Konstruktion eines qualitativen Niveaumodells im fachdidaktischen Wissen von angehenden Physiklehrkräften

Dustin Schiering¹ · Stefan Sorge¹ · Stefan Petersen¹ · Knut Neumann¹

Eingegangen: 1. November 2018 / Angenommen: 27. Juni 2019 / Online publiziert: 5. Juli 2019
© Der/die Autor(en) 2019

Zusammenfassung

Der vorliegende Artikel zielt darauf ab, qualitative Fähigkeitsniveaus für das fachdidaktische Wissen angehender Physiklehrkräfte zu formulieren. Neben dem Fachwissen und dem pädagogischen Wissen stellt das fachdidaktische Wissen den zentralen Bereich des Professionswissens von Lehrkräften dar, da es sowohl für die Qualität des Unterrichts als auch für den Lernerfolg der Schülerinnen und Schüler entscheidend ist. Diese besondere Stellung des fachdidaktischen Wissens begründet das Interesse an einer kriterienorientierten Interpretation vorhandener Wissensstrukturen im fachdidaktischen Wissen und seinen Facetten. Für eine solche Interpretation vorhandener Wissensstrukturen sind Niveaumodelle hilfreich, da sie eine Klassifikation unterschiedlicher Wissensqualität erlauben. Zur Etablierung eines Niveaumodells wurden Daten einer Querschnittserhebung von $N=200$ angehenden Physiklehrkräften mit Hilfe des Rasch-Modells analysiert. Die ermittelten Parameter wurden anschließend durch das Scale-Anchoring-Verfahren in numerische Niveaugrenzen überführt. Es lassen sich vier numerische Niveaugrenzen entlang des fachdidaktischen Wissens identifizieren. Eine Analyse der Aufgaben für jedes Fähigkeitsniveau und jede Facette liefert dann eine inhaltliche Beschreibung der Fähigkeitsniveaus. So zeigt sich beispielsweise im Wissen über Instruktionsstrategien, dass Studierende auf unteren Niveaus verschiedene Instruktionsstrategien benennen können und auf höheren Niveaus in der Lage sind, Instruktionsstrategien situativ zu reflektieren. Somit erlaubt das konstruierte Niveaumodell eine kriterienorientierte Interpretation vorhandener Wissensstrukturen im fachdidaktischen Wissen und leistet eine über die bisherigen qualitativen Beschreibungen des fachdidaktischen Wissens hinausgehende Beschreibung der Wissensstruktur, die insbesondere für die universitäre Ausbildung von Bedeutung ist.

Schlüsselwörter Fachdidaktisches Wissen · Niveaumodell · Angehende Lehrkräfte · Physik

✉ Dustin Schiering
schiering@ipn.uni-kiel.de

¹ Didaktik der Physik, IPN – Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik, Olshausenstraße 62, 24118 Kiel, Deutschland

Construction of a Competency Level Model in Pre-service Physics Teachers' Pedagogical Content Knowledge

Abstract

The present study aims at constructing a competency level model in pre-service physics teachers' pedagogical content knowledge. In addition to content knowledge and pedagogical knowledge, pedagogical content knowledge has been suggested as an important vehicle to describe teachers' professional knowledge. That is, pedagogical content knowledge shows significant impact on the quality of teaching and students' learning outcomes. Given the special role of pedagogical content knowledge, there is a great demand to assess teachers' pedagogical content knowledge according to the knowledge structure and to group teachers along their competency level. Nevertheless, no competency level model exists that enables a criteria-oriented interpretation of pedagogical content knowledge structure. For this purpose, data from a cross-sectional study of $N=200$ pre-service physics teachers was analyzed using Rasch analysis. The ability and difficulty parameters then were converted into skill levels by the scale anchoring procedure. Findings show that four competency levels can be identified. Additionally, items' analyses for each level and each facet provide a successive competency level model. For instance, students at lower levels are able to name different instructional strategies, whereas students at higher levels are able to reflect instructional strategies regarding different situations. Thus, the competency level model allows an interpretation of students' knowledge structure and therefore goes beyond previous qualitative descriptions of pedagogical content knowledge. The article discusses the implications of the findings for teacher education and development of pedagogical content knowledge.

Keywords pedagogical content knowledge · competency level model · pre-service teachers · physics

Einleitung

Die Qualität von Unterricht und damit auch das Lernen auf Seiten der Schülerinnen und Schüler werden entscheidend von der Lehrkraft beeinflusst (Hattie 2009). Zur Beschreibung der Fähigkeiten und Fertigkeiten von Lehrkräften, die für qualitativ hochwertigen Unterricht notwendig sind, hat sich im deutschsprachigen Raum das *Modell der professionellen Handlungskompetenz* durchgesetzt (Baumert und Kunter 2006). Das darin definierte Professionswissen beinhaltet die Wissensbereiche pädagogisches Wissen (PW), Fachwissen (FW), fachdidaktisches Wissen (FDW) und Organisations- und Beratungswissen. Dem FDW wird dabei die zentrale Rolle für die erfolgreiche Unterrichtsgestaltung zugeschrieben (Riese 2009). Unterstützt wird diese Annahme durch verschiedene Untersuchungen, die zeigen, dass das FDW der Lehrkraft sowohl für kognitiv aktivierenden Unterricht (Ball et al. 2001; Keller et al. 2017) als auch das Lernen der Schülerinnen und Schüler bedeutsam ist (Baumert et al. 2010; Keller et al. 2017). Um vor dem Hintergrund dieser Ergebnisse der Forderung nach einer stetigen Evaluation und Optimierung der Lehramtsausbildung nachzukommen (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) 2005), sind differenzierte Kenntnisse über die Wissensstrukturen von Lehramtsstudierenden erforderlich. Bestehende Instrumente zur Messung des FDW erlauben jedoch „nur globale Aussagen zum FDW auf der Ebene von Gesamtscores“ (Riese et al. 2017, S. 100) und damit „kei-

ne kriterienorientierte Beurteilung der jeweils vorliegenden Wissensstruktur“ (Woitkowski und Riese 2017, S. 40).

Eine Lösung bieten sogenannte Niveaumodelle, welche das Wissen kumulativ beschreiben (Klieme et al. 2007) und eine inhaltliche Klassifizierung von Personen hinsichtlich ihrer Wissensausprägung ermöglichen (Woitkowski und Riese 2017). Woitkowski und Riese (2017) haben bereits ein solches Niveaumodell für das physikalische FW von Lehramtsstudierenden im Bereich Mechanik entwickelt, bei dem sie auf Grundlage schwierigkeiterzeugender Aufgabenmerkmale aufeinander aufbauende Niveaus formulierten. Vergleichbare Niveaumodelle für das FDW existieren bisher jedoch nicht.

Diese Arbeit beschäftigt sich mit der Bestimmung qualitativer Fähigkeitsniveaus im FDW angehender Physiklehrkräfte. Hierfür wurden 200 angehende Physiklehrkräfte aus zwölf verschiedenen deutschen Universitäten mit Hilfe des im Projekt *Messung professioneller Kompetenzen in mathematischen und naturwissenschaftlichen Lehramtsstudiengängen* (Akronym: KiL) entwickelten Testinstrumentes zu ihrem FDW befragt (siehe auch Sorge et al. 2019). Die Ergebnisse der Untersuchung werden zur Charakterisierung qualitativer Fähigkeitsniveaus im FDW genutzt.

Theoretischer Hintergrund

Die professionelle Kompetenz einer Lehrkraft ist Grundlage für ihr effektives und erfolgreiches unterrichtliches Handeln (Baumert und Kunter 2006; Lipowsky 2006). Im

Zentrum professioneller Handlungskompetenz steht das Professionswissen der Lehrkräfte (Neumann et al. 2019). Auf Grundlage der Modellierung von Baumert und Kunter (2006) „herrscht weitgehend Konsens im Hinblick auf die [...] Grobstruktur des Professionswissens von (angehenden) Lehrkräften“ (Riese et al. 2015, S. 58), das in Fachwissen (FW), allgemeines pädagogisches Wissen (PW) und fachdidaktisches Wissen (FDW) unterteilt wird (Baumert und Kunter 2006). Das PW umfasst beispielsweise das Wissen über Klassenführung und Entwicklung der Lernenden (Baumert und Kunter 2006), das FW unter anderem das Wissen über systematische Lösungswege fachspezifischer Probleme sowie fachspezifische Arbeitsweisen (Shulman 1986).

Für das FDW sind die Beschreibungen und Konzeptualisierungen hingegen weniger einheitlich (Kirschner 2013; Riese 2009). Die Konzeptionen und die Entwicklungen der Modelle des FDW beziehen sich größtenteils auf Shulmans (1986) Definition des *pedagogical content knowledge* (Riese et al. 2015). So kann FDW als das Wissen beschrieben werden, mit dessen Hilfe Lehrende Fachinhalte aufbereiten können, um sie den Lernenden verständlich zu machen (Shulman 1987).

Facetten des FDW

Das FDW besteht wie auch das PW oder FW aus verschiedenen Teilaspekten, sogenannten Facetten. Abhängig vom Rahmenmodell des Professionswissens und der Konzeptualisierung des FDW zählen verschiedene Autoren verschiedene Facetten zum FDW. Für Shulman (1987) umfasst das FDW die Facetten Wissen über Lernschwierigkeiten der Schülerinnen und Schüler sowie Wissen über Instruktionsstrategien und Repräsentationen. Darauf aufbauend erweitert Grossman (1990) die innere Struktur des FDW um das Wissen über Curriculum und die Orientierungen gegenüber naturwissenschaftlichem Unterricht. Begründet wird dies damit, dass Lehrkräfte das Curriculum benötigen und nutzen, um ihren Unterricht zu strukturieren (Grossman 1990). Auf Basis der Überlegung von Tamir (1988) ergänzen Magnusson et al. (1999) die Facetten des FDW um das Wissen über Assessment (Diagnostik und Leistungsmessung). Als Argument dafür wird angeführt, dass Lehrkräfte Wissen über Assessment benötigen, um Leistungen der Schülerinnen und Schüler und damit ihren Unterricht angemessen zu evaluieren. Für Kirschner (2013) umfasst das FDW von Physiklehrkräften hingegen die Facetten Wissen über (fachliche) Konzepte, über Experimente und über Vorstellungen der Schülerinnen und Schüler. Riese et al. (2017) beziehen in ihrer Untersuchung zum FDW die Facetten Instruktionsstrategien, Vorstellungen der Schülerinnen und Schüler, fachdidaktische Konzepte sowie die Facette Experimente und Vermittlung eines angemessenen Wissenschafts-

verständnisses ein und folgen mit dieser Modellierung des FDW dem Modell von Gramzow (2015). Die Konferenz der Kultusminister formuliert in ihren inhaltlichen Anforderungen für die Fachdidaktik Physik in der Lehrerbildung sieben Facetten, darunter der Umgang mit Heterogenität im Physikunterricht, Motivation und Interesse sowie die Diagnose von Lernschwierigkeiten (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) 2018).

