des Lehramtes physikbezogene Lernprozesse von Schüler*innen?. Eine Videostudie zur Mechanik
ISBN 978-3-8325-5534-4 44.50 EUR (open access)
- 342 Ines Komor (2022): Förderung des symbolisch-mathematischen Modellverständnisses in der Physikalischen Chemie
ISBN 978-3-8325-5546-7 46.50 EUR
- 343 Verena Petermann (2022): Überzeugungen von Lehrkräften zum Lehren und Lernen von Fachinhalten und Fachmethoden und deren Beziehung zu unterrichtsnahem Handeln
ISBN 978-3-8325-5545-0 47.00 EUR (open access)
- 344 Jana Heinze (2022): Einfluss der sprachlichen Konzeption auf die Einschätzung der Qualität instruktionaler Unterrichtserklärungen im Fach Physik
ISBN 978-3-8325-5553-5 42.00 EUR (open access)
- 345 Jannis Weber (2022): Mathematische Modellbildung und Videoanalyse zum Lernen der Newtonschen Dynamik im Vergleich
ISBN 978-3-8325-5566-5 68.00 EUR (open access)
- 346 Fabian Sterzing (2022): Zur Lernwirksamkeit von Erklärvideos in der Physik. Eine Untersuchung in Abhängigkeit von ihrer fachdidaktischen Qualität und ihrem Einbettungsformat
ISBN 978-3-8325-5576-4 52.00 EUR (open access)
- 347 Lars Greitemann (2022): Wirkung des Tablet-Einsatzes im Chemieunterricht der Sekundarstufe I unter besonderer Berücksichtigung von Wissensvermittlung und Wissenssicherung
ISBN 978-3-8325-5580-1 50.00 EUR
- 348 Fabian Poensgen (2022): Diagnose experimenteller Kompetenzen in der laborpraktischen Chemielehrer*innenbildung
ISBN 978-3-8325-5587-0 48.00 EUR
- 349 William Lindlahr (2023): Virtual-Reality-Experimente. Entwicklung und Evaluation eines Konzepts für den forschend-entwickelnden Physikunterricht mit digitalen Medien
ISBN 978-3-8325-5595-5 49.00 EUR
- 350 Bert Schlüter (2023): Teilnahmemotivation und situationales Interesse von Kindern und Eltern im experimentellen Lernsetting KEMIE
ISBN 978-3-8325-5598-6 43.00 EUR
- 351 Katharina Nave (2023): Charakterisierung situativer mentaler Modellkomponenten in der Chemie und die Bildung von Hypothesen. Eine qualitative Studie zur Operationalisierung mentaler Modellkomponenten für den Fachbereich Chemie
ISBN 978-3-8325-5599-3 43.00 EUR
- 352 Anna B. Bauer (2023): Experimentelle Kompetenz Physikstudierender. Entwicklung und erste Erprobung eines performanzorientierten Kompetenzstrukturmodells unter Nutzung qualitativer Methoden
ISBN 978-3-8325-5625-9 47.00 EUR (open access)

