

Christoph Vogelsang/Andreas Borowski/David Buschhüter/
 Patrick Enkrott/Maren Kempin/Christoph Kulgemeyer/Peter Reinhold/
 Josef Riese/Horst Schecker/Jan Schröder

Entwicklung von Professionswissen und Unterrichtsperformanz im Lehramtsstudium Physik

Analysen zu valider Testwertinterpretation¹

Zusammenfassung: Angehende Physiklehrkräfte sollen im Rahmen ihres Studiums fachliches und fachdidaktisches Wissen erwerben, welches die Gestaltung lernförderlichen Unterrichts ermöglicht. Es ist allerdings empirisch nur wenig geklärt, wie sich dieses Wissen im Laufe des Studiums entwickelt und ob es zur Ausbildung von Handlungsfähigkeiten beiträgt. Um derartige Wirkungsaussagen treffen zu können, müssen Instrumente entwickelt werden, die eine valide Testwertinterpretation zulassen. In diesem Beitrag werden auf Basis von im Projekt Profile-P+ entwickelten Instrumenten Validitätsanalysen zur längsschnittlichen Entwicklung des Professionswissens von Physiklehramtsstudierenden im Verlauf des Bachelorstudiums und ihrer Fähigkeiten zur Planung und Reflexion von Physikunterricht sowie zum Erklären von physikalischen Sachverhalten vor und nach dem Praxissemester dargestellt. Neben Wissenstests kamen standardisierte Performanztests zum Einsatz. Die vorliegenden Ergebnisse sprechen dafür, dass die erhobenen Messwerte im Sinne von Wirkungsaussagen interpretiert werden können.

Schlagworte: Physik, Lehrerbildung, Professionswissen, Performanztest, Längsschnitt

1. Einleitung

1.1 Problemstellung

In einem Lehramtsstudium sollen angehende Lehrkräfte Kompetenzen erwerben, die eine notwendige Grundlage für erfolgreiches Handeln im Unterricht bilden (vgl. Terhart, 2012). Ausführliche Lerngelegenheiten zum unterrichtlichen Handeln sind zwar strukturell erst im Vorbereitungsdienst verortet, aber durch die Einführung von Praxissemestern wurden in vielen Bundesländern praktische Ausbildungsanteile auch verstärkt in das Studium integriert. Es ist jedoch unklar, ob und wie sich handlungsnahe Fähigkeiten im Verlauf solcher Praxisphasen entwickeln. Vorliegende Wirksamkeitsanalysen der Lehrerbildung untersuchen Kompetenzen von Studierenden meist in Querschnittsanalysen. Um ein detailliertes Bild zu gewinnen, ist es notwendig, bereits die universitären

¹ Das Projekt Profile-P+ wird gefördert im Rahmen des BMBF-Schwerpunktprogramms KoKoHs (FKZ: 01PK15005A-D).

Glieder einer vereinfachten Professionalisierungskette im Längsschnitt zu betrachten: *Bereitstellung von Lerngelegenheiten im Studium* → *Wissenserwerb im Studium* → *Erprobung und Reflexion in studentischen Praxisphasen* → *Erwerb und Reflexion erster Handlungsrouniten im Vorbereitungsdienst* → *Ausbau und Revision unterrichtlicher Handlungskompetenz in der Berufseingangsphase*. Hieraus ergibt sich die Frage: Inwiefern entwickeln sich Aspekte professioneller Kompetenz und unterrichtsnaher Fähigkeiten im Studienverlauf? Um sie untersuchen zu können, müssen Argumente gesammelt werden, die eine valide Interpretation von erhobenen Testwerten im Sinne dieser Frage stützen.

1.2 Ziele

Das Projekt *Profile-P+* soll einen Beitrag zur Beantwortung dieser Frage liefern. Zum einen wird die Entwicklung des Professionswissens von angehenden Physiklehrkräften längsschnittlich im Verlauf der ersten beiden Studienjahre des Bachelorstudiums untersucht. Zum anderen wird die Entwicklung des Wissens sowie der Fähigkeiten zur Bewältigung dreier Standardanforderungen für Physiklehrkräfte (*Unterrichtsplanung, Unterrichtsreflexion, Erklären von Physik*) im Verlauf eines Praxissemesters analysiert. In diesem Beitrag werden der theoretische Rahmen, das Studiendesign, die verwendeten Instrumente sowie erste Ergebnisse erläutert. Außerdem werden Ergebnisse berichtet, die die Annahme stützen, hinsichtlich der oben genannten Frage valide Testwertinterpretationen vornehmen zu können.

2. Theoretischer Hintergrund

2.1 Professionelle Kompetenz von Lehrkräften

Blömeke, Gustafson und Shavelson (2015) beschreiben Kompetenz als Kontinuum, das kognitive und affektiv-motivationale Dispositionen als Basis für situationsspezifische Fähigkeiten betrachtet, die wiederum Performanz in komplexen unterrichtlichen Handlungssituationen ermöglichen. Einen zentralen Bestandteil dieser Dispositionen bildet das professionelle Wissen, für das sich in der an Shulman (1986) anschließenden Forschung die Unterteilung in die drei Dimensionen fachliches, fachdidaktisches Wissen und pädagogisches Wissen etabliert hat (vgl. Baumert & Kunter, 2006).

Fachliches Wissen (FW) umfasst Wissen über die Inhalte und Methoden des zu unterrichtenden Faches sowie epistemologische Aspekte. Für die Domäne der Physik liegen unterschiedliche Modellierungen des Fachwissens vor (vgl. Woitkowski & Borowski, 2017), die meist eine Unterscheidung der ‚Tiefe‘ des Wissens nach curricularen Stufen (Schulwissen, universitäres Wissen) enthalten. Wir erweitern diese Struktur um eine spezifische Form des Fachwissens, von der angenommen wird, dass sie insbesondere für Lehrkräfte eine notwendige Basis für das Unterrichten ist: sogenanntes

vertieftes Schulwissen (vgl. Woehlecke et al., 2017). *Fachdidaktisches Wissen (FDW)* bezeichnet Wissen, das Lehrkräfte benötigen, um fachliche Inhalte adressatengerecht aufzubereiten und zu strukturieren. Hierfür wurden sowohl national, als auch international eine Vielzahl an Modellierungen vorgeschlagen, die bspw. aktuell im *Refined Consensus Model of PCK (pedagogical content knowledge)* zusammengeführt wurden (vgl. Hume, Cooper & Borowski, 2019). Konkretisierungen für die Domäne Physik beinhalten dabei meist Wissen über Schülerkonzeptionen und -vorstellungen, über Instruktionsstrategien und zum Umgang mit Experimenten. *Pädagogisches Wissen (PW)* beinhaltet fachunabhängiges Wissen über allgemeindidaktische Prinzipien und pädagogische Konzepte. Auch zu diesem Wissensbereich existieren verschiedene Modelle (vgl. Baumert & Kunter, 2006), die aber alle meist unterrichtsbezogenes Wissen zur Klassenführung oder zur Motivierung beinhalten.

2.2 Professionswissen und Lehrerhandeln

Unterrichten stellt eine zentrale Handlungsanforderung an Lehrkräfte dar. Dabei wird angenommen, dass Lehrkräfte für die Gestaltung lernwirksamen Unterrichts über adäquates professionelles Wissen verfügen müssen (vgl. Baumert & Kunter, 2006). Für den Physikunterricht konnten allerdings bisher sowohl national als auch international wenig Zusammenhänge zwischen physikbezogenem Professionswissen und der Unterrichtsqualität oder dem Lernzuwachs von Schülerinnen und Schülern beobachtet werden (vgl. Vogelsang & Cautet, 2017; z. B. Ohle, Boone & Fischer, 2014). Dieses Ergebnis ist unabhängig davon, auf welche Art und Weise Unterrichtsqualität operationalisiert wird, z. B. als hochinferentes Rating (vgl. Korneck, Krüger & Szogs, 2017) oder durch z. B. Analyse der inhaltlichen Sachstruktur (vgl. Liepertz & Borowski, 2019).