Sowohl die Facetten fachdidaktische Konzepte sowie Experimente und Vermittlung eines angemessenen Wissenschaftsverständnisses bei Riese et al. (2017), als auch die Facetten (fachliche) Konzepte und Experimente bei Kirschner (2013) können der übergeordneten Facette Instruktionsstrategien von Magnusson et al. (1999) zugeordnet werden (Kirschner 2013). Auf gleiche Weise lassen sich die von der Konferenz der Kultusminister formulierten sieben Facetten (KMK 2018) in dem Modell von Magnusson et al. (1999) wiederfinden. So wird der Umgang mit Heterogenität im Physikunterricht in den Instruktionsstrategien, Motivation und Interesse in Vorstellungen der Schülerinnen und Schüler und die Diagnose von Lernschwierigkeiten im Wissen über Assessment adressiert. Diesbezüglich heben auch Kind (2009) sowie Park und Oliver (2008) hervor, dass das Modell von Magnusson et al. (1999) als Konsens bei der Modellierung des FDW verstanden werden kann („Magnusson et al. (1999) propose a model that seems to encompass what is needed.“; Kind 2009, S. 198).

Die im Modell von Magnusson et al. (1999) genannten Orientierungen gegenüber dem naturwissenschaftlichen Unterricht als eigene Facette *innerhalb* des FDW zu sehen, ist nicht unproblematisch. So sind nach Abell (2007) die Orientierungen eher als *Interaktion* von Wissen, Beliefs und Werten zu verstehen. Auch im Modell professioneller Handlungskompetenz nach Baumert und Kunter (2006) stellen Orientierungen, Überzeugungen und Werthaltungen eigene Aspekte professioneller Kompetenz dar. Deshalb wird im weiteren Verlauf davon abgesehen, Orientierungen gegenüber dem naturwissenschaftlichen Unterricht als Facette des FDW zu berücksichtigen. Als Facetten des FDW werden in dieser Studie daher das Wissen über Curriculum, Vorstellungen der Schülerinnen und Schüler, Assessment und Instruktionsstrategien (Magnusson et al. 1999) berücksichtigt.

Entwicklung des FDW

Für die Entwicklung des FDW bieten sich im Laufe der eigenen Schulzeit, des Studiums und der Arbeit als Lehrkraft verschiedene Lerngelegenheiten (Kunter et al. 2011). Diese lassen sich in formale (beispielsweise in Form von Vorlesungen oder Seminaren) und informelle Lerngelegenheiten (beispielsweise selbst erteilte Nachhilfe) gruppieren (Sorge

et al. 2019). Lortie (1975) räumt insbesondere den Erinnerungen an die eigene Schulzeit als informelle Lerngelegenheit eine besondere Stellung ein. Im Vergleich zu Novizen aus anderen Disziplinen wie Jura oder Medizin haben angehende Lehrkräfte bereits Unterrichtsstunden von erfahrenen Lehrkräften beobachtet (Borg 2004). Eben diese Erfahrungen können die professionelle Kompetenz angehender Lehrkräfte maßgeblich formen („apprenticeship of observation“, Lortie 1975). Grossman (1990) bewertet diesen Umstand in der Entwicklung des FDW als durchaus problematisch. So ist das auf diese Weise erlernte FDW verbunden mit eigenen Unterrichtserinnerungen, was zu einem Konservatismus in der Lehre führen kann (Grossman 1990). Darüber hinaus besteht die Gefahr, dass Lehrkräfte ihre eigenen Erfahrungen aus der Schulzeit für repräsentativ halten. Als Konsequenz daraus könnte sich beispielsweise das Wissen über die Präkonzepte der Schülerinnen und Schüler alleine aus den eigenen (vergangenen) Präkonzepten speisen (Grossman 1990). Angehende Lehrkräfte lernen somit unbewusst in informellen Lerngelegenheiten und sind nicht notwendigerweise in der Lage, dieses Wissen zu explizieren (Lortie 1975; Grossman 1990). Man könnte in diesem Zusammenhang also von einem *intuitiven* FDW (iFDW) sprechen, das eine angehende Lehrkraft bereits zu Beginn ihrer universitären Ausbildung besitzt.

Bezüglich der Wirksamkeit verschiedener Lerngelegenheiten zeigen diverse Studien (z. B. Riese 2009; Kleickmann et al. 2013), dass formale Lerngelegenheiten erwartungsgemäß einen positiven Einfluss auf die Entwicklung des FDW haben. Im Allgemeinen kann deshalb gesagt werden, dass Studierende in der zweiten Hälfte ihres Studiums über ein signifikant höheres FDW verfügen sollten als Studierende zu Beginn ihres Studiums (z. B. Kleickmann und Anders 2011; Riese et al. 2017; Sorge et al. 2019). Erste Ansätze, diese Entwicklung im FDW angehender Lehrkräfte qualitativ zu beschreiben, bietet Berliner (1988), der die Expertise von Lehrkräften in fünf Stufen unterteilt: novice, advanced beginner, competent, proficient und expert. Darauf aufbauend modellieren Schneider und Plasman (2011) die Entwicklung des FDW von Studierenden bis hin zu erfahrenen Lehrkräften. Zur inhaltlichen Ausdifferenzierung der Entwicklung für einzelne Facetten nach Magnusson et al. (1999) ziehen sie insgesamt 91 Forschungsartikel aus den Jahren 1986 bis 2010 zum FDW von Lehrkräften mit unterschiedlicher Berufserfahrung (preservice, new, some experience, much experience, leader) heran. Beispielhaft sei im Folgenden die Entwicklung für die Facette *Instruktionsstrategien* dargestellt:

Progression. Inquiry strategies are activities that are hands on or that lead to „discovery“, are difficult to enact, and may be inappropriate for students → Inquiry strategies are primarily opportunities

to collect data through observations or experimentation and can be teacher centered → Inquiry strategies are opportunities for student to pose questions or collect and work with their own data, and traditional lessons can be converted to inquiry lessons → Inquiry strategies include students posing questions, designing investigations, collecting evidence, and making claims (with instruction). (Schneider und Plasman 2011, S. 550–551)

Über die verschiedenen Facetten hinweg charakterisieren Schneider und Plasman (2011) die Entwicklung des FDW wie folgt: „Overall teacher thinking appears to progress first to thinking about learners, then to thinking about teaching, and finally to building a repertoire“ (Schneider und Plasman 2011, S. 555). Somit konnten Schneider und Plasman (2011) die Entwicklung des FDW generisch beschreiben und für einzelne Facetten nach Magnusson et al. (1999) ausdifferenzieren. Sie stellen aber auch fest, dass es große Unterschiede in der Anzahl und Qualität der ausgewerteten Forschungsartikel gab. Beispielsweise waren vergleichsweise viele Artikel zum Wissen über Präkonzepte der Schülerinnen und Schüler verfügbar. Die ohnehin wenigen Artikel über das Wissen zu Assessment hingegen beschäftigten sich unterdessen eher mit dem was Lehrkräfte tun, anstatt mit dem was sie wissen (Schneider und Plasman 2011). Zudem beschreiben Schneider und Plasman (2011) die Entwicklung über die ganze berufliche Laufbahn von Lehrkräften. Insofern stellt die Arbeit von Schneider und Plasman (2011) zwar einen geeigneten Rahmen für die Modellierung der Entwicklung des FDW dar, erfordert aber eine weitere Auswertung im Hinblick auf die Charakterisierung der Entwicklung des FDW *über das Studium hinweg*.

Fragestellungen

Das FDW einer Lehrkraft hat einen bedeutenden Einfluss für das Lernen der Schülerinnen und Schüler, die kognitive Aktivierung sowie die Unterstützung von Lernprozessen (z. B. Keller et al. 2017). Die gezielte Förderung des FDW erfordert ein Modell, das es erlaubt, das aktuelle FDW der Lehramtsstudierenden einzuschätzen und Vorgaben macht, welche Aspekte des FDW im nächsten Schritt zu behandeln sind; das heißt, eine systematische Förderung erfordert ein Niveaumodell des FDW. Die Vorarbeiten von Schneider und Plasman (2011) stellen eine gute Basis dafür dar, sind jedoch mit Blick auf das Lehramtsstudium nicht hinreichend ausdifferenziert. Um dieses Forschungsdefizit aufzulösen, lautet die erste Fragestellung dieser Arbeit:

FS 1 Welche Fähigkeitsniveaus lassen sich im fachdidaktischen Wissen (FDW) angehender Physiklehrkräfte beschreiben?

In Hinblick auf eine optimale Unterstützung beim Übergang von der Schule an die Universität sind insbesondere Studienanfänger von Interesse. Den Überlegungen von Lortie (1975) und Grossman (1990) folgend liegt es nahe, dass Studienanfänger bereits über ein spezielles FDW verfügen. Auch die Ergebnisse von Riese et al. (2017) lassen vermuten, dass Studierende des ersten Studienjahres über ein FDW verfügen, wenngleich sie keine universitären fachdidaktischen Veranstaltungen besucht haben. Dieses postulierte iFDW könnte also eine Basis für zukünftige Lehrveranstaltungen sein. Somit ergibt sich die zweite Fragestellung:

FS 2 In welchem Umfang zeigen Studienanfänger intuitives fachdidaktisches Wissen (iFDW)?

Methoden

Entwicklung des Testinstruments

Das Projekt *Messung professioneller Kompetenzen in mathematischen und naturwissenschaftlichen Lehramtsstudiengängen* (Akronym: KiL) hatte sich zum Ziel gesetzt, ein Instrument zu konstruieren, welches eine Messung des universitär erworbenen Professionswissens bei Lehramtsstudierenden für mathematisch-naturwissenschaftliche Fächer ermöglicht (Kleickmann et al. 2014). Im Test werden die Wissensdomänen PW, FW und FDW erfasst, wobei sich FW und FDW in den Fächern Biologie, Chemie, Physik und Mathematik fachspezifisch unterscheiden. Da der Test die Wissensstrukturen im Studium abbilden soll, wird ein breites Spektrum an Wissensfacetten und -niveaus adressiert. Darum umfasst der Test zum FDW angehender Physiklehrkräfte die Facetten Vorstellungen der Schülerinnen und Schüler, Instruktionsstrategien, Curriculum und Assessment (Magnusson et al. 1999) in verschiedenen Inhaltsbereichen wie Mechanik, Elektrizitätslehre, Atom- und Kernphysik und weiteren Inhaltsbereichen (für Details siehe Sorge et al. 2019).

