



# Was ist Wissen, was ist Können? Deutungen des Kompetenzbegriffs und deren psychometrische Konsequenzen im Kontext von Fachwissen und Variablenkontrollstrategie

Andreas Nehring<sup>1</sup> · Martin Schwichow<sup>2</sup>

Eingegangen: 8. September 2019 / Angenommen: 23. Mai 2020 / Online publiziert: 3. Juni 2020  
© Der/die Autor(en) 2020

## Zusammenfassung

Der vorliegende Beitrag untersucht, inwiefern sich verschiedene Deutungen des Kompetenzbegriffs in psychometrische Modelle überführen lassen und welche Konsequenzen daraus für das Assessment in quantitativen Studien erwachsen könnten. Auf Grundlage von Verhältnisbestimmungen von deklarativem Fachwissen und prozessorientierten Kompetenzen werden dazu Between- und Within-Item-IRT-Modelle definiert und zur Re-Analyse eines Datensatzes einer Querschnittsstudie zur Variablenkontrollstrategie ( $n = 990$ ) angewendet. Bei Betrachtung der Personenschätzer für die Kompetenzausprägungen, die aus den Modellen generiert werden, zeigt sich, dass ca. ein Drittel der Schülerinnen und Schüler, je nach Modell, unterschiedlichen Quartilen der Verteilung zugewiesen werden. Gleichzeitig wird deutlich, dass die Modelle zu unterschiedlichen Informationen hinsichtlich der Zunahme an Wissen oder prozessorientierten Kompetenzen über Jahrgänge hinweg wie auch hinsichtlich von Geschlechterunterschieden in der Kompetenzausprägung führen können. Die Ausprägungen prozessorientierter Kompetenzen von Mädchen fallen aufgrund oder vielmehr trotz des (im Mittel) geringeren Fachwissens höher aus, wenn der Psychometrie ein unabhängiges, aber fachwissensrelatives Kompetenzverständnis zugrunde gelegt wird. Der Beitrag diskutiert diese Befunde mit Blick auf Konsequenzen für Grundlagenforschung, Systemmonitoring und Unterrichtsentwicklung.

**Schlüsselwörter** Kompetenzmessung · Summatives Assessment · IRT-Modellierung · Variablen-Kontroll-Strategie

---

**Zusatzmaterial online** Zusätzliche Informationen sind in der Online-Version dieses Artikels (<https://doi.org/10.1007/s40573-020-00113-y>) enthalten.

✉ Andreas Nehring  
nehring@idn.uni-hannover.de

Martin Schwichow  
martin.schwichow@ph-freiburg.de

<sup>1</sup> Institut für Didaktik der Naturwissenschaften, Leibniz Universität Hannover, Am Kleinen Felde 30, 30167 Hannover, Deutschland

<sup>2</sup> Pädagogische Hochschule Freiburg, Kunzenweg 21, 79117 Freiburg, Deutschland

## What Is Knowledge, What Is Ability? Interpreting the Notion of Competence and Investigating the Psychometrical Consequences in the Context of Content Knowledge and the Control-of-Variables-Strategy

### Abstract

With regard to the correlation of science knowledge and the ability to apply the control-of-variables strategy, the present paper investigates whether conceptualizing the role of knowledge for competences differently can motivate different psychometric models and whether using these models may lead to different information in quantitative studies. Using a sample from a cross sectional study ( $n=990$ ) on the correlation of knowledge and the control-of-variables-strategy as part of experimental competence in physics, we compared the results of applying between- und within-item-IRT-models that refer to different conceptualizations. Based on quartiles of the competence distributions, we found that only 67.4% of the students will be categorized into the same quartiles. Moreover, the models lead to different information about the competence development across years as well as about gender differences. We discuss these findings with regard to basic research, to system monitoring and to the development of teaching practices focusing on competences.