Dies kann ein Indiz für einen fehlenden Einfluss des erhobenen Wissens auf die Qualität des Unterrichtshandelns sein. Ein anderer Grund könnte aber auch sein, dass realer Unterricht sehr viele heterogene Kontextfaktoren enthält, die vorhandene Zusammenhänge verdecken (vgl. Kulgemeyer & Riese, 2018). Daher sollte die Handlungsqualität kontrollierter erfasst werden. Hierzu kann man sich an Konzepte aus der Ausbildung und Prüfung von Medizinerinnen und Medizinern anlehnen (vgl. Miller, 1990). Bei *Objective Structured Clinical Examinations (OSCE)* werden typische Standardanforderungen des jeweiligen Berufs simuliert, häufig mit Hilfe von Schauspielerinnen und Schauspielern. Es handelt sich also um Situationen, die realen Handlungsanforderungen möglichst nahekommen, aber zugleich so standardisiert sind, dass eine hohe Vergleichbarkeit gewährleistet ist. Mit *scoring sheets* lassen sich hohe Testgütestandards erreichen (vgl. Walters, Osborn & Raven, 2005). Ein solcher Performanztest zum *Erklären von Physik* wurde in Form einer Rollenspielsituation bereits im Vorgängerprojekt *Profile-P* entwickelt und eingesetzt. Dabei zeigten sich die vermuteten Zusammenhänge zwischen Erklärfähigkeit und professionellem Wissen (Kulgemeyer & Riese, 2018).

2.3 Entwicklung von Professionswissen und Handlungsqualität im Lehramtsstudium

Ergebnisse zur Entwicklung professionellen Wissens über das Physiklehramtsstudium hinweg liegen aus bisherigen Studien nur wenig vor. Querschnittserhebungen zeigen, dass Studierende höherer Semester ein höheres FW und FDW aufweisen als Studierende unterer Semester (vgl. z. B. Riese & Reinhold, 2012). Sorge, Keller, Petersen und Neumann (2018) konnten eine Zunahme von FW und FDW auch längsschnittlich im Verlauf eines Studienjahres zeigen. Es fehlen derzeit jedoch Längsschnittstudien, in denen die Entwicklung von FW, FDW und PW über längere Studienverläufe gemeinsam untersucht wird, ebenso wie Untersuchungen zu damit in Verbindung stehenden Handlungsqualitäten, wie z. B. bei der Unterrichtsplanung und -reflexion. Solche Längsschnitte über den Studienverlauf stellen eine hohe Anforderung an die verwendeten Instrumente dar: es müssen z. B. Boden- und Deckeneffekte für Gruppen unterschiedlicher Expertise vermieden werden, ebenfalls kann das mehrfache Ausfüllen von Testinstrumenten Auswirkungen auf Testmotivation oder *test-taking-effort* haben, die die Messung stören können (vgl. z. B. Liu, Bridgeman & Adler, 2012).

3. Forschungsfragen

Ausgehend von den oben beschriebenen Forschungsdesiderata wird in diesem Beitrag die übergreifende Frage behandelt:

Können Veränderungen von Testwerten zu professionellem Wissen und zu Fähigkeiten in den genannten Bereichen valide auf zunehmenden Studienfortschritt in der Lehrerbildung bezogen werden?

Es sollen daher Argumente diskutiert werden, inwiefern vorgenommene Messungen und die daraus abgeleiteten Testwerte valide Schlussfolgerungen hinsichtlich zweier Teilforschungsfragen zulassen.

- Wie verändert sich professionelles Wissen im Verlauf eines Lehramtsstudiums im Fach Physik in den ersten beiden Studienjahren sowie im Verlauf eines Praxissemesters?
- Wie entwickeln sich die Fähigkeiten zur Unterrichtsplanung, zum Erklären von Physik sowie zur Unterrichtsreflexion im Verlauf eines Praxissemesters im Fach Physik?

4. Design

Zur Untersuchung dieser Fragen werden zwei unterschiedliche Längsschnittkohorten herangezogen. In der ersten Kohorte wurde das professionelle Wissen von Bachelorstudierenden des Lehramts Physik in den Bereichen FW, FDW und PW (Instrument zur Erfassung des PW von Riese & Reinhold, 2012) zu drei Messzeitpunkten (Studienbeginn, Beginn 3. Semester, Beginn 5. Semester) an zehn Hochschulen in Deutschland und einer Hochschule in Österreich im *survey*-Design erhoben (Abb. 1). Dabei handelt es sich um freiwillige Gelegenheitsstichproben, wobei an den meisten Standorten nahezu Vollerhebungen erreicht werden konnten (bei insgesamt geringen Studierendenzahlen in diesen Studiengängen). Zusätzlich werden weitere Personenmerkmale als Kontrollvariablen betrachtet: mathematische Fähigkeiten und Fertigkeiten, Einstellungen zum Erklären, Angaben zu Lerngelegenheiten im Bachelorstudium und im Praxissemester sowie demografische Daten.

Insgesamt wurde pro Messzeitpunkt eine Testzeit von 180 Minuten benötigt, wobei die Befragungen meist verpflichtend im Rahmen zweier Lehrveranstaltungen à 90 Minuten mit wenigen Tagen Abstand stattfanden.

In einer zweiten Kohorte wurden das professionelle Wissen von Masterstudierenden (FW, FDW, PW), alle Kontrollvariablen (s.o.) sowie die Fähigkeiten in den drei o.g. Standardsituationen mit Hilfe von Performanztests vor und nach dem Absolvieren eines Praxissemesters (als Quasi-Intervention) an den Universitäten Aachen, Bremen, Paderborn und Potsdam erfasst (Abb. 1). Bis auf den Performanztest zum Erklären, der nur innerhalb einer freiwilligen Teilstichprobe eingesetzt wurde, handelt es sich

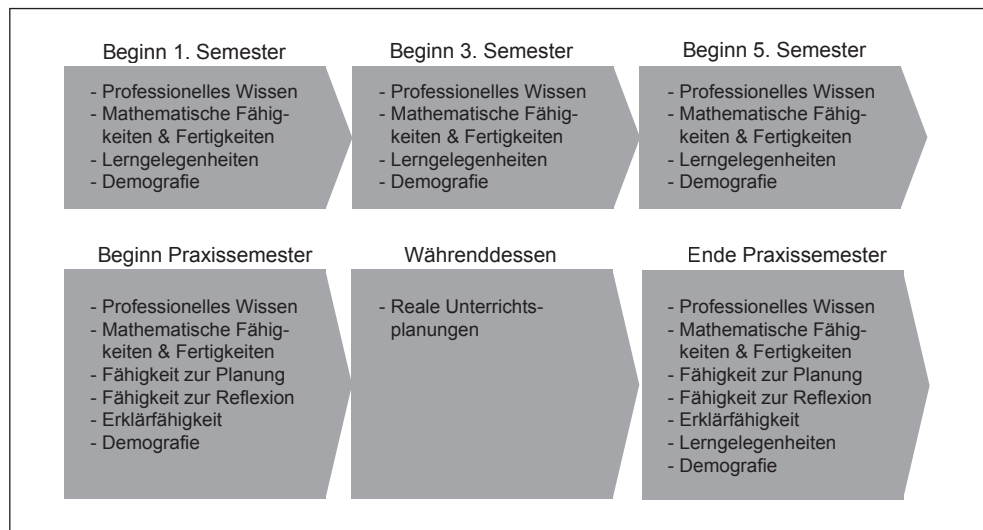


Abb. 1: Studiendesign: oben Längsschnitt während des Bachelorstudiums, unten über das Praxissemester im Masterstudium

im Wesentlichen um Vollerhebungen zum jeweiligen Studienzeitpunkt, so dass Stichprobeneffekte minimiert wurden. Die umfangreichen Datenerhebungen erforderten zu den beiden Messzeitpunkten im Abstand von ca. einem halben Jahr jeweils fünf Befragungstermine, die im Rahmen universitärer Praxisbegleitveranstaltungen innerhalb von jeweils ca. 3 Wochen realisiert wurden. Insgesamt betrug die Testzeit pro Person und Zeitpunkt 330 Minuten. Die kontinuierlich über mehrere Praxissemesterdurchgänge (max. vier, je nach Standort) erhobenen Daten wurden zur Analyse zusammengeführt und auf Kohorteneffekte überprüft (siehe Abschnitt 6).