Das Wissen über Vorstellungen der Schülerinnen und Schüler umfasst im Testinstrument unter anderem typische Fehl-, beziehungsweise Präkonzepte von Schülerinnen und Schülern zu physikalischen Phänomenen und physikspezifische Denkweisen sowie Besonderheiten im Lernen physikalischer Inhalte (z. B. die Theorie des Conceptual Change). Der Bereich Instruktionsstrategien beinhaltet Strategien zur Unterrichtsgestaltung vor dem Hintergrund spezieller Lerngruppen oder spezieller physikalischer Inhalte. Außerdem wird adressiert, welche Methoden, Medien und Arbeitsweisen im Unterricht eingesetzt werden. Das Wissen über das

Curriculum schließt unter anderem die Kenntnis der Bildungsstandards Physik ein, beispielsweise das Wissen von Kompetenzen, die Schülerinnen und Schüler im Physikunterricht erwerben sollen. Im Bereich Assessment umfasst das Testinstrument verschiedene Formen der Bewertung und Beurteilung von Schülerinnen und Schüler sowie Kriterien für Aufgaben zur Leistungsmessung und Diagnostik.

Alle Aufgaben zum FDW wurden auf Grundlage verschiedener Standards zur Lehramtsausbildung von einem Team ausgewiesener Fachdidaktikerinnen und Fachdidaktiker neu entwickelt. Nach ersten Machbarkeitsstudien im Jahr 2011, der Pilotierung im Jahr 2012 und Expertenratings wurden schließlich 39 Aufgaben für die Hauptstudie ausgewählt (davon 12 Aufgaben zum Wissen über Vorstellungen der Schülerinnen und Schüler, 11 zum Wissen über Instruktionsstrategien und jeweils 8 zum Wissen über Curriculums und Assessment), von denen 18 offene Aufgaben, 15 Aufgaben im Multiple-Choice-Format, 2 Wahr-Falsch-Aufgaben, 3 Zuordnungsaufgaben sowie eine Aufgabe im short-answer-Format vorliegen.

Die Hauptstudie fand am Ende des Sommersemesters 2013 an 12 deutschen Hochschulen statt. Die Testung dauerte vier Stunden mit zwei 15-minütigen Pausen und wurde von Mitarbeiterinnen und Mitarbeitern des Projektes durchgeführt.

Stichprobe

Die Stichprobe der Hauptstudie umfasst $N=200$ Lehramtsstudierende (41 % weiblich) des Faches Physik aus verschiedenen Semestern. Das durchschnittliche Alter der Stichprobe beträgt 23,7 Jahre ($SD=3,0$ Jahre), die durchschnittliche Studienzeit entspricht 2,85 Jahren ($SD=1,31$ Jahre). 142 Probanden studieren mit dem Ziel des Gymnasiallehramts. Für die genauere Stichprobenverteilung siehe Abb. 1. Die Gruppe der Studierenden im ersten

Abb. 1 Verteilung der Studierenden über die Studienjahre. Mit ≥ 5 sind alle Studierenden bezeichnet, die zum Zeitpunkt der Befragung im fünften oder einem höheren Studienjahr sind

Studienjahr umfasst $N = 30$ Probanden (47 % weiblich). Das durchschnittliche Alter dieser Subgruppe beträgt 22,5 Jahre ($SD = 2,7$ Jahre). Der Test zum FDW zeigt für die Stichprobe insgesamt eine gute WLE-Reliabilität von 0,73.

Konstruktion der Fähigkeitsniveaus

Testverfahren, die das Wissen von Probanden lediglich in Form von Gesamtpunktzahlen kommunizieren, besitzen eine limitierte Aussagekraft. Dieses Problem lässt sich durch die Formulierung sogenannter *Niveaus* auflösen. Dabei versteht man unter einem *Niveau* eine kriteriumsorientierte Beschreibung der fachbezogenen Fähigkeiten auf Grundlage der Punktzahlen bzw. numerischen Ergebnisse eines Tests (Hartig 2007). Einige Autoren wie zum Beispiel Woitkowski und Riese (2017) oder Hartig (2007) sprechen in diesem Zusammenhang auch von sogenannten *Kompetenzniveaus*. Da durch das hier verwendete Testinstrument aber eher (kognitive) Fähigkeiten erfasst werden, werden im Verlauf *Fähigkeitsniveaus* im fachdidaktischen Wissen angehender Physiklehrkräfte diskutiert.

Die grundlegende Schwierigkeit in der Formulierung solcher Niveaus liegt in der Festsetzung der eigentlichen Niveaugrenzen, wobei diverse Verfahren in der Literatur diskutiert werden: Bookmark-Verfahren, regressionsanalytischer Ansatz, Scale-Anchoring Verfahren (für eine Übersicht siehe z.B. Woitkowski und Riese 2017). Im Bookmark-Verfahren (Mitzel et al. 2001) werden von einer Kommission aus Expertinnen und Experten Aufgaben nach ihrer Schwierigkeit beurteilt (z.B. Stanat et al. 2012). Das Gremium setzt entlang der Aufgaben Markierungen (sog. bookmarks), an denen für eine Kompetenzstufe gerade noch kompetente Personen diese Aufgaben richtig lösen würde (Stanat et al. 2012). Schecker et al. (2016) führen zum Bookmark-Verfahren aus, dass die Setzung der Stufen einen hohen Inferenzgrad besitzt und eine Objektivität nur begrenzt möglich ist. Um die Problematik der subjektiven Einschätzungen durch eine Expertenkommission zu vermeiden, verwenden Woitkowski und Riese (2017) für die Konstruktion eines Niveaumodells einen sogenannten regressionsanalytischen Ansatz. Bei diesem Verfahren werden Items gemäß ihrer zum Lösen nötigen Kognition (z.B. gemäß ihrer Komplexitätsstufe) durch Regressionsanalysen gruppiert (Woitkowski und Riese 2017). Dieses Verfahren bedarf jedoch eines theoretisch fundierten Modells zu schwierigkeiterzeugenden Merkmalen, da die Aufgaben a priori kategorisiert werden. Für Aufgaben zum FDW existiert bislang in der Literatur jedoch kein etabliertes Modell für schwierigkeiterzeugende Merkmale.

Das hier verwendete Scale-Anchoring-Verfahren, welches beispielsweise in TIMSS und PIRLS 2011 (Mullis 2012) aber auch in ähnlicher Form bei PISA 2009 (OECD 2012) angewandt wurde, umgeht beide Problematiken. Ziel

des Scale-Anchoring-Verfahrens ist es, numerische Niveaugrenzen entlang der Personenfähigkeiten, bzw. Aufgabenschwierigkeiten zu entwickeln (gemäß des Rasch-Modells befinden sich Personenfähigkeiten und Aufgabenschwierigkeiten auf einer gemeinsamen Skala, siehe zum Beispiel Bond und Fox 2007). Dazu werden Personengruppen gebildet, gruppenweise die Lösungswahrscheinlichkeiten für einzelne Aufgaben berechnet und Aufgaben anhand ihrer Lösungswahrscheinlichkeiten in den einzelnen Gruppen zusammengefasst (Mullis 2012). Aus diesen Aufgabengruppen ergeben sich die numerischen Niveaugrenzen. Stehen diese numerischen Grenzen der Fähigkeitsniveaus fest, können typische Aufgaben an diesen Grenzen analysiert werden, um diese (rein numerischen) Grenzen nun auch inhaltlich zu charakterisieren. So ergeben sich qualitative Fähigkeitsniveaus. Die Entwicklung von Niveaugrenzen in dieser Arbeit geschieht in fünf Schritten und orientiert sich an dem Verfahren von Mullis (2012).

Im *ersten Schritt* werden die aus einer Rasch-Analyse (Rasch 1960) der Testergebnisse ermittelten Personenfähigkeiten auf eine praktikablere Skala von 300 bis 700 Fähigkeitspunkten transformiert (vgl. Mullis 2012). Die mittlere Personenfähigkeit beträgt nach Transformation 517,7 Fähigkeitspunkte ($SD = 80,6$ Fähigkeitspunkte).

Im *zweiten Schritt* werden entlang der Fähigkeitsskala drei Abschnitte markiert (hier 400–450 Punkte, 500–550 Punkte und 600–650 Punkte). Probanden, die entsprechend ihrer Fähigkeit in einen dieser Abschnitte zugeordnet werden können, werden identifiziert und in korrespondierenden Personengruppen (1, 2 und 3) zusammengefasst. Die Anzahl der definierten Abschnitte richtet sich nach der Anzahl der intendierten Niveaugrenzen. Die Breite der Abschnitte hängt stark von der gesamten Stichprobe und der Standardabweichung ihrer Fähigkeiten ab. Einerseits ist zu beachten, dass über die Abschnitte (und damit die korrespondierenden Personengruppen) hinweg ein möglichst heterogenes Fähigkeitsspektrum abgebildet wird. So können im Folgenden möglichst unterschiedliche Personen zur Analyse von Aufgaben genutzt werden. Andererseits sollten die Abschnitte so gewählt werden, dass ein möglichst homogenes Fähigkeitsspektrum der Probanden innerhalb eines Abschnitts (und damit innerhalb der korrespondierenden Personengruppe) vorliegt und die einzelnen Abschnitte (und damit die korrespondierenden Personengruppen) dennoch genügend Personen umfassen (Beaton und Allen 1992).

Auf Grund der relativ geringen Stichprobengröße wird für diese Arbeit festgelegt, dass jeder Abschnitt (und damit jede korrespondierende Personengruppe) mindestens 20 Probanden (10 % der Gesamtstichprobe) umfassen soll. Tab. 1 zeigt für die drei markierten Abschnitte die korrespondierenden Personengruppen mit Probandenanzahl N , mittlerer Fähigkeit M und Standardabweichung SD sowie das Ergebnis eines t-Testes zwischen den Gruppen. Die Stan-

Tab. 1 Zusammenfassung der drei Personengruppen im Rahmen des Scale-Anchoring-Verfahrens

Personengruppen	Abschnitt auf der Fähigkeitsskala	<i>N</i>	<i>M</i>	<i>SD</i>	t-Wert
1	400–450	26	431,7	12,1	30,7*** (zwischen 1 und 2)
2	500–550	45	527,6	13,7	24,5*** (zwischen 2 und 3)
3	600–650	23	621,7	15,6	47,2*** (zwischen 1 und 3)

*** $p < 0,001$

dardabweichungen der drei Personengruppen zeigen, dass es sich innerhalb der Personengruppen um ein relativ homogenes Fähigkeitsspektrum handelt. Mit dem t-Test kann überprüft werden, ob der Abstand zwischen den Personengruppen groß genug ist, um von verschiedenen Fähigkeitsspektren auszugehen. Alle drei Personengruppen zeigen einen signifikanten Unterschied zueinander und besitzen somit einen ausreichend großen Abstand.