- 353 Jan Schröder (2023): Entwicklung eines Performanztests zur Messung der Fähigkeit zur Unterrichtsplanung bei Lehramtsstudierenden im Fach Physik
ISBN 978-3-8325-5655-6 46.50 EUR
- 354 Susanne Gerlach (2023): Aspekte einer Fachdidaktik Körperpflege. Ein Beitrag zur Standardentwicklung
ISBN 978-3-8325-5659-4 45.00 EUR
- 355 Livia Murer (2023): Diagnose experimenteller Kompetenzen beim praktisch-naturwissenschaftlichen Arbeiten. Vergleich verschiedener Methoden und kognitive Validierung eines Testverfahrens
ISBN 978-3-8325-5657-0 41.50 EUR (open access)
- 356 Andrea Maria Schmid (2023): Authentische Kontexte für MINT-Lernumgebungen. Eine zweiteilige Interventionsstudie in den Fachdidaktiken Physik und Technik
ISBN 978-3-8325-5605-1 57.00 EUR (open access)
- 357 Julia Ortmann (2023): Bedeutung und Förderung von Kompetenzen zum naturwissenschaftlichen Denken und Arbeiten in universitären Praktika
ISBN 978-3-8325-5670-9 37.00 EUR (open access)
- 358 Axel-Thilo Prokop (2023): Entwicklung eines Lehr-Lern-Labors zum Thema Radioaktivität. Eine didaktische Rekonstruktion
ISBN 978-3-8325-5671-6 49.50 EUR
- 359 Timo Hackemann (2023): Textverständlichkeit sprachlich variiertes physikbezogener Sachtexte
ISBN 978-3-8325-5675-4 41.50 EUR (open access)
- 360 Dennis Dietz (2023): Vernetztes Lernen im fächerdifferenzierten und integrierten naturwissenschaftlichen Unterricht aufgezeigt am Basiskonzept Energie. Eine Studie zur Analyse der Wirksamkeit der Konzeption und Implementation eines schulinternen Curriculums für das Unterrichtsfach „Integrierte Naturwissenschaften 7/8“
ISBN 978-3-8325-5676-1 49.50 EUR
- 361 Ann-Katrin Krebs (2023): Vielfalt im Physikunterricht. Zur Wirkung von Lehrkräftefortbildungen unter Diversitätsaspekten
ISBN 978-3-8325-5672-3 65.50 EUR (open access)
- 362 Simon Kaulhausen (2023): Strukturelle Ursachen für Klausurmisserfolg in Allgemeiner Chemie an der Universität
ISBN 978-3-8325-5699-0 37.50 EUR (open access)
- 363 Julia Eckoldt (2023): Den (Sach-)Unterricht öffnen. Selbstkompetenzen und motivationale Orientierungen von Lehrkräften bei der Implementation einer Innovation untersucht am Beispiel des Freien Explorierens und Experimentierens
ISBN 978-3-8325-5663-1 48.50 EUR (open access)

- 364 Albert Teichrow (2023): Physikalische Modellbildung mit dynamischen Modellen
ISBN 978-3-8325-5710-2 58.50 EUR (open access)
- 365 Sascha Neff (2023): Transfer digitaler Innovationen in die Schulpraxis. Eine explorative Untersuchung zur Förderung der Implementation
ISBN 978-3-8325-5687-7 59.00 EUR (open access)
- 366 Rahel Schmid (2023): Verständnis von Nature of Science-Aspekten und Umgang mit Fehlern von Schüler*innen der Sekundarstufe I. Am Beispiel von digital-basierten Lernprozessen im informellen Lernsetting Smartfeld
ISBN 978-3-8325-5722-5 53.50 EUR (open access)
- 367 Dennis Kirstein (2023): Individuelle Bedingungs- und Risikofaktoren für erfolgreiche Lernprozesse mit kooperativen Experimentieraufgaben im Chemieunterricht. Eine Untersuchung zum Zusammenhang von Lernvoraussetzungen, Lerntätigkeiten, Schwierigkeiten und Lernerfolg beim Experimentieren in Kleingruppen der Sekundarstufe I
ISBN 978-3-8325-5729-4 50.50 EUR (open access)
- 368 Frauke Düwel (2024): Argumentationslinien in Lehr-Lernkontexten. Potenziale englischer Fachtexte zur Chromatografie und deren hochschuldidaktische Einbindung
ISBN 978-3-8325-5731-7 63.00 EUR (open access)
- 369 Fabien Güth (2023): Interessenbasierte Differenzierung mithilfe systematisch variiertes Kontextaufgaben im Fach Chemie
ISBN 978-3-8325-5737-9 48.00 EUR (open access)
- 370 Oliver Grewe (2023): Förderung der professionellen Unterrichtswahrnehmung und Selbstwirksamkeitsüberzeugungen hinsichtlich sprachsensibler Maßnahmen im naturwissenschaftlichen Sachunterricht. Konzeption und Evaluation einer video- und praxisbasierten Lehrveranstaltung im Masterstudium
ISBN 978-3-8325-5738-6 44.50 EUR (open access)
- 371 Anna Nowak (2023): Untersuchung der Qualität von Selbstreflexionstexten zum Physikunterricht. Entwicklung des Reflexionsmodells REIZ
ISBN 978-3-8325-5739-3 59.00 EUR (open access)
- 372 Dominique Angela Holland (2023): Bildung für nachhaltige Entwicklung (BNE) kooperativ gestalten. Vergleich monodisziplinärer und interdisziplinärer Kooperation von Lehramtsstudierenden bei der Planung, Durchführung und Reflexion von Online-BNE-Unterricht
ISBN 978-3-8325-5760-7 47.00 EUR (open access)
- 373 Renan Marcello Vairo Nunes (2024): MINT-Personal an Schulen. Eine Untersuchung der Arbeitssituation und professionellen Kompetenzen von MINT-Lehrkräften verschiedener Ausbildungswege
ISBN 978-3-8325-5778-2 51.00 EUR (open access)
- 374 Mats Kieserling (2024): Digitalisierung im Chemieunterricht. Entwicklung und Evaluation einer experimentellen digitalen Lernumgebung mit universeller Zugänglichkeit
ISBN 978-3-8325-5786-7 45.50 EUR