5. Instrumente

Alle Instrumente beziehen sich auf den physikalischen Inhaltsbereich Mechanik, um eine inhaltliche Passung zwischen Wissens- und Handlungsbereichen herzustellen. Es werden für FDW und die handlungsnahe Erklärfähigkeit zwei breit erprobte und etablierte Instrumente eingesetzt (Riese et al., 2015). Der Test zum FW wurde übernommen, aber unter stärkerer Berücksichtigung des vertieften Schulwissens weiterentwickelt (u. a. da die Hypothese überprüft werden soll, dieser Bereich des Wissens sei zentral für unterrichtliches Handeln). Daher wird im Folgenden auf die neu- bzw. weiterentwickelten Tests zum FW sowie zur Unterrichtsplanung und -reflexion fokussiert.

5.1 Test zur Erfassung des Fachwissens

Das dem Test zu Grunde liegende Wissensmodell (vgl. Riese et al., 2015) unterscheidet drei Subdimensionen (Abb. 3). *Schulwissen (SW)* beinhaltet physikalische Inhalte, die unmittelbarer Bestandteil schulischer Curricula sind. *Universitäres Wissen (UW)* umfasst dagegen physikalische Inhalte, die an Universitäten für Abschlussprüfungen im Fach herangezogen werden. Das Modell wurde weiterentwickelt und in der Kategorie des *vertieften Schulwissens (VSW)* ausgeschärft. Es beinhaltet Wissen zum (1) *Erkennen von Herleitungs- und Lösungsansätzen*, (2) *Identifizieren von Gemeinsamkeiten und Unterschieden physikalischer Konzepte* und (3) zum *Umgang mit Grenzen physikalischer Modelle*. VSW ist im Gegensatz zu SW und UW, die curricular bestimmt werden können, durch theoretische Annahmen zu fachlichen Anforderungen an Lehrkräfte definiert, wenn sie Unterricht auf fachlicher Ebene planen, durchführen und reflektieren (z. B. zur Elementarisierung fachwissenschaftlicher Aussagen).

Diese Modellanpassungen bezüglich des vertieften Schulwissens erforderten auch eine Weiterentwicklung des Testinstruments. Hierzu wurden der bestehende Test mit Items etablierter Instrumente ergänzt bzw. ausgetauscht (vgl. z. B. Sorge et al., 2018) sowie neue Itemsets für das VSW entwickelt (Beispiel in Abb. 3).

Das Testinstrument umfasst 48 *single-choice*-Items (SW: 20, UW: 13, VSW: 12), die sich auf Inhalte der Mechanik beziehen (z. B. Kinematik, Newtonsche Axiome, Energieerhaltung). Es ist für eine Testzeit von 50 Minuten ausgelegt.

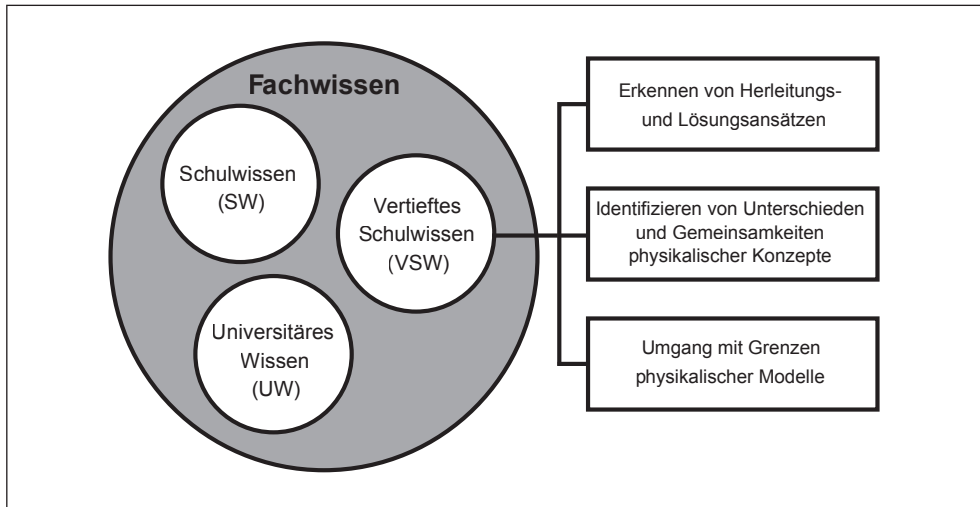


Abb. 2: Modell physikalischen Fachwissens

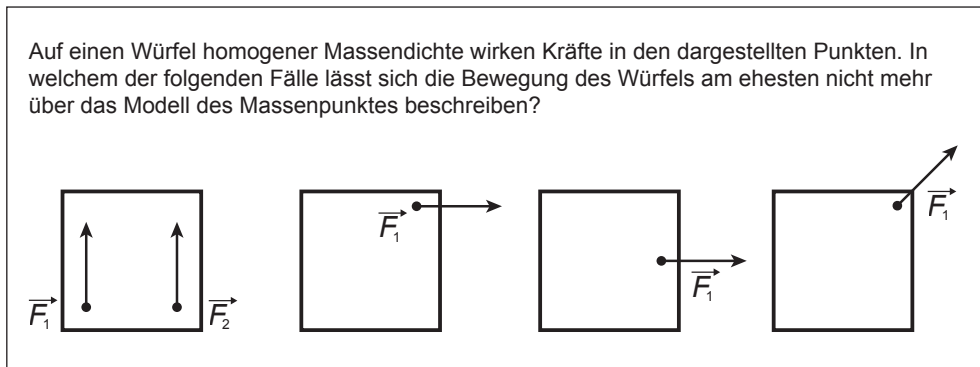


Abb. 3: Beispielim: VSW – Umgang mit Grenzen physikalischer Modelle

5.2 Performanztest zur Planung von Physikunterricht

Unterrichtsplanung ist eine zentrale Anforderungssituation an Physiklehrkräfte und wird häufig als zirkulärer kognitiver Entscheidungsprozess modelliert (vgl. Shavelson & Stern, 1981). Sie erfordert den Entwurf einer professionell begründeten Abfolge von Handlungen, die zum Erreichen eines Ziels beitragen sollen. Vor diesem Hintergrund wurde ein Performanztest entwickelt, der eine standardisierte Erfassung der Planungsfähigkeit ermöglicht. Darin erhalten die Studierenden die Aufgabe, innerhalb einer Testzeit von 60 Minuten für eine fiktive Lerngruppe eine Unterrichtsstunde mit vorgegebenen Lernzielen zum dritten Newtonschen Axiom schriftlich zu planen. Die Aufgabe ist

in eine typische Anforderungssituation des Praxissemesters eingebettet: Eine Praxismentorin bietet der Praktikantin bzw. dem Praktikanten an, die nächste Stunde zu ‚übernehmen‘ und verlangt zuvor eine kurze schriftliche Planung. Die Studierenden erhalten Auszüge aus zwei typischen schulischen Physiklehrwerken sowie einer oft genutzten Online-Quelle als standardisierte Hilfsmittel. Um die Planungen der Studierenden systematischer vergleichen zu können, erhalten sie ein vorstrukturiertes Planungspapier, das in Form kurzer Prompts bestimmte Planungselemente vorgibt (kurze Darstellung des Fachinhalts, erwartete Schülerkognitionen, zentrale Fragestellung, gewählte Aufgaben und Experimente, Tafelbild; vgl. Abb. 4), ein Raster für einen Verlaufsplan beinhaltet und Begründungen einfordert.