Im *dritten Schritt* wird nun für jede Aufgabe der prozentuale Anteil an Probanden pro Personengruppe berechnet, die die Aufgabe korrekt gelöst haben. So zeigt sich beispielsweise, dass die Aufgabe 1 von rund 38 % der Probanden aus Personengruppe 1, 67 % der Probanden aus Personengruppe 2 und 87 % der Probanden aus Personengruppe 3 richtig gelöst wurde. Bei polytomen Aufgaben muss jedoch genauer differenziert werden. Hat eine Aufgabe beispielsweise vier richtige Antworten, so wird der prozentuale Anteil an Probanden berechnet, die mindestens eine, mindestens zwei, mindestens drei und genau vier richtige Antworten gegeben haben (vgl. Mullis 2012). Es handelt sich demgemäß um kumulierte Wahrscheinlichkeiten. Dies ist zum einen sinnvoll, da Probanden mit zwei richtigen Antworten eben auch eine richtige Antwort gegeben haben und zum anderen, da es so bei besonders leichten Aufgaben nicht dazu kommen kann, dass eine korrekte Antwort unwahrscheinlicher ist als vier korrekte Antworten. Polytome Aufgaben mit n möglichen Punkten werden also als n verschiedene Aufgaben interpretiert. Aus diesem Grund ist im Verlauf häufig von 56 (statt 39) Aufgaben die Rede.

Im *vierten Schritt* werden mit Hilfe der prozentualen Anteile an Probanden mit korrekter Lösung die Aufgaben in Aufgabengruppen nach folgenden Kriterien zusammengefasst (Mullis 2012):

- Eine Aufgabe wird zur Aufgabengruppe I zugeordnet, falls $\geq 55\%$ der Personen aus Personengruppe 1 die Aufgabe richtig gelöst haben.
- Eine Aufgabe wird zur Aufgabengruppe II bzw. III zugeordnet, falls $\geq 55\%$ der Personen aus Personengruppe 2 bzw. 3 und $\leq 50\%$ der Personen aus Gruppe 1 bzw. 2 die Aufgabe richtig gelöst haben.
- Eine Aufgabe wird zur Aufgabengruppe >III zugeordnet, falls $< 50\%$ der Personen aus Personengruppe 3 die Aufgabe richtig gelöst haben.

Beispielsweise wird Aufgabe 1 somit zur Aufgabengruppe II zugeordnet.

Mit der Bezeichnung der Aufgabengruppe >III (nicht etwa Aufgabengruppe IV) soll verdeutlicht werden, dass es auf Grundlage der drei Personengruppen nicht möglich ist, diese Aufgabengruppe weiter auszudifferenzieren. Es ist somit unklar, ob es sich im Falle der Gesamtpopulation beispielsweise um zwei Aufgabengruppen (etwa IV und V) handelt. Folglich kann nur bestimmt werden, dass diese Aufgaben über der Aufgabengruppe III anzusiedeln sind. Nichtsdestotrotz können mit der Aufgabengruppe >III Aussagen über die Wissensstruktur getroffen werden (siehe Abschnitt Inhaltliche Beschreibung der Fähigkeitsniveaus).

Im *fünften Schritt* wird abschließend die durchschnittliche Aufgabenschwierigkeit jeder Aufgabengruppe berechnet. Durch diese vier durchschnittliche Aufgabenschwierigkeiten werden dann die unteren Grenzen der vier Fähigkeitsniveaus I, II, III und >III definiert (Hartig 2007; Woitkowski und Riese 2017). Dieser Schritt ist im Scale-Anchoring-Verfahren von Mullis (2012) nicht vorhanden, da hier mit der Festsetzung der Personengruppen (zweiter Schritt) auch die Grenzen der Fähigkeitsniveaus festgesetzt werden. Woitkowski und Riese (2017) führen hierzu aus, dass die so identifizierten Fähigkeitsniveaus hinsichtlich ihrer Qualität und Interpretierbarkeit „stark von den (u. U. zufällig) an den Niveaugrenzen liegenden Items abhängt“ (Woitkowski und Riese 2017, S. 42). Um dieser Schwäche zu begegnen, werden die Grenzen der Fähigkeitsniveaus nicht direkt durch die Personengruppen definiert, sondern durch die durchschnittliche Aufgabenschwierigkeit innerhalb der jeweiligen Fähigkeitsniveaus.

Durch eine inhaltliche Analyse von Aufgaben können abschließend die (rein) numerischen Grenzen der Fähigkeitsniveaus inhaltlich interpretiert werden. Diese Interpretation der Niveaus geschieht anhand der Aufgaben, welche sich an der jeweiligen Niveaugrenze befinden (Beaton und Allen 1992; Hartig 2007). Mit Hilfe dieser Aufgaben kann dann das Wissen von Studierenden „beschrieben werden, deren Leistung innerhalb des Niveaus liegt“ (Hartig 2007, S. 87). Zur Festlegung, welche Aufgaben sich an einer Niveaugrenze befinden, existieren in der Literatur keine etablierten Richtwerte. Hier werden als Begrenzung der Umgebung die jeweiligen halben Standardabweichungen der Aufgabenschwierigkeiten in einem Niveau genutzt. Zum einen kann so sichergestellt werden, dass Aufgaben zwi-

schen zwei Niveaugrenzen nicht zur Beschreibung eines Niveaus herangezogen werden. Denn so wäre keine ausreichend scharfe Trennung zwischen benachbarten Niveaus mehr möglich (Hartig 2007). Zum anderen deckt das Intervall Niveaugrenze plus bzw. minus ihrer halben Standardabweichungen eine ausreichend hohe Anzahl an Items zur inhaltlichen Charakterisierung der Niveaus ab.

Im Anschluss an die Konstruktion der Fähigkeitsniveaus ist es unverzichtbar, ihre Qualität festzustellen. Woitkowski und Riese (2017) stellen hierzu drei Qualitätskriterien zusammen. So ist zunächst sicherzustellen, dass der Abstand der Niveaus (bzw. Niveaugrenzen) untereinander groß genug ist, sodass „eine Versuchsperson am unteren Rand eines Niveaus die zur Charakterisierung des nächst höheren Niveaus herangezogenen Anforderungsmerkmale tatsächlich nicht hinreichend sicher beherrscht“ (Woitkowski und Riese 2017, S. 43). Dieses *Abstandskriterium* kann hierbei mit Hilfe der Lösungswahrscheinlichkeiten überprüft werden. Demnach muss zum Beispiel eine Person mit einer Niveau-I-Fähigkeit eine ausreichend niedrige Lösungswahrscheinlichkeit für eine typische Aufgabe des Niveaus II besitzen. Die üblich tolerierten Werte reichen von einer Wahrscheinlichkeit von 20 % (Klieme et al. 2000) bis 30 % (Beaton und

Allen 1992). Darüber hinaus muss überprüft werden, ob sich die Merkmale benachbarter Niveaus nicht überlappen, bzw. sich genügend stark voneinander unterscheiden. Das *Signifikanzkriterium* kann kontrolliert werden, indem zum einen die Merkmale benachbarter Niveaus inhaltlich verglichen werden und zum anderen überprüft wird, ob sich die empirischen Aufgabenschwierigkeiten zweier benachbarter Niveaus in einem t-Test signifikant voneinander unterscheiden. Das *Idealitätskriterium* fordert schließlich, dass die inhaltlichen Anforderungen eines Fähigkeitsniveaus genauso für die darüber liegenden Niveaus gelten. So beherrscht beispielsweise ein Proband auf dem Fähigkeitsniveau II auch die Anforderungen des Fähigkeitsniveaus I. Die konstruierten Niveaus mit ihren Anforderungsmerkmalen sind also nicht disjunkt, sondern enthalten sich vielmehr. Dieses Kriterium lässt sich anhand einer inhaltlichen Analyse der definierten Niveaus überprüfen.

Abb. 2 Wright-Map für das FDW der Stichprobe mit eingezeichneten Fähigkeitsniveaus. Die Farben Grün (Kreuz), Gelb (Viereck), Rot (Dreieck) und Grau (Stern) symbolisieren die Zugehörigkeit zu den Niveaus I, II, III und >III. Die Probandengruppe, welche schwarz eingefärbt ist, liegt unterhalb des Niveaus I

Tab. 2 Verteilung der Probanden über die Fähigkeitsniveaus

Fähigkeitsniveau	Anzahl an Probanden	Mittleres Studienjahr
Unter I	3	1,67 (<i>SD</i> = 0,58)
I	44	2,43 (<i>SD</i> = 1,35)
II	104	2,86 (<i>SD</i> = 1,30)
III	49	3,29 (<i>SD</i> = 1,17)
>III	0	–
Gesamt	200	2,85 (<i>SD</i> = 1,31)

Ergebnisse

Auswertung der Niveaugrenzen

Zur Konstruktion der Niveaugrenzen im FDW wurden mit Hilfe der prozentualen Anteile richtiger Antworten gemäß dem oben beschriebenen Verfahren 51 der 56 Aufgaben in vier Aufgabengruppen (I, II, III, >III) einsortiert. Für jede Aufgabengruppe wurde im nächsten Schritt die durchschnittliche Aufgabenschwierigkeit bestimmt. Auf diese Weise können vier Niveaugrenzen identifiziert werden. Anschließend wurde mit einer Wright-Map die Einteilung der Aufgaben und Personen in die Fähigkeitsniveaus visualisiert (siehe Abb. 2). Wie erwartet streuen die Aufgaben um die jeweilige Niveaugrenze (also dem Mittelwert der Aufgabenschwierigkeiten), wobei es dennoch möglich ist, dass die Schwierigkeit einer Aufgabe im Rasch-Modell vergleichsweise hoch (bzw. niedrig) ist, während die Aufgabe einem niedrigen (bzw. hohen) Fähigkeitsniveau zugeordnet wurde. Möglich wird dies beispielsweise, wenn Personen mit niedrigerer Fähigkeit diese Aufgabe wahrscheinlicher korrekt lösen als Personen mittlerer Fähigkeit (Aufgabe 27) oder wenn die Lösungswahrscheinlichkeit der fähigsten Personen gerade knapp oberhalb der definierten 55 % liegt (Aufgabe 35). Die Personenverteilung entlang der Fähigkeitsniveaus (siehe Tab. 2) zeigt, dass sich die Mehrzahl der Studierenden (52 %) innerhalb des Niveaus II befindet. Außerdem ist zu erkennen, dass das durchschnittliche Studienjahr mit den Fähigkeitsniveaus steigt. Höhere Niveaus wurden also tendenziell von Studierenden höherer Studienjahre erreicht.

Mit dem Abstandskriterium wird überprüft, ob eine für das Fähigkeitsniveau *typische* Person tatsächlich eine *durchschnittliche* Aufgabe des nächsthöheren Niveaus un-

sicher genug löst. Dabei entspricht die *typische* Person, bzw. *durchschnittliche* Aufgabe eines Fähigkeitsniveaus genau der mittleren Aufgabenschwierigkeit des Niveaus und damit der numerischen Niveaugrenze. Tab. 3 zeigt die Lösungswahrscheinlichkeit zum jeweils nächsthöheren Fähigkeitsniveau. Zu erkennen ist, dass die Lösungswahrscheinlichkeiten zwischen Niveau I und II und zwischen Niveau III und >III unter 30 % liegen. Demnach liegen die Fähigkeitsniveaus hinsichtlich der Personen- und Aufgabenparameter hinreichend weit auseinander. Zwischen dem Niveau II und III liegt eine Lösungswahrscheinlichkeit von 30,8 %, was somit noch toleriert werden kann.