**Keywords** Assessment of competence · Summative assessment · IRT modelling · Control-of-variables strategy

### Einleitung

Die Modellierung und Erfassung von Kompetenzen gehören zu den zentralen Forschungsschwerpunkten der empirisch forschenden Fachdidaktiken wie auch der empirischen Bildungsforschung (Gesellschaft für Didaktik der Chemie und Physik 2019; Schecker und Parchmann 2006). Dabei zählt die Untersuchung von Geschlechterunterschieden in der Ausprägung naturwissenschaftlicher Kompetenz zu den häufig analysierten Merkmalen des Bildungssystems. Diese Unterschiede führen potentiell zu Disparitäten bei der Berufswahl und ziehen damit Konsequenzen in der individuellen und gesellschaftlichen Nutzung von kreativen und ökonomischen Potentialen nach sich (Pant et al. 2013; Reiss et al. 2016). In den Fachdidaktiken der naturwissenschaftlichen Fächer werden – spätestens seit der Einführung nationaler Bildungsstandards (KMK 2005a, 2005c, 2005b) – zahlreiche Modelle und Tests entwickelt, die sowohl im Large-Scale-Bereich des Systemmonitorings (Pant et al. 2013; Wellnitz et al. 2012) als auch im Bereich naturwissenschaftsdidaktischer Beobachtungs- und Interventionsstudien eingesetzt werden (Arnold et al. 2014). Im Fokus zahlreicher Studien (z. B. Grünkorn et al. 2014; Schwichow et al. 2016; Nehring 2014; Vorholzer et al. 2018) steht dabei die Beherrschung naturwissenschaftlicher Arbeitsweisen, wie z. B. der Umsetzung von Experimenten oder der Nutzung von Modellen, als zentrales Ziel des naturwissenschaftlichen Unterrichts weltweit (Hodson 2014).

Dementsprechend kommt den Prozessen der experimentellen Erkenntnisgewinnung, zu denen Fähigkeiten zur Planung, Durchführung und Interpretation aussagefähiger Experimente gehören, eine grundlegende Bedeutung zu. Um beim Experimentieren eindeutige Aussagen über Ursache-Wirkungs-Beziehungen zu erhalten, manipulieren Wissenschaftlerinnen und Wissenschaftler aktiv den

Beobachtungsgegenstand, sodass sie zwei Bedingungen vergleichen können, die sich nur hinsichtlich einer Variablen unterscheiden. Wird nur eine Variable verändert und alle weiteren Variablen konstant gehalten, so kann untersucht werden, ob die veränderte Variable einen Einfluss auf eine abhängige Variable hat. Das Kontrollieren potenziell konfundierender Variablen (alternative Ursache-Wirkungs-Beziehungen) gewährleistet eine Identifizierung kausaler Zusammenhänge und begründet die höhere Aussagekraft (Validität) von Experimenten im Vergleich zu reinen Beobachtungen (Woodward 2003). Dieses grundlegende Prinzip der experimentellen Erkenntnisgewinnung wird als Variablen-Kontroll-Strategie (VKS) bezeichnet und umfasst mehrere Teilfähigkeiten. Neben der Fähigkeit zum Planen kontrollierter Experimente (Teilfähigkeit Planung kurz PL) umfasst die VKS die Fähigkeiten zur Interpretation kontrollierter Experimente (Teilfähigkeit Interpretation kurz IN), zur Identifizierung kontrollierter Experimente aus einer Auswahl an kontrollierten und unkontrollierten Experimenten (Teilfähigkeit Identifizieren, kurz ID), sowie das Verständnis der fehlenden Aussagekraft unkontrollierter Experimente (Teilfähigkeit Verständnis kurz VER, Chen und Klahr 1999). Aufgrund der zentralen Rolle der VKS in der wissenschaftlichen Erkenntnisgewinnung ist sie ein eigenständiges Lernziel naturwissenschaftlichen Unterrichts (Schwichow et al. 2016).

### Theoretischer Hintergrund

#### Von Wissen über Teilfähigkeiten zur Kompetenz

In der häufig zitierten Definition des Kompetenzbegriffs von Weinert (2001) wird Kompetenz allgemein als eine Menge an Fähigkeiten und Fertigkeiten sowie affektiven Voraus-