Ausgehend vom Modell der didaktischen Rekonstruktion (vgl. Kattman, Duit, Gropengießer & Borowski, 1997), verschiedenen Planungsratgebern sowie fachdidaktischen Lehrwerken wurde ein Manual zur Beurteilung der Planungsqualität der Studierendenantworten entwickelt. Es enthält 26 zwei- oder dreistufige Kodierkategorien für die Planungsbereiche *Fachlicher Inhalt*, *Kompetenzziele*, *Lernvoraussetzungen*, *Elementarisierungen*, *Kontext*, *Experimente* und *Aufgaben*. Die Beurteilung der einzelnen Bereiche erfolgt sowohl lokal (bezogen auf die korrekte und nachvollziehbare Darstellung eines bestimmten Planungselements) als auch global (bezogen auf die Passung des Elements zur gesamten Planung). Alle Beurteilungen der Kodierkategorien werden zu einem quantitativen Gesamtscore durch Summation der 26 Einzelwerte zusammengeführt.

Skizzieren Sie für eine Kreidetafel ein Tafelbild (oder alternativ einen Sicherungstext) mit Skizze, das

- 1.) gemeinsam mit den Schülerinnen und Schülern als Sicherung entwickelt werden könnte und
- 2.) das den Schülerinnen und Schülern zum Lernen und Wiederholen dienen kann.

(1. Demo - Exp) Skizze: Beob: - SuS 1 rollt weg - übt Kraft auf SuS 2 aus - dieser übt ebenfalls Kraft auf SuS 1 aus => Federkraftmessor auslenkt - Kräfte gleich groß	3. Newton'sches Axiom: <u>Das Wechselwirkungsprinzip</u> ▷ actio = reactio übt Körper A eine Kraft $\vec{F}_{A \rightarrow B}$ auf Körper B aus, so wirkt auch dieser eine gleichgroße entgegengesetzt gerichtete Kraft auf Körper A aus. Anm.: - Wechselwirkungskräfte greifen an zwei Körper an	Beispiele - Mensch auf Erde - Trampolin mit Mensch - weitere Bsp. (auf Zugklappe RS)
--	---	--

Abb. 4: Planungsbeispiel zum Prompt Tafelbild

5.3 Performanztest zur Reflexion von Physikunterricht

Unterrichtsreflexion dient der Analyse der Wirkungen von Unterricht sowie seiner Weiterentwicklung (vgl. Plöger, Scholl & Seifert, 2015) und kann als Selbst- oder Fremdreflexion erfolgen. Modelle zur Beschreibung von Unterrichtsreflexion konzeptualisieren diese meist in Form von Reflexionsstufen bzw. aufeinander folgender Schritte (vgl. z.B. Wyss, 2013). Gelungene Reflexionen beinhalten demnach (1) die adäquate *Beschreibung* von Unterrichtssituationen, (2) eine *Bewertung* bzw. ein Urteil über die Handlungen in dieser Situation, (3) das Formulieren von *Handlungsalternativen* sowie (4) die Ableitung von *Konsequenzen* (z.B. für die eigene Professionalisierung). Die vorgenommenen Reflexionsschritte können dabei sowohl begründet (z.B. mit Bezug zu fachdidaktischen Konzepten) als auch unbegründet erfolgen. Für die Erfassung der Reflexionsfähigkeit wurde ein spezifisches Modell zur Reflexion von Physikunterricht (Nowak, Kempin, Kulgemeyer & Borowski, 2019) entwickelt, das erreichte Reflexionsstufen abbildet und dabei unterscheidet, auf welcher Wissensbasis die jeweilige Reflexionshandlung beruht (Abb. 5).

Da eine Unterrichtsreflexion in alltäglichen beruflichen Situationen meist in mündlicher Form erfolgt, wurde ein onlinebasierter Performanztest entwickelt, der ein Reflexionsgespräch in Form von kurzen, mit Hilfe von Schauspielerinnen und Schauspielern erstellten Videosequenzen authentisch simuliert (vgl. Abb. 6). Die Sequenzen enthalten kontrollierte Reflexionsanlässe (z.B. ein ungünstig gewähltes Unterrichtsbeispiel). Die