Neben dem Abstandskriterium wurde darüber hinaus geprüft, ob sich die dem Fähigkeitsniveau zugeordneten Aufgaben bezüglich ihrer Schwierigkeit signifikant unterscheiden. Die in der Tab. 3 aufgeführten t-Werte zeigen, dass sich die Aufgabenschwierigkeiten zwischen den jeweiligen Niveaus signifikant unterscheiden. Somit enthalten die jeweiligen Niveaus tatsächlich signifikant unterschiedliche Aufgabenschwierigkeiten.

Inhaltliche Beschreibung der Fähigkeitsniveaus

Die konstruierten (numerischen) Niveaugrenzen werden anschließend genutzt, um inhaltliche Aussagen über die vorhandenen Wissensstrukturen zu tätigen. Hierfür wurden jene Aufgaben ausgewählt, die an, bzw. in der Nähe einer Niveaugrenze liegen (Hartig 2007). Anschließend wird für diese Aufgaben analysiert, welches FDW prototypisch vorhanden sein muss, um sie korrekt zu lösen. Der Inhalt dieser Aufgaben sowie die Anforderungen an das FDW, diese Aufgaben richtig zu lösen, liefert dann die inhaltliche Beschreibung des Fähigkeitsniveaus. Die *Nähe* an einer Niveaugrenze lässt sich durch die entstandenen Intervalle (für das Fähigkeitsniveau I [–1,66; –1,13], für das Fähigkeitsniveau II [–0,53; –0,40], für das Fähigkeitsniveau III [0,25; 0,57], für das Fähigkeitsniveau >III [1,10; 1,96]) charakterisieren (siehe auch Abb. 2). Die Anzahl der identifizierten Aufgaben pro Facette und Niveau sind in Tab. 4 abgebildet, die daraus resultierende inhaltliche Beschreibung liefern die in Tab. 5 (im Anhang) formulierten Fähigkeitsniveaus. Im Anhang finden sich exemplarisch niveaucharakterisierende Aufgaben.

Tab. 3 Zusammenfassung der Fähigkeitsniveaus zur Überprüfung der Qualitätskriterien

Fähigkeitsniveau	Aufgabenanzahl	Mittlere Aufgabenschwierigkeit/ Niveaugrenze	t-Wert	Lösewahrscheinlichkeit des nächsthöheren Niveaus
I	15	–1,40	6,32*** (zwischen I und II)	26,9 %
II	11	–0,40	6,66*** (zwischen II und III)	30,8 %
III	11	0,41	4,75*** (zwischen III und >III)	24,6 %
>III	16	1,53	–	–

*** $p < 0,001$

Tab. 4 Anzahl von Aufgaben pro Facette und Fähigkeitsniveau, die zur inhaltlichen Beschreibung genutzt werden

Fähigkeitsniveau	Vorstellungen der Schülerinnen und Schüler	Instruktionsstrategien	Curriculum	Assessment
I	2	1	4	0
II	2	1	2	0
III	1	3	0	2
>III	2	1	2	2

Für die Prüfung des Idealitätskriteriums ist es sinnvoll, die Fähigkeitsniveaus facettenweise zu diskutieren. Betrachtet man also die Facette zum Wissen über **Vorstellungen der Schülerinnen und Schüler**, zeigt sich auf dem ersten Niveau, dass Studierende eher mit allgemeinphysikalischen Vorstellungen vertraut sind. So sind Probanden beispielsweise in der Lage, die wissenschaftliche (und damit adäquate) Eigenschaft eines physikalischen Modells gegen die prominente Vorstellung der Schülerinnen und Schülern, dass der „Wahrheitsgehalt“ das wichtigste Merkmal eines physikalischen Modells ist, abzugrenzen. Außerdem sind Studierende mit typischen Fehlvorstellungen der Mechanik vertraut (siehe Anhang, Aufgabe 12). Auf dem zweiten Niveau sind Studierende mit weiteren themenspezifischen Vorstellungen (wie zum Beispiel der Elektrizitätslehre) von Schülerinnen und Schülern vertraut und in der Lage, durch Experimente die Vorstellungen der Schülerinnen und Schüler bzgl. naturwissenschaftlicher Arbeitsweisen zu fördern (etwa: „Für das Lernen über naturwissenschaftliche Arbeitsweisen muss das Nachvollziehen des experimentellen Aufbaus, die Durchführung und Auswertung eines Experimentes fokussiert werden“). Studierende des dritten Fähigkeitsniveaus können potentielle Fehlvorstellungen in Repräsentationen (bspw. aus Graphen und Abbildungen aus Schulbüchern) erkennen. Zwischen dem Niveau I und II zeigt sich ein sukzessiver Wissensaufbau, da Studierende auf dem zweiten Niveau ihre allgemeinphysikalischen Vorstellungen (Fähigkeitsniveau I) nun auf konkrete Fachinhalte beziehen können. Auch im Vergleich zwischen Niveau II und Niveau III zeigt sich der aufbauende Charakter. Studierenden des zweiten Niveaus ist es vorerst nur möglich, typische Fehlvorstellung zu *benennen*; sie reproduzieren gewissermaßen die Lerninhalte, die sie in den Fachdidaktikveranstaltungen präsentiert bekommen haben. Auf dem dritten Fähigkeitsniveau jedoch ist es ihnen nun möglich, *potentielle* Fehlvorstellungen zu *erkennen*. Dies geht über eine reine Wiederholung der Lerninhalte hinaus, da die Studierenden Abbildungen selbstständig mit Hilfe der gelernten Inhalte analysieren und bewerten müssen. Für diese Facette ist das Idealitätskriterium somit erfüllt.

Auf der Ebene der **Instruktionsstrategien** zeigt sich ein Zuwachs des Wissens ebenfalls deutlich. Ist den Studierenden auf dem ersten Niveau lediglich der entdeckende Unter-

richt bekannt, so wächst ihr Wissen über den sachstrukturorientierten bis hin zum kontextorientierten Unterricht von Fähigkeitsniveau zu Fähigkeitsniveau an. Die Studierenden auf dem dritten Niveau kennen demnach einen Grundstock an Instruktionsstrategien. Für diese Annahme spricht auch, dass Probanden des dritten Niveaus in der Lage sind, die in Schulbüchern verwendeten Instruktionsstrategien zu erläutern sowie verschiedene Instruktionsstrategien gegeneinander abzuwägen (hier, um eine holistische Vorstellung bei optischen Abbildungen zu überwinden). Auf dem Fähigkeitsniveau >III ist es Studierenden anschließend möglich, ihr Repertoire an Instruktionsstrategien zu *nutzen*, um sich für eine Unterrichtsstrategie (hier im Sinne eines Konzeptwechsels) zu entscheiden (siehe Anhang, Aufgabe 17). Studierende dieses Niveaus besitzen also nicht nur ein Wissen *von* verschiedenen Instruktionsstrategien, sondern auch ein reflektiertes Wissen *über* deren (geeignete) Verwendung. Das Idealitätskriterium ist also auch hier erfüllt.

Auf dem Fähigkeitsniveau I zum Wissen über das **Curriculum** besitzen Studierende ein partielles Wissen über die Bedeutung sozialer und kultureller Einflüsse der Wissenschaftsgeschichte und können diese verschiedenen Aspekte hinsichtlich ihrer Bedeutung für den Unterricht einordnen (z.B., dass auch die Thematisierung von Wissenschaftsgeschichte nötig ist, um die gesellschaftliche Bedeutung von Naturwissenschaften zu diskutieren). Zudem ist es den Studierenden möglich, Aufgaben, zum Beispiel zum elektrischen Widerstand, gemäß den Bildungsstandards in die dafür vorgesehenen Anforderungsbereiche 1 und 2 einzuordnen, während schließlich das Niveau II eine Einordnung von Aufgaben innerhalb aller drei Anforderungsbereiche umfasst. Außerdem sind Studierende des Niveaus II in der Lage, physikalische Inhalte gemäß des Spiralcurriculums anzuordnen (siehe Anhang, Aufgabe 25). Erwähnenswert ist, dass Studierende im Bereich des Curriculums erst auf dem Niveau >III in der Lage sind, eigene Unterrichtsentscheidungen auf Grundlage des Curriculums zu reflektieren. So können Studierende motivationale und inhaltliche Ziele eines außerschulischen Lernorts erst hier darlegen. Somit unterliegt die Einteilung gemäß der Fähigkeitsniveaus dem Idealitätskriterium.

Die Facette des **Assessments** ist in den Niveaus I und II nicht vorhanden. Studierende des dritten Niveaus können Aufgaben den Kompetenzbereichen der Bildungsstandards zuordnen (siehe Anhang, Aufgabe 38), Probanden des Fähigkeitsniveaus >III sind in der Lage, Aspekte naturwissenschaftlicher Kompetenz (z.B. die Variablenkontrollstrategie oder das Planen von Experimenten) zu erkennen, die durch bestimmte Aufgaben erfasst werden können. Während Studierende auf dem Niveau III also passende Aufgaben im Sinne der Kompetenzbereiche auswählen, sind Studierende des Niveaus >III in der Lage, im Sinne naturwissenschaftlicher Kompetenz (hier die Variablenkontrollstrategie) adä-

quate Aufgaben zur Leistungsmessung auszuwählen. Die Fähigkeit, Aufgaben auf Grundlage naturwissenschaftlicher Merkmale zu evaluieren, scheint demnach eher Studierenden der höheren Niveaus zu gelingen. Die formulierten Fähigkeitsniveaus im Assessment verhalten sich also erwartungsgemäß und unterliegen dem Idealitätskriterium.

Um zusätzlich zu überprüfen, ob die so charakterisierten Fähigkeitsniveaus nicht nur Aufgabenformate abbilden, sondern tatsächlich die vorhandene Wissensstruktur der Studierenden, wird eine lineare Regression von Aufgabenformat (offen vs. geschlossen) auf das zugeordnete Niveau (I, II, III, >III) berechnet. Die Regression zeigt dabei keinen signifikanten Einfluss des Aufgabenformates ($b=0,89$; $p=0,09$) auf das Fähigkeitsniveau. Somit ist davon auszugehen, dass die Fähigkeitsniveaus kein Artefakt der Aufgabenformate darstellen.