- 375 Cem Aydin Salim (2024): Die Untersuchung adaptiver Lernsettings im Themenbereich „Schwimmen und Sinken“ im naturwissenschaftlichen Unterricht
ISBN 978-3-8325-5787-4 49.00 EUR (open access)
- 376 Novid Ghassemi (2024): Evaluation eines Lehramtsmasterstudiengangs mit dem Profil Quereinstieg im Fach Physik. Erkenntnisse zu Eingangsbedingungen, professionellen Kompetenzen und Aspekten individueller Angebotsnutzung
ISBN 978-3-8325-5789-8 41.50 EUR (open access)
- 377 Martina Flurina Cavelti (2024): Entwicklung und Validierung eines Messinstruments zur Erfassung der Schülerkompetenzen im Bereich des wissenschaftlichen Skizzierens im Fach Chemie in der Sekundarstufe I
ISBN 978-3-8325-5829-1 45.00 EUR (open access)
- 378 Tom Bleckmann (2024): Formatives Assessment auf Basis von maschinellem Lernen. Eine Studie über automatisiertes Feedback zu Concept Maps aus dem Bereich Mechanik
ISBN 978-3-8325-5842-0 46.50 EUR (open access)
- 379 Jana Marlies Rehberg (2024): Das physikspezifische Mindset zum Studienbeginn. Fragebogenentwicklung und Aufbau einer Online-Intervention
ISBN 978-3-8325-5850-5 59.50 EUR (open access)
- 380 Florian Trauten (2024): Entwicklung und Evaluation von automatisierten Feedbackschleifen in Online-Aufgaben im Fach Chemie
ISBN 978-3-8325-5859-8 46.00 EUR (open access)
- 381 Johanna Dejanovikj (2024): Reflexion als Lerngegenstand im Unterricht zur Förderung von Bewertungskompetenz
ISBN 978-3-8325-5860-4 41.00 EUR (open access)
- 382 Katharina Fliester (2024): Verständlichkeit physikalischer Sachtexte. Untersuchungen zum Wirkungsgefüge zwischen sprachlicher Textgestaltung und der Behaltensleistung sowie der Textwahrnehmung im Schulfach Physik
ISBN 978-3-8325-5858-1 44.00 EUR (open access)
- 383 Stephanie Neppl (2024): Perspektivenübernahme im Physikunterricht. Explorative Interviewstudie zu einer Seminarkonzeption mit dem Schwerpunkt Perspektivenübernahme bei der Planung von Physikunterricht
ISBN 978-3-8325-5865-9 48.00 EUR (open access)
- 384 Katja Plicht (2024): Ein Physikübungskonzept zur Förderung der Problemlösekompetenz. Entwicklung und empirische Evaluation eines Strategietrainings auf der Basis von Expertisemerkmalen
ISBN 978-3-8325-5875-8 45.00 EUR (open access)
- 385 Svenja Boegel (2024): Feedback beim Experimentieren: Zur Rolle von Cognitive Load und Motivation
ISBN 978-3-8325-5911-3 45.50 EUR (open access)