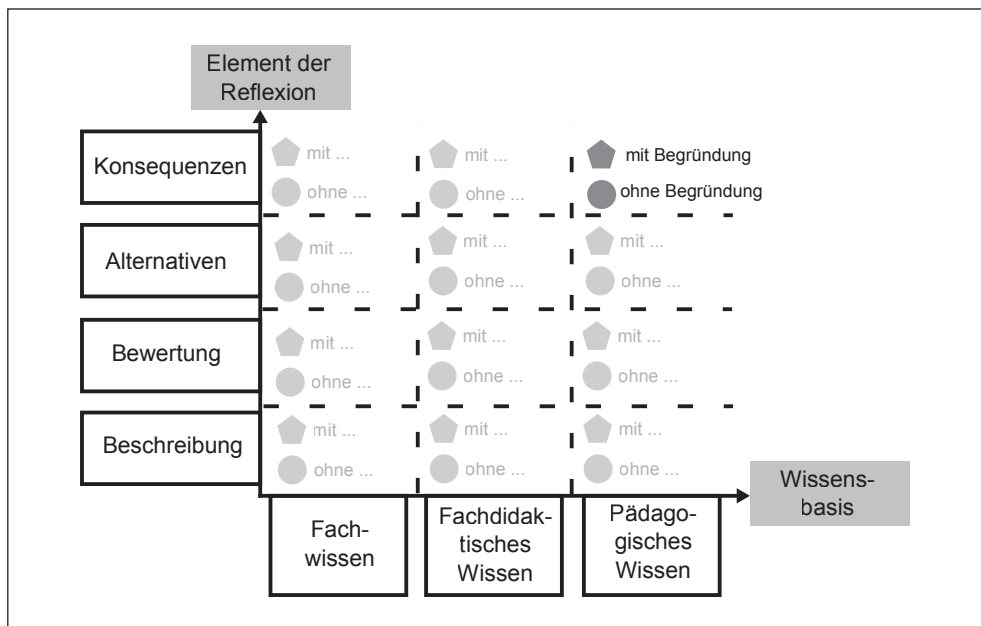


Abb. 5: Modell zur Reflexion von (Physik-)Unterricht

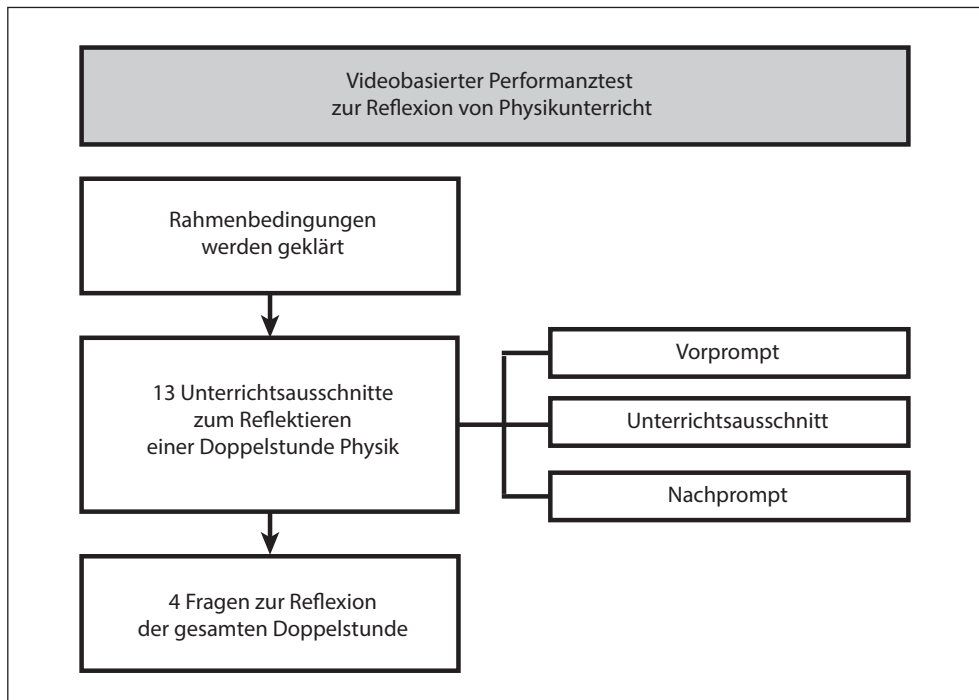


Abb. 6: Ablaufstruktur des Performanztests zur Reflexion von Physikunterricht

Probandinnen und Probanden werden dabei in eine typische Anforderungssituation des Praxissemesters versetzt, in der ein fiktiver Mitpraktikant (Robert) sie um Feedback zu seinem Unterricht im Sinne einer Fremdrelexion bittet.

Zu Beginn erhalten die Studierenden von Robert ebenfalls videobasiert Informationen zu den Rahmenbedingungen seines Unterrichts (z. B. zu angestrebten Zielen). Anschließend erfolgt die Reflexion von 13 Unterrichtsausschnitten aus einer Doppelstunde zum dritten Newtonschen Axiom mit Weiterführung zur Impulserhaltung. Die Sequenzen werden jeweils durch einen Vorprompt, in dem Robert den Ausschnitt im Stundenverlauf verortet, eingeleitet und mit einem Nachprompt abgeschlossen, in dem Robert Reflexionsanregungen gibt („Kannst du dir einen Reim auf die falschen Äußerungen der Schülerinnen und Schüler machen?“). Der Test endet mit vier Fragen zur Reflexion über die Doppelstunde als Ganzes. Die Studierenden sprechen ihre Antworten in ein Headset (Audiomitschnitt) in einer Testzeit von 70 Minuten.

Zur Auswertung wird ein Kodiermanual entwickelt, in dem die Antworten der Studierenden danach kodiert werden, ob und auf welcher Reflexionsstufe sie die definierten Reflexionsanlässe aufgreifen und reflektieren (siebenstufig von 0 = nicht angesprochen bis 7 = Formulierung einer Konsequenz *mit* Begründung). Diese Kodierungen werden summativ in einem quantitativen Maß für Reflexionsfähigkeit zusammengeführt (vgl. Abb. 5).

6. Argumente zur validen Interpretation der Testwerte

Um Argumente für eine valide Interpretation der gewonnenen Testwerte hinsichtlich der Forschungsfragen zu prüfen, wird im Projekt *Profile-P+* ein *argument based approach of validation* verfolgt (vgl. Kane, 2013). Es wird in mehreren Teiluntersuchungen jeweils angestrebt, die Validität einer Testwertinterpretation für Aussagen über Wissen bzw. Handlungsqualität argumentativ abzuwägen. Da die Testwerte primär zur Erklärung von Leistungen interpretiert werden sollen, werden im Anschluss an Messick (1995) Argumente zu verschiedenen Aspekten von Validität gesammelt. Aus Messicks (1995) Argumentklassen werden hier insbesondere Argumente zu *Inhaltsvalidität* (d. h. Abdeckung des Konstrukts durch das Instrument), *kognitiver Validität* (Passung zwischen kognitiven Prozessen bei der Testbearbeitung und intendiertem Konstrukt), *struktureller Validität* (Vergleich von theoretischer und empirisch gefundener Struktur des Konstrukts) sowie zur *externen Validität* (konvergenter bzw. divergenter Vergleich mit anderen Instrumenten) und zur *konsequentiellen Validität* (Passung zwischen Testwerten und ihrer Verwendung) gesammelt. Diese Argumentklassen sind Teil einer übergeordneten Konstruktvalidität.

6.1 Fachwissenstest

Um aus einer Veränderung der Testwerte auf die Entwicklung des schulischen (SW), universitären (UW) und vertieften schulischen Wissens (VSW) schließen zu können, sind der inhaltliche Aspekt (insb. der curriculare) und der strukturelle Aspekt von Validität besonders grundlegend. Zur Untersuchung des inhaltlichen Aspekts wurde unter der Beachtung schulischer und universitärer Curricula und Empfehlungen der Konferenz der Fachbereiche Physik (z. B. KFP, 2012) sowie der Theorie zum vertieften Schulwissen (vgl. Abschnitt 5.1) ein Manual entwickelt, mit der sich die Aufgaben durch zwei unabhängige Experten eindeutig den drei Fachwissenskategorien zuordnen lassen konnten (Cohen's $k = 0.95$). Die curriculare Trennbarkeit zwischen UW und SW wurde mithilfe einer Untersuchung von Schulbüchern bestätigt. In Interviews mit erfahrenen Lehrkräften bestätigte sich die Annahme, dass aus deren Perspektive Aufgaben zum SW und VSW gegenüber Aufgaben zum UW eine höhere Berufsrelevanz besitzen (Pfläging, Enkrott, Buschhüter & Borowski, 2019). Damit lässt sich aussagen, dass sich die Aufgabeninhalte auf die ihnen zugeordneten Fachwissensbereiche beziehen. Zudem wurden *stimulated-recall* Interviews genutzt, um sicherzustellen, dass falsche (bzw. richtige) Antworten tatsächlich mit einem fachlichen (bzw. einem korrekten Verständnis) in Verbindung standen und nicht durch Artefakte in den Aufgaben hervorgerufen wurden (substantieller Aspekt; Messick, 1995).

Um die Fachwissensentwicklung in unterschiedlichen Dimensionen untersuchen zu können, wurde zudem gezeigt, dass die Bereiche auch im statistischen Sinne trennbar sind und dass das gemessene Konstrukt nicht eine gemeinsame Skala (Fachwissen Physik) darstellt (struktureller Aspekt; vgl. Messick, 1995). Hierfür stand ein Datensatz

von insgesamt $N = 861$ Testheften von Lehramtsstudierenden des Fachs Physik beider Kohorten (vgl. Abschnitt 4) sowie Studierender des Fachbachelorstudiums der Physik zur Verfügung. Zur Strukturprüfung wird die Passung der Daten an verschiedendimensionale Raschmodelle geprüft. Tabelle 1 und 2 enthalten den Vergleich eines eindimensionalen Raschmodells mit dem theoretisch angenommenen dreidimensionalen Modell.

Es zeigt sich, dass das dreidimensionale Raschmodell die Datenstruktur signifikant angemessener beschreibt, als ein eindimensionales Modell (EAP-Reliabilität: 0.84).

Dies eröffnet die Möglichkeit, die Fachwissensentwicklung differenziert nach den Bereichen SW, UW und VSW zu betrachten.