Das FDW von Studierenden des ersten Studienjahres (iFDW)

Zur Untersuchung des postulierten iFDW ist es nötig, jene Probanden aus der Stichprobe zu wählen, die in ihrem bisherigen Studium (nahezu) keine universitären fachdidaktischen Veranstaltungen besucht haben. Zu erwähnen ist, dass die Erhebung des FDW im Laufe des Sommersemesters stattfand, somit ist nicht auszuschließen, dass Studierende des ersten Jahres bereits an ersten fachdidaktischen Lehrveranstaltungen teilgenommen haben. Aus der Wright-Map mit 90-prozentiger Lösungswahrscheinlichkeit kann nun abgelesen werden, welche jeweiligen Personenfähigkeiten nötig waren, um entsprechende Aufgaben mit einer 90-prozentigen Wahrscheinlichkeit korrekt zu lösen. Durch die durchschnittliche Fähigkeit der Teilstichprobe konnte dann festgestellt werden, welche Aufgaben von einem *typischen* (d.h. durchschnittlichen) Studierenden des ersten Studienjahres mit hoher Wahrscheinlichkeit (nämlich $\geq 90\%$) richtig gelöst werden.

Abb. 3 im Anhang zeigt die berechnete Wright-Map. Die durchschnittliche Personenfähigkeit der Teilstichprobe beträgt 0,03 Logits ($SD=0,52$). Die mittlere der drei horizontalen Linien repräsentiert die durchschnittliche Fähigkeit eines Probanden aus der Teilstichprobe, die Linien darunter bzw. darüber zeigen die Standardabweichung der mittleren Personenfähigkeit. Zu den dichotomen Aufgaben sind darüber hinaus die Standardabweichungen in Form von Fehlerbalken eingezeichnet¹. Der Abb. 3 ist zu entnehmen, dass 12 Aufgaben innerhalb bzw. unterhalb der Standardabweichung eines durchschnittlichen Studierenden des ersten

Jahres zu lokalisieren sind². Dementsprechend wurden diese Aufgaben ausreichend sicher von Studierenden gelöst, die bisher keine oder nur begrenzt universitären Lerngelegenheiten für das FDW hatten. Daher kommen diese Aufgaben für ein iFDW in Frage.

Eine genauere Betrachtung der intuitiven Aufgaben zeigte, dass es sich bei zehn der zwölf Aufgaben um polytome Aufgaben handelt, genauer um Zuordnungsaufgaben. Das ist insofern relevant, als dass durch die Aufgabe alleine nicht deutlich wird, *welche* Elemente (innerhalb einer Aufgabe) tatsächlich intuitiv richtig zugeordnet wurden. Da dies für eine Beschreibung des iFDW jedoch durchaus relevant ist, wurden in einem nächsten Schritt die Zuordnungen der Teilstichprobe analysiert und insbesondere überprüft, welche richtigen Antwortmöglichkeiten vermehrt ausgewählt wurden. Es zeigte sich, dass in vier der fünf polytomen Items die intuitiven Zuordnungen eindeutig erkennbar waren. Beispielsweise konnten in der (Zuordnungs-)Aufgabe 29 folgende Arbeitsaufträge zum elektrischen Widerstand von 28 bzw. 25 Studierenden dem richtigen Anforderungsbereich zugeordnet werden:

Geben Sie an, wie sich der elektrische Widerstand eines ohmschen Bauteils aus Strom und Spannung berechnen lässt. (Anforderungsbereich: Wiedergabe, 28 der 30 Studierenden)

Erklären Sie mit Hilfe der angestellten Betrachtung die Funktionsweise des elektrischen Heizelementes in einem Wasserkocher. (Anforderungsbereich: Transfer, 25 der 30 Studierenden)

Die restlichen vier Zuordnungen der Aufgabe 29 wurden von 3, 13, 21, bzw. 14 Probanden richtig beantwortet, kommen also als Intuition weniger in Frage. Darüber hinaus zeigte das Zuordnungsmuster der polytomen Aufgabe 10 keine eindeutig präferierte Zuordnung; keine der vier richtigen Zuordnungen wurde von mehr als der Hälfte der Probanden gewählt. Es werden im Folgenden also nur die vier polytomen und übrigen zwei dichotomen Aufgaben berücksichtigt.

Die mittlere Aufgabenschwierigkeit der ausgewählten Aufgaben für Studienanfänger beträgt $-1,61$ Logits ($SD=0,43$). Versteht man diese mittlere Aufgabenschwierigkeit als untere Grenze für ein grundlegendes Fähigkeitsniveau, etwa Niveau 0, so ließe sich dieses unter das bestehenden Fähigkeitsniveau I einsortieren. Im Vergleich zum Niveau I jedoch fällt auf, dass alle 11 Aufgaben des Niveaus 0 im Niveau I vertreten sind. Somit ist das

¹ Die polytomen Aufgaben besitzen keine Fehlerbalken, da diese für Thresholds i. A. nicht ausgegeben werden.

² Aufgabe 10.1 befindet sich streng genommen nicht mehr im Fehlerbereich der mittleren Personenfähigkeit. Jedoch kann an den Fehlerbalken der dichotomen Aufgaben abgeschätzt werden, dass sich der Fehler der Aufgabe 10.1 mit hoher Sicherheit noch im Bereich der mittleren Personenfähigkeit befindet.

Fähigkeitsniveau 0 eine echte Teilmenge des Niveaus I. Weiter sind immerhin rund 70 % der Aufgaben aus Niveau I dem Fähigkeitsniveau 0 zugeordnet worden. Daraus kann geschlossen werden, dass mit dem Fähigkeitsniveau I Wissens Elemente des iFDW erfasst werden; oder anderes formuliert: Das Fähigkeitsniveau I kennzeichnet das iFDW.

Diskussion

Die erfolgreiche Unterrichtsgestaltung und damit Anregung schulischen Lernens erfordert ein angemessenes FDW (Keller et al. 2017). Diese besondere Relevanz des FDW motiviert, eine inhaltliche Charakterisierung der Wissensstrukturen von (angehenden) Lehrkräften vornehmen zu können. Ein Niveaumodell könnte eine kriterienorientierte Interpretation des FDW angehender Lehrkräfte ermöglichen (Klieme et al. 2007). Die Entwicklung eines solchen Modells war Ziel des vorliegenden Artikels. Die aus einer Querschnitterhebung gewonnenen Daten konnten nach einer Rasch-Analyse (Rasch 1960) durch das Scale-Anchoring-Verfahren so interpretiert werden, dass sich vier Fähigkeitsniveaus entlang des FDW angehender Physiklehrkräfte identifizieren ließen. Darüber hinaus konnte eine Gruppe von Aufgaben herausgestellt werden, die bereits von Probanden des ersten Studienjahres mit hoher Wahrscheinlichkeit korrekt beantwortet werden. Dies lässt darauf schließen, dass Studierende bereits zu Studienbeginn über ein nahezu intuitives Verständnis von Fachdidaktik verfügen. Eine genauere Betrachtung dieses iFDW zeigte, dass mit dem ersten Fähigkeitsniveau im FDW Wissens Elemente des iFDW erfasst werden.

In Bezug auf die erste Fragestellung, wie sich die Fähigkeitsniveaus hinsichtlich der vier Facetten des FDW beschreiben lassen, lässt sich nach dem durchgeführten Scale-Anchoring-Verfahren zunächst festhalten, dass die Fähigkeitsniveaus die tatsächlich vorhandenen Wissensstrukturen und nicht etwa das Aufgabenformat abbilden. Außerdem ist zu erkennen, dass höhere Niveaus tendenziell von Studierenden höherer Studienjahre erreicht werden. Und obwohl den Fähigkeitsniveaus kein Stufenmodell (im Sinne einer Entwicklung) zugrunde liegt, kann dies als erstes Indiz dafür gewertet werden, dass die konstruierten Niveaus auch eine Entwicklung im FDW Lehramtsstudierender beschreiben. Mit Blick auf die Facetten ist festzustellen, dass das Reproduzieren von Lerninhalten generell den niedrigeren, reflexive Fähigkeiten eher den höheren Niveaus zugeordnet werden können. Diese Beobachtung steht im Einklang mit gängigen Modellen der *hierarchischen Komplexität*. So schlagen hierfür beispielsweise Neumann et al. (2007) in ihrer Modellierung zur physikalischen Kompetenz eine (hierarchische) Differenzierung zwischen Fakten, Zusammenhängen und übergeordneten Konzepten vor. Bern-

holt et al. (2009) unterscheiden (in der Chemie) zwischen unreflektiertem Erfahrungswissen, Fakten, Prozessbeschreibungen, Lineare Kausalität und Multivariate Interdependenz (für weitere Kompetenzmodelle im Fachwissen siehe z.B. Bernholt et al. 2009). Zwar unterscheiden sich die beiden genannten Modelle hinsichtlich ihrer Komplexitätsschemata, trotzdem lässt sich erkennen, dass das *Begründen von Zusammenhängen* stets eine höhere kognitive Anforderung darstellt als das *Nennen isolierter Fakten*. Die inhaltliche Analyse der hier dargelegten Fähigkeitsniveaus deckt sich demnach mit Modellen, denen (zumindest im FW) eine hierarchische Komplexität zu Grunde liegt. Im Speziellen zeigt sich, dass reflexive Fähigkeiten, um Instruktionsstrategien zu bewerten oder Aktivitäten motivational zu begründen, erst in den höheren Niveaus auftreten. Dies deckt sich mit den Befunden von Schneider und Plasman (2011). Zudem gelingen reflektierte Begründungen in den Facetten Curriculum und Assessment erst auf höheren Niveaus, besonders im Vergleich zu den Facetten Vorstellungen der Schülerinnen und Schüler und Instruktionsstrategien. Grund dafür könnte sein, dass Studierende im Zeitraum ihrer universitären Ausbildung seltener auf Grundlage eines Curriculums reflektieren. Dass Curriculum und Assessment im Allgemeinen also als weniger vertraute Facetten angesehen werden können, steht auch hier im Einklang mit den Resultaten von Schneider und Plasman (2011). Diese geben jedoch zusätzlich den Hinweis, dass Lehrkräfte mit wachsender Berufserfahrung tendenziell fähiger in der Facette Curriculum werden (Schneider und Plasman 2011). Somit ist zu erwarten, dass insbesondere in der zweiten Phase der Lehrkräfteausbildung dieses Wissen weiter ausgeprägt wird. Außerdem kann aus diesen Ergebnissen abgeleitet werden, dass es expliziter Lerngelegenheiten bedarf, um adäquates Wissen im Curriculum und Assessment zu konstruieren und auf unterrichtliche Entscheidungsprozesse zu beziehen. Dies ist gerade wichtig, weil Lehrkräfte insbesondere durch ein adäquates Assessment in der Lage sind, den Wissens- und Lernstand der Schülerinnen und Schüler einzuschätzen und darauf aufbauend ihr Repertoire an Instruktionsstrategien anzupassen (Park und Chen 2012).