setzungen zum Lösen von Problemen in variablen Situationen definiert. Das konkrete Zusammenspiel der einzelnen Fähigkeiten und Fertigkeiten, die beim Lösen komplexerer Probleme zu koordinieren sind, wird hingegen nicht näher beschrieben. Dies ist in einer allgemeinen Kompetenzdefinition sicher kaum möglich, da für ein erfolgreiches Lösen von Problemen spezifische Teilkompetenzen, Fähigkeiten und Wissensfacetten – je Kompetenzbereich – in verschiedener Weise interagieren. Für eine formative Kompetenz-erfassung sind jedoch Informationen über die Funktionen, Ausprägungen und Interaktionen verschiedener Teilfähigkeiten und deklarativer Wissens Elemente wünschenswert, um z. B. differenziertes Assessment betreiben oder passgenau und effiziente unterrichtliche Intervention designen zu können (Klieme und Leutner 2006). Dabei ist vor allem die Rolle des deklarativen Fachwissens für die Ausprägung spezifischer Kompetenzen von Interesse, da z. B. eine Abhängigkeit der Kompetenzausprägung von konkretem deklarativem Fachwissen eine notwendige Vermittlung dieses Wissens zur Förderung der Kompetenzen bedeuten würde.

Wenn in diesem Beitrag von „deklarativem Wissen“ gesprochen wird, soll an die Unterscheidung in deklaratives und prozedurales Wissen angeknüpft werden. Prozedurales Wissen umfasst eine Verknüpfung zu motorischen Aspekten und kann als handlungsnahes Wissen beschrieben werden. Deklaratives Wissen ist wiederum eine Wissensart, die explizit verbalisiert werden kann und Faktenwissen und konzeptuelles Wissen umfasst (Anderson und Krathwohl 2001). Das „deklarative Fachwissen (dFW)“ in diesem Artikel legt damit den Bezug auf den deklarativen Teil von Wissen in einem spezifischen Inhaltsgebiet.

Das Zusammenspiel unterschiedlicher Wissens Elemente und Fähigkeiten sei am Beispiel experimenteller Kompetenz konkretisiert: Im Prozess der experimentellen Erkenntnisgewinnung sind, neben epistemischen Vorstellungen (Neumann und Kremer 2013), deklaratives Fachwissen über das Inhaltsgebiet der Experimente als auch prozesshafte Fähigkeiten, wie die Fähigkeit zur gezielten Anwendung der VKS, notwendig und aufeinander abzustimmen. So werden z. B. Hypothesen auf Grundlage des Vorwissens der Experimentierenden gebildet, die wiederum Voraussetzung für die Planung kontrollierter Experimente sind. Des Weiteren bedürfen die Auswahl und Handhabung von Messgeräten und die Interpretation von Ergebnissen, neben den entsprechenden Fähigkeiten und Fertigkeiten, auch deklaratives Fachwissen (Schwchow und Nehring 2018). Auf der Ebene intra-individueller Prozesse ist empirisch belegt, dass themenspezifisches Fachwissen eine erfolgreiche experimentelle Erkenntnisgewinnung fördert. Bereits 1991 zeigten Arbeiten von Schauble et al. (1991) im Kontext der Elektrizitätslehre, dass Lernende mit höherem Vorwissen differenziertere und genauere Hypothesen bilden und letztlich auch häufiger die Gesetzmäßigkeiten der Kirch-

hoffschen Regeln<sup>1</sup> experimentell herausarbeiten konnten. Weniger Vorwissen führte zu eher explorativen Experimenten, die weniger gezielt geplant waren und auf deren Grundlage auch seltener die fachlich korrekten Zusammenhänge gefunden wurden.

Bisher jedoch werden prozessorientierte Kompetenzen wie auch das Fachwissen als inhaltsorientierte Kompetenz sowohl auf curricularer Ebene (Fachwissen inhaltsorientierte Kompetenz; KMK 2005a, 2005c, 2005b) als auch auf psychometrisch-empirischer Ebene (Kampa und Köller 2016; Klos et al. 2008) als getrennte Konstrukte angesehen. Entsprechende Studien zeigen allerdings, dass zwischen deklarativem Fachwissen und den experimentellen Kompetenzen substantielle Korrelationen bestehen (ebenda). Dass die Berücksichtigung des Fachwissens bei der Erfassung experimenteller Kompetenzen durchaus einen Effekt auf Schlussfolgerungen hat, zeigen Analysen von Längsschnittstudien. Werden nur die Lösung von Aufgaben zu experimentellen Kompetenzen betrachtet, finden sich deutliche Kompetenzzuwächse über die Jahrgangsstufen, die nahelegen, dass Schülerinnen und Schüler über ihre Schullaufbahn hinweg einen Zuwachs an experimentellen Kompetenzen erfahren (Henke 2007; Scherer 2012). Analysen, die hingegen auch das deklarative Fachwissen der Schülerinnen und Schüler berücksichtigen, zeigen, dass es gerade nicht klar ist, ob diese Kompetenzausprägungen älterer Schülerinnen und Schüler auf zunehmendes Fachwissen oder auf zunehmende prozessorientierte Kompetenzen selbst zurückzuführen sind (Nehring et al. 2015; Schwchow und Nehring 2018). Dementsprechend könnten ungenaue Informationen zum Fachwissenserwerb und genuinem Kompetenzerwerb im Laufe des Schulbesuchs und damit zur Wirkung des naturwissenschaftlichen Unterrichts vorliegen.