Als Bedingung für die Messbarkeit des Wissens zu den unterschiedlichen Messzeitpunkten ist es zudem wichtig, dass sich das Wissen auch zu den einzelnen Zeitpunkten reliabel erheben lässt: In den einzelnen Substichproben ergaben sich zufriedenstellende Reliabilitäten im Bereich von 0.65–0.84 (EAP-Rel.). Lediglich hinsichtlich des ersten Messzeitpunkts des Bachelors beim UW lag der Wert bei 0.57 (EAP-Rel.), was jedoch mit der bis zu diesem Zeitpunkt geringen Ausprägung des UW erklärt werden kann. Die Lehramts-Stichprobe scheint darüber hinaus hinsichtlich der durchschnittlichen Abiturnoten stabil zu bleiben, so ist lediglich eine leichte Verbesserung von 2.02 auf 1.90 von Messzeitpunkt 1 zu 2 zu beobachten.

Die ersten Daten des Längsschnitts der Bachelorstudierenden, die mittels eines individuellen Probandencodes vom 1. Semester über das 3. Semester verbunden werden konnten, zeigen darüber hinaus signifikante Fähigkeitszuwächse auf allen postulierten Dimensionen mit großen Effektstärken (Tab. 4). Zur Datenauswertung wurde eine

	1D-Modell	3D-Modell
AIC	43705	43466
BIC	43934	43709
χ^2 -Test	$\chi^2 = 248.91$; $p < 0.001$	

Tab. 1 Vergleich von 1D- und 3D-Modell bzgl. des Fachwissens

Subdimension	EAP
Schulwissen	0.84
Universitäres Wissen	0.76
Vertieftes Schulwissen	0.76

Tab. 2 Reliabilität der FW-Subdimensionen (3D Modell)

Subdimension	Bachelor		Master		t	p	d
	MW	SD	MW	SD			
SW	-0.099	1.111	0.886	1.136	-5.17	< 0.001	0.88
UW	-0.057	0.809	0.556	0.795	-4.62	< 0.001	0.76
VSW	-0.036	0.798	0.518	1.006	-3.33	0.002	0.68

Tab. 3 Vergleich von Lehramtsstudierenden des Bachelors ($N = 143$) und des Masters ($N = 39$) zum jeweils ersten Messzeitpunkt. Berichtet werden mittlere Personenparameter sowie t -Tests

Dimension	Theta-t ₁		Theta-t ₂		t	p	Cohen's d
	M	SD	M	SD			
SW	-0.08	0.85	0.58	0.85	5.95	< .001	0.88
UW	-0.005	0.55	0.62	0.66	6.31	< .001	0.93
VSW	-0.13	0.49	0.37	0.55	6.36	< .001	0.94

Tab. 4 Längsschnittlicher Vergleich der Bachelor Lehramtsstudierenden vom 1. zum 3. Semester (N = 46)

längsschnittlich geführte Raschanalyse nach Hartig und Kühnbach (2006) genutzt. Dabei werden die Messzeitpunkte bzw. Personen in einer gemeinsamen Schätzung skaliert und Personen, die zu mehreren Messzeitpunkten teilgenommen haben, als sogenannte virtuelle Personen festgesetzt.

6.2 Performanztest zur Planung von Physikunterricht

Bei der Entwicklung des Kodiermanuals wurden zur Untersuchung der Inhaltsvalidität bereits Aspekte curricularer Validität berücksichtigt: Der fachliche Inhalt der zu planenden Unterrichtsstunde ist zentrales Thema schulischer Curricula und gängiger Lehrwerke. Dies wurde auch durch insgesamt sieben Interviews mit Lehrkräften und Fachleitungen für das Fach Physik bestätigt. Zur konvergenten Validierung (Aspekt der strukturellen Validität) wurde geprüft, ob der relative Gesamtscore auf Basis der Kodierkategorien zu den gleichen Beurteilungen führt, wie sie Fachleiter der zweiten Ausbildungsphase aufgrund ihrer professionellen Erfahrungen (ohne Kenntnis der Kategorien) vornehmen. Hierzu bewerteten drei Fachleiter drei Planungsergebnisse des Performanztests, die eine hohe Varianz im Gesamtscore aufwiesen, in typischer Form von Noten (ohne Kenntnis des Bewertungsmanuals). Wie Tabelle 5 zeigt, kamen die Fachleiter dabei zu ähnlichen Einschätzungen der Planungsqualität und der gleichen Qualitätsrangfolge wie der mithilfe des Manuals ermittelte Score.

	Planung A	Planung B	Planung C	
Score gemäß Manual	46 %	17 %	73 %	Qualitätsscore (max. 100 %)
Fachleiter 1	4	6	2	Schulnoten von 1 bis 6
Fachleiter 2	3 bis 4	4 bis 5	2–	
Fachleiter 3	3 bis 4	5	1 bis 2	
Fachleiter Modalwert	4	5	2	

Tab. 5 Ergebnisse der Expertenbefragung dreier Fachleiter

Für eine Überprüfung externer Validität wurden für $N = 17$ Studierende die im Rahmen des Tests angefertigten Unterrichtsplanungen mit jeweils einer schriftlichen Unterrichtsplanung verglichen, die im Praxissemester real erstellt wurde.

Zum Vergleich des Planungsverhaltens wurde mithilfe der qualitativen Inhaltsanalyse nach Mayring (2010) ein Kriterienkatalog mit 66 Kriterien entwickelt. Dabei zeigte sich, dass sich das Planungsverhalten der Studierenden für eine reale Lerngruppe in Vollständigkeit und Umfang nicht wesentlich von den Planungen innerhalb des Performanztests unterscheidet.

Insgesamt konnten über beide Messzeitpunkte vor und nach dem Praxissemester 160 Planungsdokumente erhoben werden, von denen bislang (März 2019) Prä-Post-Daten für $N = 63$ Studierende zusammengefügt werden konnten (der Dropout von ca. 22 % resultiert aus individuellen Fehlzeiten, Studienabbrüche waren während des Praxissemesters nicht festzustellen). Um zu prüfen, ob sich durch den Dropout die Charakteristika der Stichprobe verändern, wurden die demografischen Daten der Probanden für beide Messzeitpunkte verglichen. Dabei ergaben sich keine signifikanten Unterschiede. Zudem wurde für die Substichprobe eine kurze Skala zur Testmotivation ausgewertet (3 Items, $\alpha = 0.93$), wobei sich keine signifikante Änderung zwischen den Messzeitpunkten zeigt.

Zum jetzigen Zeitpunkt sind Prä-Post-Daten von $N = 25$ der aktuell 63 matchbaren Studierenden kodiert. Eine erste Auswertung von Ergebnissen vor und nach dem Praxissemester auf Basis der 26 Kodierkategorien (vgl. Abschnitt 5.2) zeigt einen signifikanten Zuwachs im Summenscore der Planungsfähigkeit mit mittlerer Effektstärke ($d = 0.47$; $p < 0.05$) bei ausreichender Skalenreliabilität ($\alpha = 0.73$). Da Planung von Unterricht einen zentralen Lerninhalt während des Praxissemesters darstellt, kann dies als Argument für Instruktionssensitivität des Instruments gewertet werden (konsequentielle Validität; Polikoff, 2010). Für 30 Planungsdokumente wurde bislang eine Doppelcodierung vorgenommen und ein Gwet's AC₁ von 0.68 erreicht, was einer akzeptablen Übereinstimmung entspricht. Durch weitere Ausschärfungen des Manuals soll die Übereinstimmung weiter verbessert werden.