Bezüglich der zweiten Forschungsfrage zur Existenz eines iFDW kann festgestellt werden, dass Studierende des ersten Studienjahres tatsächlich über ein nicht triviales FDW verfügen, ohne bzw. nur erste formale fachdidaktische Lerngelegenheiten wahrgenommen zu haben. Dieses Ergebnis deckt sich prinzipiell mit den Beobachtung von Riese et al. (2017). Woraus Studierende dieses iFDW jedoch beziehen, kann hier nicht beantwortet werden. Die Vermutung von Lortie (1975) und Grossman (1990), besonders die eigene Schulzeit als Quelle für iFDW anzusehen, scheint jedoch plausibel. Für zukünftige Forschung im Bereich des iFDW ist es erstrebenswert, ein Testinstrument zu entwerfen, welches in der Lage ist, das iFDW differenziert

und detailliert direkt zu Studienbeginn zu erfassen. Auf diese Weise ließe sich das iFDW gegenüber des Fähigkeitsniveaus II schärfer abgrenzen und adäquater beschreiben, um so die Grundlage für einen fließenderen Übergang von der Schule zur Universität zu schaffen.

Bei der Beschreibung der Fähigkeitsniveaus ist zu beachten, dass eine vergleichsweise geringe Anzahl an Aufgaben pro Facette genutzt wurde, um die Niveaus inhaltlich zu interpretieren. Dies liegt darin begründet, dass nur wenige Aufgaben *einer* Facette an der Niveaugrenze liegen. Es besteht somit die Gefahr, nur einen Ausschnitt der tatsächlichen Fähigkeiten jeden Niveaus und Facette erfasst zu haben. Die hier aufgeführten Beschreibungen der Fähigkeitsniveaus sollten folglich eher als Teil einer (möglicherweise umfassenderen) Niveaubeschreibung verstanden werden. Außerdem muss beachtet werden, dass die Aufgaben im Testinstrument größtenteils einen deutlichen Bezug zu Fachinhalten aufwiesen, die Niveaus jedoch möglichst generalisiert, frei von Fachinhalten, formuliert wurden. Eine weitere Limitation liegt im Scale-Anchoring-Verfahren und der Wahl der ursprünglichen Fähigkeitsabschnitte (zweiter Schritt). Zur Untersuchung, inwieweit die gewählten Abschnitte auf der Fähigkeitsskala arbiträr erfolgten, wurden hierfür die drei Abschnitte, bzw. Personengruppen nach links, bzw. rechts linear transformiert (also 375–425, 475–525, 575–625 Fähigkeitspunkte, bzw. 425–475, 525–575, 625–675 Fähigkeitspunkte) und anschließend die Fähigkeitsniveaus analysiert. Dabei zeigt sich für dichotome Aufgaben, dass die lineare Transformation der drei Personengruppen nach links keinen Einfluss auf den Inhalt der Niveaus hat. Die Transformation in Richtung höherer Fähigkeit würde jedoch die inhaltliche Beschreibung der Niveaus I und III verändern, zeigt gleichermaßen aber eine weniger homogene Abbildung der Stichprobe. Somit liefern die hier dargestellten Fähigkeitsniveaus eine erste Ausgangslage für die weitere Untersuchung der Qualität des FDW (angehender) Lehrkräfte.

Außerdem wäre denkbar, dass die erfassten Aufgaben für das iFDW weniger durch informelle (fachspezifische) Lerngelegenheiten im FDW (z.B. die eigene Schulzeit) lösbar sind, sondern vielmehr mit einem Alltagsverständnis vom Lernen allgemein. Jedoch zeigen beispielsweise die Aufgaben 26 und 29, dass Studierende die Vermittlung einer naturwissenschaftlichen Grundbildung im Physikunterricht zutreffend beschreiben können (z.B., dass Wissenschaftsgeschichte fachliches Lernen unterstützen kann, da Schülerinnen und Schüler anhand historischer Konzepte ihrer eigenen Vorstellungen bewusst werden) und physikalische Aufgaben gemäß Bildungsstandards in Anforderungsbereiche zuordnen. Da es sich hier um curriculare Anforderungen und Kompetenzen handelt, die Schülerinnen und Schüler im Physikunterricht erwerben sollen (also FDW) liegt es nahe, dass dieses Wissen im eigenen Physikunterricht erworben

wurde, also über generelles Alltagsverständnis vom Lernen hinausgeht. Dennoch lässt sich nicht mit Sicherheit sagen, ob das hier festgestellte iFDW tatsächlich jenes Wissen erfasst, welches der eigenen Schulzeit entstammt. Dies sollte zukünftig mit einem speziellen dafür konstruierten Testinstrument überprüft werden. Zudem sind die Ergebnisse zum iFDW dahingehend limitiert, dass die Erhebung des FDW eher am Ende des Sommersemesters stattfand. Somit kann nicht völlig ausgeschlossen werden, dass Studierende des ersten Jahres nicht doch formale Lerngelegenheiten im FDW wahrgenommen haben.

Prospektiv kann das hier konstruierte Niveaumodell dafür genutzt werden, die Ausbildung angehender Lehrkräfte an den Universitäten weiter zu verbessern. Einerseits können Curricula und Ausbildungspläne des Lehramtsstudiums künftig an das Niveaumodell angelehnt werden, um eine konsistente fachdidaktische Ausbildung zu sichern. So sollte die fachdidaktische Ausbildung auf der einen Seite alle Facetten simultan fördern, um Kohärenz im FDW zu gewährleisten (Park und Chen 2012). Auf der anderen Seite jedoch legen die hier beschriebenen Ergebnisse nahe, dem Wissen über Curriculum und Assessment vermehrt Aufmerksamkeit in den universitären Lerngelegenheiten zu schenken. Universitäten sollten demnach Lerngelegenheiten anbieten, die ein Erschließen der höheren Niveaus ermöglichen, bzw. unterstützen. Darüber hinaus können die Fähigkeitsniveaus genutzt werden, um Lehrveranstaltungen auf ihre Lernwirksamkeit hin zu überprüfen und gegebenenfalls anzupassen, um die Veranstaltungen im fachdidaktischen Bereich weiter zu optimieren und damit die Qualität der Lehrkräfteausbildung zu verbessern.

Anhang

Übersicht der Fähigkeitsniveaus

Tab. 5 Inhaltliche Charakterisierung der verschiedenen Fähigkeitsniveaus nach Facetten sortiert

Fähigkeitsniveau I (Personenfähigkeit θ im Bereich $-1,31 \leq \theta < -0,35$; 27 % der Probanden)

Vorstellungen der Schülerinnen und Schüler

Studierende grenzen die Beschreibung eines Modells gegen die gängigen Vorstellungen von Schülerinnen und Schülern ab, indem sie Modelle nicht als richtig oder falsch klassifizieren, sondern als „passend“ für die Beschreibung eines Phänomens. (Aufgabe 11.2)

Studierende erläutern typische Fehlvorstellungen im Bereich der Newtonsche Axiome. (Aufgabe 12)

Instruktionsstrategien

Studierende kennen typische Merkmale des entdeckenden Physikunterrichts. (Aufgabe 20.1)

Curriculum

Studierende sind in der Lage, die Rolle der Wissenschaftsgeschichte für Physikunterricht zu bewerten. (Aufgabe 26.3 und 26.4)

Studierende können bei operationalisierten Aufgaben gemäß der Bildungsstandards zwischen den Anforderungsbereichen „Wiedergeben“ und „Anwenden“ unterscheiden. (Aufgabe 29.1 und 29.2)

Fähigkeitsniveau II (Personenfähigkeit θ im Bereich $-0,35 \leq \theta < 0,46$; 52 % der Probanden)

Vorstellungen der Schülerinnen und Schüler

Studierende benennen typische und untypische Fehlvorstellungen im Bereich der Elektrizitätslehre. (Aufgabe 1)

Studierende sind in der Lage, die Vorstellungen der Schülerinnen und Schüler zu naturwissenschaftliche Arbeitsweisen durch ein Experiment zu fördern. (Aufgabe 10.1)

Instruktionsstrategien

Studierende kennen typische Merkmale des sachstrukturorientierten und entdeckenden Physikunterrichts. (Aufgabe 20.2)

Curriculum

Studierende benennen Inhalte (bspw. aus der Elektrizitätslehre), die gemäß des Spiralcurriculums angeordnet werden können. (Aufgabe 25)

Studierende können bei operationalisierten Aufgaben gemäß der Bildungsstandards zwischen den drei Anforderungsbereichen unterscheiden. (Aufgabe 29.3)

Fähigkeitsniveau III (Personenfähigkeit θ im Bereich $0,46 \leq \theta < 1,54$; 19 % der Probanden)

Vorstellungen der Schülerinnen und Schüler

Studierende sind in der Lage, basale Fehlvorstellungen zu erkennen, die durch Abbildungen hervorgerufen werden können. (Aufgabe 7.1)

Instruktionsstrategien

Studierende können die in Physikschulbüchern verwendete Instruktionsstrategie erläutern. (Aufgabe 13)

Studierende kennen typische Merkmale des sachstrukturorientierten, kontextorientierten und entdeckenden Physikunterrichts. (Aufgabe 20.3)

Studierende können ihr Wissen über verschiedene Instruktionsstrategien anwenden, indem sie diese zum Erreichen eines spezifischen Lernzieles gegeneinander abwägen. (Aufgabe 21)

Assessment

Studierende können Aufgaben den Kompetenzbereichen der Bildungsstandards zuordnen. (Aufgabe 38)

Fähigkeitsniveau > III (Personenfähigkeit θ im Bereich $\theta \geq 1,54$; 0 % der Probanden)

Vorstellungen der Schülerinnen und Schüler

Studierende sind in der Lage, mögliche Quellen für Fehlvorstellungen zu identifizieren, die durch Abbildungen hervorgerufen werden können. (Aufgabe 7.2)

Studierende sind in der Lage, diverse Vorstellungen von Schülerinnen und Schülern zu physikalischen Experimenten (bspw. über naturwissenschaftlichen Arbeitsweise oder die Gewinnung von Erkenntnissen über die Natur der Naturwissenschaften) durch Experimente zu fördern. (Aufgabe 10.4)

Instruktionsstrategien

Studierende sind in der Lage, basierend auf Äußerungen von Schülerinnen und Schülern Unterrichtssituationen zum Konzeptwechsel des Kraftbegriffes zu schaffen. (Aufgabe 17)

Curriculum

Studierende sind in der Lage, außercurriculare Aktivitäten zu begründen (bspw. außerschulische Lernorte hinsichtlich inhaltlicher und motivationaler Ziele). (Aufgabe 30 und 31)

Assessment

Studierende sind in der Lage, Validität in Bezug auf eine Klassenarbeit in Physik zu erläutern. (Aufgabe 36)

Studierende erkennen Aspekte naturwissenschaftlicher Kompetenz, die durch bestimmte Aufgaben erfasst werden können. (Aufgabe 37)

Beispiele für niveaucharakterisierende Aufgaben

Aufgabe 12

Fähigkeitsniveau I – Vorstellungen der Schülerinnen und Schüler

Schülerinnen und Schülern fällt es oft schwer, die Newtonschen Axiome zur Lösung konkreter Aufgaben anzuwenden.