Vor dem Hintergrund einer evidenzbasierten Weiterentwicklung von Unterricht ist diese „Diffusität“ als problematisch anzusehen, da mit der Einführung von kompetenzbasierten Bildungsstandards durch die KMK (2010, S. 9) auch eine „Entwicklungsfunktion“ beschrieben worden ist. Diese bestehe in „einer schrittweisen Erweiterung von bisher weit verbreiteten Unterrichtsskripten, die sich mehr oder weniger ausschließlich auf die Vermittlung fachlicher Inhalte konzentrieren, also um die Entwicklung von Fähigkeiten und Fertigkeiten“ (KMK 2010, S. 9). Verfahren zur Erfassung von experimentellen Kompetenzen, die aufgrund der hohen Zusammenhänge zwischen Fachwissen und Kompetenzen weniger differenzielle Informationen über beide Konstrukte zulassen, könnten zu Fehlschlüssen über die

<sup>1</sup> Die Kirchhoffschen Regeln beschreiben den Spannungsabfall und Stromfluss in Stromkreisen. Die Maschenregel besagt, dass sich sämtliche Spannungen in einem geschlossenen Stromkreis (einer Masche) zu Null addieren. Die Knotenregel besagt wiederum, dass sämtliche in einen Punkt hineinfließende Ströme auch wieder aus ihm hinausfließen.

Wirkungen eines auf prozessorientierte Kompetenzen fokussierten Unterrichts kommen, da potentiell auch ein stark an Fachwissen orientierter Unterricht mit einem Zuwachs auf kompetenzorientierten Skalen führen könnte. Um zwischen Fachwissenserwerb und der Entwicklung experimenteller Kompetenzen differenzieren zu können, müssen zunächst theoretische Modelle zum Zusammenspiel der beiden Konstrukte beim Experimentieren aufgestellt werden.

### Perspektiven auf das Zusammenspiel von Fachwissen und prozessorientierten Kompetenzen

Dabei sind zwei unterschiedliche Sichtweisen denkbar. Beide Sichtweisen gehen von einer Trennung zwischen Fachwissen und prozessorientierten Kompetenzen aus, unterscheiden sich jedoch in den zugrunde liegenden Annahmen über deren Verhältnis zueinander, was Konsequenzen für die psychometrische Modellierung hat.

Da es sich im empirischen Teil dieses Artikels um eine Re-Analyse bestehender Daten handelt, beziehen sich die folgenden Ausführungen auf die Anwendung der VKS (operationalisiert in vier Teilfähigkeiten nach Chen und Klahr 1999) als spezifischen Teil experimenteller Kompetenz und deklaratives Fachwissen in den Bereichen Wärme- und Elektrizitätslehre:

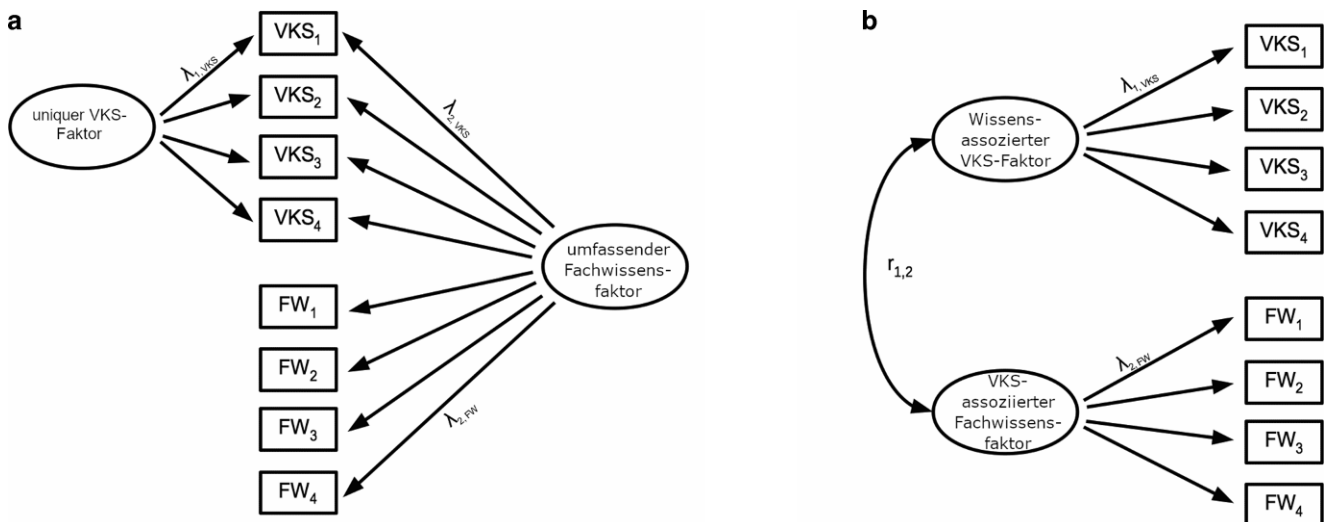
1. Wird von einem *unabhängigen Verhältnis* ausgegangen, so wäre für ein erfolgreiches Experimentieren sowohl die Anwendung von deklarativem Fachwissen als auch der VKS notwendig. Jedoch bestünden beide Konstrukte insofern in einem unabhängigen Verhältnis, als sie sich nicht gegenseitig ausgleichen können. Ein Mangel an deklarativem Fachwissen kann beim Lösen eines experimentellen Problems nicht durch eine höhere VKS-Fähigkeit ausgeglichen werden. Im umgekehrten Sinne kann eine geringe VKS-Fähigkeit nicht durch stark ausgeprägtes Fachwissen kompensiert werden. Die VKS-Fähigkeit ist damit stärker als unabhängig vom Fachwissen anzusehen.
2. Die zweite Sichtweise auf den Zusammenhang zwischen deklarativem Fachwissen und VKS-Fähigkeiten beschreibt ein *korreliert-abhängiges Verhältnis*. Deklaratives Fachwissen und VKS-Fähigkeiten erhöhen die Wahrscheinlichkeit, experimentelle Probleme zu lösen, und sind verschiedene und unabhängige Konstrukte. Sie können sich aber bei der Problemlösung gegenseitig positiv beeinflussen. Ein Mangel an Fachwissen kann durch höhere VKS-Fähigkeiten ausgeglichen werden. Auch wenn Wissen bei dieser zweiten Sichtweise nicht notwendig für die erfolgreiche Anwendung der VKS sein muss, kann es dennoch die Ausprägung der Fähigkeit zur VKS-Anwendung erhöhen.

Auf Grundlage einfacherer Korrelationen kann bisher nicht unterschieden werden, welche der beiden Sichtweisen für den Zusammenhang zwischen Fachwissen und VKS eher zutreffend ist, da beide Sichtweisen eine Korrelation zwischen beiden Konstrukten nach sich ziehen.