6.3 Performanztests zur Reflexion von Physikunterricht

Analog zum Performanztest zur Planung wurde auch für den Test zur Reflexion die inhaltliche Validität durch Curriculums- und Lehrwerksanalysen sowie Experteninterviews abgesichert. Aus 154 audiografierten Reflexionen konnten Prä-Post-Daten von bislang $N = 41$ Studierenden zusammengeführt werden (Dropout: ca. 47 %). Auch für diese Stichproben ergaben sich keine signifikanten Unterschiede bzgl. der demografischen Zusammensetzung. Für die zusammengeführten $N = 41$ Studierenden wurde ebenfalls die kurze Skala zur Testmotivation ausgewertet ($\alpha = 0.903$). Hier zeigt sich eine signifikante Minderung der Motivation zwischen den Zeitpunkten, $t(64) = 2.47$, $p < 0.05$, $d = -0.308$.

Zur Prüfung der Anwendbarkeit des Kodiermanuals und der Interraterreliabilität erfolgte bei 17 % der vorliegenden Audiografien eine Kodierung von drei unabhängigen

	Kodierer 1 vs. Kodierer 2	Kodierer 1 vs. Kodierer 3	Kodierer 2 vs. Kodierer 3
Paarweise prozentuale Übereinstimmung	94.1 %	97.0 %	90.4 %
Gemittelte paarweise prozentuale Übereinstimmung	93.3 %		

Tab. 6 Übereinstimmung von drei Kodierenden für die ersten beiden Videosequenzen

Raterinnen und Ratern für die ersten fünf Reflexionssequenzen. Dabei wird ein paarweises Gwet's AC_1 zwischen 0.882 und 0.973 und ein gemitteltes von 0.904 erreicht (Tab. 6), was einer sehr guten Übereinstimmung entspricht.

Betrachtet man in einer ersten Auswertung die Ergebnisse zur Entwicklung der Reflexionsfähigkeit für $N = 41$ Studierende für die ersten fünf (von 13, vgl. Abschnitt 5.3) Reflexionssequenzen vor und nach dem Praxissemester, ergeben sich zum jetzigen Stand keine signifikanten Veränderungen in der Qualität der Reflexionen ($d = 0.18$, $p = 0.274$).

7. Ausblick und zusammenfassende Wertung der Validitätsargumente

Im vorliegenden Projekt wird für die Domäne der Physik eine gleichzeitige, längsschnittliche Analyse von Wissens- und Fähigkeitsfacetten realisiert. Die vorliegenden Ergebnisse liefern Argumente, die für gute Inhaltsvalidität, kognitive Validität, strukturelle Validität, externe Validität und konsequentielle Validität der Verfahren sprechen. Es fehlt allerdings noch eine Validitätsklasse nach Messick (1995): die Generalisierbarkeit, die zentral ist, um aus den Ergebnissen auch Rückschlüsse auf andere Hochschulen ziehen zu können. Es ist zu hoffen, dass aufgrund der Studienanlage eine hohe Übertragbarkeit der Ergebnisse gewährleistet ist, da Universitäten mit sehr unterschiedlichen Studienprogrammen Teil der Stichprobe sind. Dies muss aber nach Abschluss des Projekts weiter kritisch analysiert werden. Ferner müssen bei den Performanztests noch weitere Analysen bis zur finalen Kodierung abgewartet werden, die zur Verbreiterung der Stichprobe sowie zur Erhöhung der Reliabilität führen können und so auch helfen, kleinere Effekte abzubilden. Es stehen außerdem noch weitere Studien zur Validität aus, z. B. zu Prüfungen der prognostischen Validität des FW-Tests (Korrelation zwischen FW und Modulnoten im Studium). Für alle Instrumente werden zentrale Argumente für die Konstruktvalidität zudem noch anhand eines umfangreichen nomologischen Netzwerks überprüft.

Zum Einsatz von Performanztests in der Lehrerbildungsforschung gibt es wenig theoretische oder empirische Bezugspunkte. Wir sehen jedoch durch so erhobene Fähigkeiten ein großes Potential dafür, zu überprüfen, ob universitär vermitteltes Professionswissen wirklich eine zentrale Disposition für Handeln darstellt. Durch den Einsatz der Instrumente zum Professionswissen sowie der Performanztests vor und nach dem Pra-

xissemester können durch Analysen dieser zwei Wellen eines *Cross-lagged-panels* erste Rückschlüsse auf einen möglichen Kausalzusammenhang zwischen Professionswissen und Performanz gezogen werden. Die in diesem Beitrag präsentierten Argumente für die valide Interpretation der Testwerte sprechen dafür, dass diese Analysen fundiert vorgenommen werden können.

Begrenzungen der Aussagekraft der Studie sind u. a. in der aufwändigen Erhebung zu finden: Studierende tendieren mindestens beim Reflexionstest bei späteren Messzeitpunkten dazu, eine niedrigere Testmotivation zu zeigen, was möglicherweise dazu führt, einen Zuwachs in den Fähigkeiten nicht auflösen zu können. Weitere Einschränkungen liegen darin, dass Testinstrumente im gesamten Verlauf des Studiums eingesetzt werden. Nicht nur sind Boden- und Deckeneffekte bei identischen Messungen für Personen zu Beginn und am Ende des Studiums möglich, steigende Expertise kann auch zu einer Veränderung von Testlösestrategien führen. Dies soll in der Folge insbesondere für die schriftlichen Testinstrumente kritisch überprüft werden, indem z. B. die Stabilität der Zusammenhänge der Aspekte des Professionswissens analysiert wird.

Literatur

- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Blömeke, S., Gustafsson, J., & Shavelson, R. (2015). Beyond Dichotomies. *Zeitschrift für Psychologie*, 223(1), 3–13.
- Hartig, J., & Kühnbach, O. (2006). Schätzung von Veränderung mit „plausible values“ in mehrdimensionalen Rasch-Modellen. In A. Ittel & H. Merckens (Hrsg.), *Veränderungsmessung und Längsschnittstudien in der empirischen Erziehungswissenschaft* (S. 27–44). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hume, A., Cooper, R., & Borowski, A. (Hrsg.) (2019). *Repositioning Pedagogical Content Knowledge in Teachers' Professional Knowledge*. Singapore: Springer.
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50(1), 1–73.
- Kattmann, U., Duit, R., Gropengießer, H., & Komorek, M. (1997). Das Modell der Didaktischen Rekonstruktion – Ein Rahmen für naturwissenschaftsdidaktische Forschung und Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 3(3), 3–18.
- KFP = Konferenz der Fachbereiche Physik (2012). *Empfehlung der Konferenz der Fachbereiche Physik zum Umgang mit den Mathematikkenntnissen von Studienanfängern der Physik (Fassung vom 28. 02. 2012)*. <https://www.kfp-physik.de/dokument/KFP-Empfehlung-Mathematikkenntnisse.pdf> [07. 05. 2019].
- Korneck, F., Krüger, M., & Szogs, M. (2017). Professionswissen, Lehrerüberzeugungen und Unterrichtsqualität angehender Physiklehrkräfte unterschiedlicher Schulformen. In H. Fischler & E. Sumfleth (Hrsg.), *Professionelle Kompetenz von Lehrkräften der Chemie und Physik* (S. 113–133). Berlin: Logos.
- 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), 1–26.
- Liepert, S., & Borowski, A. (2019). Testing the Consensus Model: relationships among physics teachers' professional knowledge, interconnectedness of content structure and student achievement. *International Journal of Science Education*, 41(7), 890–910.