Betrachten Sie die folgende Situation: Ein kleiner Körper der Masse m bewegt sich reibungsfrei und mit konstanter Geschwindigkeit $\vec{v}$ nach rechts. Auf den Körper wirken dabei drei Kräfte. Zwei davon sind eingezeichnet. Sie bitten die Schülerinnen und Schüler, die dritte Kraft einzuzichnen.

Welche physikalisch falsche Antwort würden Sie von den Schülerinnen und Schülern erwarten?

Mögliche korrekte Antworten:

- [Kraftpfeil in Bewegungsrichtung] – Es muss eine Kraft für die Bewegung verantwortlich sein.
- [Kraftpfeil als Summe von F_1 und F_2] – Reflexartiges Zeichnen eines Kräfteparallelogramms.
- [Kraftpfeil als Verlängerung von F_1] – Summe aller Kräfte muss in Bewegungsrichtung zeigen.

Aufgabe 25

Fähigkeitsniveau II – Curriculum

Mit dem Spiralcurriculum bezeichnet man die wiederkehrende, aber tiefergehende Behandlung von Inhalten in späteren Jahrgangsstufen. Zum Beispiel wird die Elektrizitätslehre in der Grundschule, der Mittelstufe und der Oberstufe behandelt.

Nennen Sie Inhalte der Elektrizitätslehre aus den drei verschiedenen Jahrgangsstufen (z.B. Klasse 4, 8 und 12), die jeweils im Sinne eines Spiralcurriculums nacheinander angeordnet sind.

Auswahl möglicher korrekter Antworten:

- 4. Klasse: Leiter & Nichtleiter, 8. Klasse: Ohmsches Gesetz, 12. Klasse: Bändermodell
- 4. Klasse: einfacher Stromkreis, 8. Klasse: verzweigter Stromkreis, 12. Klasse: Schwingkreis
- 4. Klasse: Pole von Magneten, 8. Klasse: Magnetfeldlinien, 12. Klasse: magnetisches Feld / Elektromagnetismus

Aufgabe 38

Fähigkeitsniveau III – Assessment

Betrachten Sie zu Lösung der untenstehenden Frage die folgende Aufgabe:

Eine Metallscheibe mit einem Loch in der Mitte wird aufgeheizt und dehnt sich dadurch aus.

Wird das Loch beim Aufheizen größer, kleiner oder ändert sich seine Größe nicht?

Welchem Kompetenzbereich der Bildungsstandards ist diese Aufgabe zugeordnet?

Fachwissen	☒
Erkenntnisgewinnung	☐
Kommunikation	☐
Bewertung	☐

Aufgabe 17

Fähigkeitsniveau >III – Instruktionsstrategien

Im Physikunterricht:

Aylin: „Klar brauche ich Kraft, wenn ich Fahrrad fahre. Ich muss doch immer in die Pedale treten, um nicht langsamer zu werden.“

Michelle: „Irgendwie braucht es doch einen Grund, dass man langsamer wird. Ich glaube, das liegt an der Reibung.“

Frau F. ist die Lehrerin und will in das Gespräch eingreifen. Sie möchte einen Konzeptwechsel anbahnen, um die Entwicklung einer adäquaten Vorstellung des Kraftbegriffs zu fördern. Wie müsste sie am ehesten vorgehen, wenn sie dabei die Anknüpfungsstrategie verfolgen will?

- Sie müsste die geäußerte Idee aufgreifen, dass auf das Fahrrad Kraft übertragen wird. Damit kann sie den Energieerhaltungssatz motivieren. ☐
- Sie müsste die geäußerte Idee aufgreifen, dass die Reibung die Tretkraft behindert. Damit kann sie das dritte Newtonsche Axiom („Actio gleich Reactio“) motivieren. ☐
- Sie müsste die geäußerte Idee aufgreifen, dass jede Änderung eines Zustandes eine Ursache benötigt. Damit kann sie das zweite Newtonsche Axiom ($F = dp/dt$) motivieren. ☒
- Sie müsste ein Experiment durchführen, das einen kognitiven Konflikt hervorruft. Damit kann sie den Kraftbegriff physikalisch korrekt motivieren. ☐

Wright-Map der Studienanfänger

Abb. 3 Wright-Map mit 90-prozentiger Lösungswahrscheinlichkeit der Teilstichprobe (ersten Studienjahr). Die horizontalen Linien entsprechen der durchschnittlichen Fähigkeit $\pm$ SD eines Studierenden des ersten Studienjahres. Polytome Aufgaben besitzen keine Fehlerbalken

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Literatur

- Abell, S.K. (2007). Research on science teacher knowledge. In S.K. Abell & N.G. Lederman (Hrsg.), *Handbook of research on science education* (S. 1105–1149). Mahwah: Erlbaum.
- Ball, D.L., Lubienski, S., & Mewborn, D. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Hrsg.), *Handbook of research on teaching*. Washington, DC: American Educational Research Association.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- 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.
- Beaton, A.E., & Allen, N.L. (1992). Interpreting scales through scale anchoring. *Journal of Educational Statistics*, 17(2), 191–204.
- Berliner, D.C. (1988). *The development of expertise in pedagogy*. New Orleans: American Assoc. of Colleges for Teacher Education.
- Bernholt, S., Parchmann, I., & Commons, M.L. (2009). Kompetenzmodellierung zwischen Forschung und Unterrichtspraxis. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 219–245.
- Bond, T.G., & Fox, C.M. (2007). *Applying the Rasch model. Fundamental measurement in the human sciences*. Mahwah: Lawrence Erlbaum.
- Borg, M. (2004). The apprenticeship of observation. *ELT Journal*, 58(3), 274–276.
- Gramzow, Y. (2015). *Fachdidaktisches Wissen von Lehramtsstudierenden im Fach Physik. Modellierung und Testkonstruktion*. Zugl.: Paderborn, Univ., Diss., 2014. Berlin: Logos.
- Grossman, P.L. (1990). *The making of a teacher: teacher knowledge and teacher education: teachers college*.
- Hartig, J. (2007). Skalierung und Definition von Kompetenzniveaus. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)* (S. 83–99). Weinheim: Beltz.
- Hattie, J. (2009). *Visible Learning. A synthesis of over 800 meta-analyses relating to achievement*. London, New York: Routledge.
- Keller, M.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
- Kind, V. (2009). Pedagogical content knowledge in science education: perspectives and potential for progress. *Studies in Science Education*, 45(2), 169–204.
- Kirschner, S. (2013). *Modellierung und Analyse des Professionswissens von Physiklehrkräften*. Berlin: Logos.
- Kleickmann, T., & Anders, Y. (2011). Lernen an der Universität. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 305–315). Münster: Waxmann.
- Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., & Baumert, J. (2013). Teachers' content knowledge and pedagogical content knowledge. *Journal of Teacher Education*, 64(1), 90–106.
- Kleickmann, T., Großschedl, J., Harms, U., Heinze, A., Herzog, S., Hohenstein, F., Köller, O., Kröger, J., Lindmeier, A., Loch, C., Mahler, D., Möller, J., Neumann, K., Parchmann, I., Steffensky, M., Taskin, V. & Zimmermann, F. (2014). Professionswissen von Lehramtsstudierenden der mathematisch-naturwissenschaftlichen Fächer – Testentwicklung im Rahmen des Projekts KiL. *Unterrichtswissenschaft*, 42(3), 280–288.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E., & Vollmer, H.J. (2007). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Bonn: BMBF – Referat Öffentlichkeitsarbeit.
- Klieme, E., Baumert, J., Köller, O., & Bos, W. (2000). Mathematische und naturwissenschaftliche Grundbildung. Konzeptuelle Grundlagen und die Erfassung und Skalierung von Kompetenzen. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit. TIMSS/III Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn*, (Bd. 1, S. 85–133). Wiesbaden: VS.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern* (S. 47–70). Weinheim: Beltz.
- Lortie, D. (1975). *Schoolteacher: a sociological study*. London: University of Chicago Press.
- Magnusson, S., Borko, H., & Krajcik, J. (1999). Nature, sources and development of pedagogical content knowledge for science teaching. In *Examining pedagogical content knowledge: the construct and its implications for science education* (S. 95–132).
- Mitzel, H.C., Lewis, D.M., Patz, R.J., & Green, D.R. (2001). The bookmark procedure: psychological perspectives. In G.J. Cizek (Hrsg.), *Setting performance standards: concepts, methods, and perspectives* (S. 249–281). Mahwah: Lawrence Erlbaum.
- Mullis, I.V.S. (2012). Using scale anchoring to interpret the TIMSS and PIRLS 2011 achievement scales. In M. O. Martin & I. V. S. Mullis (Hrsg.), *Methods and procedures in TIMSS and PIRLS 2011*. Chestnut Hill: TIMSS & PIRLS International Study Center, Boston College.
- Neumann, K., Kauertz, A., Lau, A., Notarp, H. & Fischer, H.E. (2007). Die Modellierung physikalischer Kompetenz und ihrer Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 101–121.
- Neumann, K., Kind, V. & Harms, U. (2019) Probing the amalgam: the relationship between science teachers' content, pedagogical and pedagogical content knowledge. *International Journal of Science Education*, 41(7), 847–861.
- OECD (2012). Proficiency scale construction. In *PISA 2009 technical report* (S. 257–277). Paris: OECD Publishing.
- Park, S., & Chen, Y.-C. (2012). Mapping out the integration of the components of pedagogical content knowledge (PCK): examples from high school biology classrooms. *Journal of Research in Science Teaching*, 49(7), 922–941.
- 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.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.

- Riese, J. (2009). *Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkräften*. Berlin: Logos.
- 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. In S. Blömeke & O. Zlatkin-Troitschanskaia (Hrsg.), *Kompetenzen von Studierenden* (S. 55–79). Weinheim: Beltz Juventa.
- Schecker, H., Neumann, K., Theyßen, H., Eickhorst, B., & Dickmann, M. (2016). Stufen experimenteller Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 197–213.
- Schneider, R. M., & Plasman, K. (2011). Science teacher learning progressions: a review of science teachers' pedagogical content knowledge development. *Review of Educational Research*, 81(4), 530–565.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (2005). Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der Kultusministerkonferenz vom 16.12.2004. *Zeitschrift für Pädagogik*, 51(2), 280–290.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (2018). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung. Beschluss der Kultusministerkonferenz vom 11.10.2018*. Bonn: KMK.
- Shulman, L. S. (1986). Those who understand: knowledge growth in teaching. *Educational Researcher*, 15(4), 4–14.
- Shulman, L. S. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Educational Review*, 57(1), 1–22.
- 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.
- Stanat, P., Anand Pant, H., Böhme, K., & Richter, D. (2012). *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011*. Münster: Waxmann.
- Tamir, P. (1988). Subject matter and related pedagogical knowledge in teacher education. *Teaching and Teacher Education*, 4(2), 99–110.
- Woitkowski, D., & Riese, J. (2017). Kriterienorientierte Konstruktion eines Kompetenzniveaumodells im physikalischen Fachwissen. *Zeitschrift für Didaktik der Naturwissenschaften*, 23(1), 39–52.