### Kompetenzverständnisse und deren psychometrische Konsequenzen

Die unterschiedlichen theoretischen Sichtweisen auf das Verhältnis von deklarativem Fachwissen und VKS-Fähigkeit als Teil experimenteller Kompetenzen können durch verschiedene latente Modelle abgebildet werden. Ein *unabhängiges Verhältnis* zwischen beiden Konstrukten kann durch einen Faktor, auf dem sowohl Fachwissens- als auch VKS-Items laden, repräsentiert werden (Abb. 1a). Dieser Faktor erklärt Varianz in beiden Itemtypen und bildet damit zu großen Teilen das deklarative Fachwissen bzw. die deklarativen Fachwissensanteile ab, die zum Lösen von Fachwissens-Items als auch VKS-spezifischen Items notwendig sind („umfassender Fachwissensfaktor“). Um dennoch zwischen Fachwissen und Fähigkeit zur Anwendung der VKS zu unterscheiden, laden die VKS-Aufgaben zusätzlich noch auf einen unabhängigen VKS-Faktor. Dieses als „uniquer VKS-Faktor“ bezeichnete latente Konstrukt ist mit dem Fachwissensfaktor unkorreliert. Damit erhöhen Ausprägungen auf dem Fachwissensfaktor und dem uniquer VKS-Faktor unabhängig voneinander die Wahrscheinlichkeit, VKS-spezifische Items zu lösen. Gleichzeitig wird aber auch angenommen, dass Fachwissensanteile hilfreich sind, um VKS-spezifische Items zu lösen. Diese Modellierung entspricht einem Verständnis prozessorientierter Kompetenzen als, dem Fachwissen gegenüber, eigenständigen Merkmalen („unique Kompetenz“).

Modelle diesen Typus werden auch als genestete Modelle oder mehrdimensionale Within-Item Modelle bezeichnet, da durch die Ladung der VKS-Aufgaben auf beide Faktoren die Mehrdimensionalität innerhalb derselben Aufgaben und nicht zwischen getrennten Aufgaben modelliert wird (Hartig und Höhler 2008). Sie sind Teil der Klasse von Bi-Faktormodellen, in denen die Varianz eines Items durch mehrere Faktoren erklärt wird (Reise et al. 2007). Within-Item Modelle bieten die Möglichkeit, mehrdimensionale Datenstrukturen zu repräsentieren, wobei die Faktorladungen der Items Aufschluss darüber geben, wie viel Varianz durch welchen Faktor erklärt wird (Chen et al. 2006). Im konkreten Fall dieses Artikels kann damit eine Verhältnisbestimmung des Einflusses des umfassenden deklarativen Fachwissensfaktors und des unigen VKS-Faktors vorgenommen werden (siehe auch Methodenteil). Die Anwendung dieser Modelltypen stellt, mit Blick auf den Stand der kompetenzorientierten naturwissenschaftsdidaktischen For-



**Abb. 1** Darstellung des Within-Item (a) und Between-Item Modelles (b)

scheidung, eine Innovation dar (siehe auch Abschnitt Modellierungen in der bisherigen fachdidaktischen Forschung).

Ein *korreliert-abhängiges* Verhältnis kann durch zweidimensionale Modelle, in denen jede Aufgabe entweder nur auf den Fachwissens- oder nur auf den VKS-Faktor lädt und die von einer Korrelation zwischen beiden Faktoren ausgehen, abgebildet werden (Abb. 1b). Da die Mehrdimensionalität in diesem Fall nicht innerhalb derselben Aufgaben, sondern zwischen verschiedenen Aufgaben modelliert wird, firmieren die Modelle auch unter der Bezeichnung Between-Item Modelle (Hartig und Höhler 2008). Diese Modellierung entspricht eher dem Stand der naturwissenschaftsdidaktischen Kompetenzforschung (siehe auch Abschnitt Modellierungen in der bisherigen fachdidaktischen Forschung). Sie entspricht einem Kompetenzverständnis, das prozessorientierte Kompetenzen zwar als vom Fachwissen eigenständige Konstrukte begreift, aber den genuinen spezifischen Anteil der prozessorientierten Kompetenz nicht eigenständig abbildet, sondern auf der Annahme beruht, dass geringe Ausprägungen der VKS-Fähigkeit durch ein höheres Fachwissen ausgeglichen werden kann. Damit bilden diese Modellierungen eigentlich ein fachwissens-integriertes Kompetenzverständnis ab und setzen, ohne es zu explizieren, die Trennung beider Konstrukte nicht konsequent um. Between-Item Modelle können vor diesem Hintergrund als eine alternative Parametrisierung von Within-Item Modellen verstanden werden.