- 386 Laura Pannullo (2025): Wahlmöglichkeiten beim Experimentieren: Entwicklung und Erprobung eines Konzeptes für Experimente in inklusiven Lerngruppen im Physikunterricht
ISBN 978-3-8325-5916-8 46.50 EUR (open access)
- 387 Matthias Schweinberger (2025): „Stumme Videos“. Aufmerksamkeitslenkende Moderation von Demonstrationsexperimenten
ISBN 978-3-8325-5930-4 50.50 EUR
- 388 Rike Große-Heilmann (2025): Entwicklung fachdidaktischen Wissens zum Einsatz digitaler Medien im Fach Physik
ISBN 978-3-8325-5943-4 51.00 EUR (open access)

Vollständige Übersicht unter: <https://www.logos-verlag.de/spcl>

Alle erschienenen Bücher können unter der angegebenen ISBN direkt online (<http://www.logos-verlag.de>) oder telefonisch (030 - 42 85 10 90) beim Logos Verlag Berlin bestellt werden.

Studien zum Physik- und Chemielernen

Herausgegeben von Martin Hopf und Mathias Ropohl

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 im deutschsprachigen Raum.

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:

Univ.-Prof. Dr. Martin Hopf
Universität Wien,
Österreichisches Kompetenzzentrum
für Didaktik der Physik,
Porzellangasse 4, Stiege 2,
1090 Wien, Österreich,
Tel. +43-1-4277-60330,
e-mail: martin.hopf@univie.ac.at

Prof. Dr. Mathias Ropohl
Didaktik der Chemie,
Fakultät für Chemie,
Universität Duisburg-Essen,
Schützenbahn 70, 45127 Essen,
Tel. 0201-183 2704,
e-mail: mathias.ropohl@uni-due.de

Angesichts der Potentiale, die der Einsatz digitaler Medien im Physikunterricht bietet, bedarf es entsprechender fachdidaktischer Lerngelegenheiten in der Lehrkräftebildung. Die begleitende Untersuchung des Erwerbs digitaler Kompetenzen wie dem fachdidaktischen Wissen (FDW) zum Einsatz digitaler Medien erfolgt üblicherweise mittels Selbsteinschätzungen, welche jedoch hinsichtlich ihrer mangelnden Validität und Fachspezifität kritisiert werden.

In der vorliegenden Arbeit wird daher die Entwicklung eines fachdidaktischen Leistungstests zur validen Erfassung des FDW zum Einsatz digitaler Medien fokussiert. Der Leistungstest wird zur Evaluation fachdidaktischer Seminare im Prä-Post-Design bei Physiklehramtsstudierenden eingesetzt ($N = 70$). Durch anknüpfende retrospektive Interviews mit einem Teil der Studierenden wird analysiert, welche Seminarelemente mit Veränderungen in den Testantworten zusammenhängen, um lernförderliche oder lernhinderliche Elemente für den Wissenserwerb zu identifizieren.

Die Ergebnisse der Arbeit zeigen einen Wissenszuwachs im FDW mit kleinem Effekt. Aus der Identifikation lernförderlicher Seminarelemente in der Interviewanalyse können wichtige Gestaltungsmerkmale für zukünftige Lerngelegenheiten zum Einsatz digitaler Medien abgeleitet werden, wie beispielsweise eine kritische Reflexion von Medieneinsätzen und -beispielen sowie die Möglichkeit des eigenen Ausprobierens und Implementierens der Medien.

Logos Verlag Berlin

ISBN 978-3-8325-5943-4