- Liu, O. L., Bridgeman, B., & Adler, R. M. (2012). Measuring Learning Outcomes in Higher Education: Motivation matters. *Educational Researcher*, 41(9), 352–362.
- Mayring, P. (2010). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Weinheim und Basel: Beltz.
- Messick, S. (1995). Validity of Psychological Assessment. Validation of inferences from persons' responses as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749.
- Miller, G. E. (1990). The Assessment of Clinical Skills/Competence/Performance. *Academic Medicine*, 65(9), 563–567.
- Nowak, A., Kempin, M., Kulgemeyer, C., & Borowski, A. (2019). Reflexion von Physikunterricht. In C. Maurer (Hrsg.), *Naturwissenschaftliche Bildung als Grundlage für berufliche und gesellschaftliche Teilhabe* (Jahrestagung 2018 in Kiel der Gesellschaft für Didaktik der Chemie und Physik) (S. 838–841). Regensburg: Universität Regensburg.
- Ohle, A., Boone, W. J., & Fischer, H. E. (2015). Investigating The Impact of Teachers' Physics CK on Students Outcomes. *International Journal of Science and Mathematics Education*, 13(6), 1211–1233.
- Pfläging, M., Enkrott, P., Buschhüter, D., & Borowski, A. (2019). Einschätzung der beruflichen Relevanz physikalischen Fachwissens. In C. Maurer (Hrsg.), *Naturwissenschaftliche Bildung als Grundlage für berufliche und gesellschaftliche Teilhabe* (Jahrestagung 2018 in Kiel der Gesellschaft für Didaktik der Chemie und Physik) (S. 739–742). Regensburg: Universität Regensburg.
- Polikoff, M. S. (2010). Instructional Sensitivity as a Psychometric Property of Assessments. *Educational Measurement: Issues and Practice*, 29(4), 3–14.
- Plöger, W., Scholl, D., & Seifert, A. (2015). Analysekompetenz – ein zweidimensionales Konstrukt?!. *Unterrichtswissenschaft*, 43(2), 166–184.
- Riese, J., & Reinhold, P. (2012). Die professionelle Kompetenz angehender Physiklehrkräfte in verschiedenen Ausbildungsformen. *Zeitschrift für Erziehungswissenschaft*, 15(1), 111–143.
- Riese, J., Kulgemeyer, C., Zander, S., Borowski, A., Fischer, H., Gramzow, Y., Reinhold, P., Schecker, H., & Tomczyszyn, E. (2015). Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik. In S. Blömeke & O. Zlatkin-Troitschanskaia (Hrsg.), *Kompetenzen von Studierenden* (61. Beiheft der Zeitschrift für Pädagogik, S. 55–79). Weinheim: Beltz.
- Shavelson, R. J., & Stern, P. (1981). Research on Teachers' Pedagogical Thoughts, Judgments, Decisions, and Behavior. *Review of Educational Research*, 51(4), 455–498.
- Shulman, L. (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*. (Jahrestagung 2017 der Gesellschaft für Didaktik der Chemie und Physik) (S. 114–117). Regensburg: Universität Regensburg.
- Terhart, E. (2012). Wie wirkt Lehrerbildung? Forschungsprobleme und Gestaltungsfragen. *Zeitschrift für Bildungsforschung*, 2(1), 3–21.
- Vogelsang, C., & Cauter, E. (2017). Wie valide sind Professionswissenstests? – Zum Zusammenhang von erfasstem Wissen, Unterrichtshandeln und Unterrichtserfolg. In H. Fischler & E. Sumfleth (Hrsg.), *Professionelle Kompetenz von Lehrkräften der Chemie und Physik* (S. 77–96). Berlin: Logos.
- Walters, K., Osborn, D., & Raven, P. (2005). The Development, Validity and Reliability of a Multimodality Objective Structured Clinical Examination in Psychiatry. *Medical Education*, 39(3), 292–298.

- Woehlecke, S., Massolt, J., Goral, J., Hassan-Yavuz, S., Seider, J., Borowski, A. Fenn, M., & Glowinski, I. (2017). Das erweiterte Fachwissen für den schulischen Kontext als fachübergreifendes Konstrukt und die Anwendung im universitären Lehramtsstudium. *Beiträge zur Lehrerinnen- Und Lehrerbildung*, 35(3), 413–426.
- Woitkowski, D., & Borowski, A. (2017). Fachwissen im Lehramtsstudium Physik. In H. Fischler & E. Sumfleth (Hrsg.), *Professionelle Kompetenz von Lehrkräften der Chemie und Physik* (S. 57–76). Berlin: Logos.
- Wyss, C. (2013). *Unterricht und Reflexion. Eine mehrperspektivische Untersuchung der Unterrichts- und Reflexionskompetenz von Lehrkräften*. Münster: Waxmann.

Abstract: In pre-service teacher education programs prospective physics teachers should acquire professional knowledge that enables them to carry out effective instruction. However, there is little empirical evidence with regard to the development of professional knowledge in the courses of their studies – it is even unclear what knowledge has an impact on teaching quality. In order to be able to analyse these questions, instruments must be developed that permit a valid interpretation of tests cores. Based on instruments developed in project Profile-P+, this article presents a validity argument for the interpretation of test scores for the development of pre-service physics teachers' professional knowledge during a bachelor degree program. We also develop validity arguments for the interpretation of tests cores for the development of their skills to plan and reflect on physics lessons and to explain physics during teaching practice. In addition to knowledge tests, standardized performance tests were used. The results of the analysis suggest that the measured values can be interpreted in the intended sense.

Keywords: Physics, Teacher Education, Professional Knowledge, Performance Assessment, Longitudinal Analysis

Anschrift der Autor_innen

Dr. Christoph Vogelsang, Universität Paderborn,
Zentrum für Bildungsforschung und Lehrerbildung – PLAZ Professional School,
Warburger Str. 100, 33098 Paderborn, Deutschland
E-Mail: christoph.vogelsang@upb.de

Prof. Dr. Andreas Borowski, Universität Potsdam,
AG Didaktik der Physik,
Karl-Liebknecht-Str. 24/25, 14476 Potsdam-Golm, Deutschland
E-Mail: andreas.borowski@uni-potsdam.de

Dr. David Buschhüter, Universität Potsdam,
AG Didaktik der Physik,
Karl-Liebknecht-Str. 24/25, 14476 Potsdam-Golm, Deutschland
E-Mail: buschhue@uni-potsdam.de

M. Ed. Patrick Enkrott, Universität Potsdam,
AG Didaktik der Physik,
Karl-Liebknecht-Str. 24/25, 14476 Potsdam-Golm, Deutschland
E-Mail: enkrott@uni-potsdam.de

M. Ed. Maren Kempin, Universität Bremen,
Institut für Didaktik der Naturwissenschaften, Abteilung Physikdidaktik,
Postfach 330 440, 28334 Bremen, Deutschland
E-Mail: maren.kempin@uni-bremen.de

PD Dr. Christoph Kulgemeyer, Universität Bremen,
Institut für Didaktik der Naturwissenschaften, Abteilung Physikdidaktik,
Postfach 330 440, 28334 Bremen, Deutschland
E-Mail: kulgemeyer@physik.uni-bremen.de

Prof. Dr. Peter Reinhold, Universität Paderborn,
AG Didaktik der Physik,
Warburger Str. 100, 33098 Paderborn, Deutschland
E-Mail: peter.reinhold@upb.de

Prof. Dr. Josef Riese, Rheinisch-Westfälische Technische Hochschule Aachen,
Didaktik der Physik und Technik,
Sommerfeldstraße 14, 52074 Aachen, Deutschland
E-Mail: riese@physik.rwth-aachen.de

Prof. Dr. Horst Schecker, Universität Bremen,
Institut für Didaktik der Naturwissenschaften, Abteilung Physikdidaktik,
Postfach 330 440, 28334 Bremen, Deutschland
E-Mail: schecker@physik.uni-bremen.de

M. Ed. Jan Schröder, Rheinisch-Westfälische Technische Hochschule Aachen,
Didaktik der Physik und Technik,
Sommerfeldstraße 14, 52074 Aachen, Deutschland
E-Mail: jkschroeder@physik.rwth-aachen.de