Werden die Varianzen der Faktoren in beiden Modellen auf eins fixiert, die Faktorladungen aller Aufgaben gleichgesetzt, die auf einen Faktor laden, und wird die Korrelation zwischen beiden Faktoren im Within-Modell auf null gesetzt, so sind die Modelle hinsichtlich ihrer Freiheitsgrade identisch und hinsichtlich ihrer Passung zu den Daten vergleichbar (Blömeke und Suhl 2010). Eine Unterscheidung

zwischen den Modellen auf Grundlage informationstheoretischer Kriterien ist folglich nicht möglich.

Allerdings unterscheidet sich die theoretische Bedeutung der Faktoren in beiden Modellen. Im Within-Modell repräsentiert der gemeinsame Faktor, auf dem sowohl die Fachwissens- als auch die VKS-Items laden, das deklarative Fachwissen, das sowohl zum Lösen der Fachwissens- als auch der VKS-Items dient („umfassender Fachwissensfaktor“). Der unique VKS-Faktor erklärt die verbliebene Varianz auf den VKS-Items und beschreibt damit den Anteil der VKS-Fähigkeit, der nicht mit dem Fachwissen zusammenfällt. Die Varianz auf dem VKS-Faktor ist damit um die Varianz aus dem Fachwissen bereinigt. Die VKS-Fähigkeit kann damit spezifischer und vom Fachwissen unabhängiger abgebildet werden.

Im Between-Modell hingegen beschreiben die Fachwissens- und VKS-Faktoren zwei empirisch trennbare Konstrukte, zwischen denen ein Zusammenhang in Form einer Korrelation besteht. Im Unterschied zur Within-Modellierung kann nicht unterschieden werden, inwiefern die gezeigte Performanz auf den VKS-Items auf ein höheres Fachwissen oder eine genuin höhere Ausprägung der VKS-Fähigkeiten zurückzuführen ist. Damit stellt der VKS-Faktor bei genauerer Betrachtung einen integrativen Faktor dar, der sowohl spezifische Anteile der VKS-Fähigkeit als auch des Fachwissens umfasst.

Zusammenfassend liegt den Modellen ein unterschiedliches Verständnis von VKS-Kompetenz und ihrem Verhältnis zum deklarativem Fachwissen (dFW) zugrunde: Im Between-Item Modell stehen das dFW und die VKS-Kompetenz in einer zwar getrennten, aber korrelierten Beziehung zueinander. „Kompetentsein“ kann in diesem Sinne sowohl durch ein höheres dFW als auch durch eine höhere spezifi-



**Tab. 1** Perspektiven auf Kompetenz und entsprechende psychometrische Modelle

Psychometrische Bezeichnung	Anzahl Dimensionen	Theoretische Perspektive auf Kompetenzen/ Kompetenzverständnis	Implikation: latenter Faktor 1	Implikation: latenter Faktor 2
<b>Within-Item Modell</b>	2	<i>Unabhängiges Verhältnis:</i> VKS-Kompetenz und deklaratives Fachwissen sind getrennt, jedoch befähigt die VKS-Kompetenz zum Problemlösen über das zur Verfügung stehende deklarative Fachwissen hinaus	<i>Umfassender deklarativer Fachwissensfaktor:</i> Fachwissensanteile der deklarativen Fachwissensitems und VKS-Items	<i>Uniquer VKS-Faktor:</i> VKS Kompetenz, die um das deklarative Fachwissen bereinigt ist
<b>Between-Item Modell</b>	2	<i>Korreliert-abhängiges Verhältnis:</i> Kompetenzen und deklaratives Fachwissen sind getrennt, aber assoziiert und korreliert. „Kompetentsein“ kann sowohl durch ein höheres deklaratives Fachwissen als auch durch eine höhere Ausprägung der VKS-Kompetenz ermöglicht werden	<i>VKS-assoziiertes Fachwissensfaktor:</i> Mit VKS-Kompetenzen assoziiertes deklaratives Fachwissen	<i>Wissensassoziierter VKS-Faktor:</i> Kompetenzausprägungen können durch deklaratives Fachwissen erhöht werden k.

sche Ausprägung der VKS-Kompetenz ermöglicht werden. Das dFW ist damit implizit Teil der VKS-Kompetenz.

Dem entgegen steht das Kompetenzverständnis eines Within-Item Modells: Hier geht das dFW nicht in die Ausprägung der VKS-Fähigkeit ein. „Kompetentsein“ kann durch dFW erklärt werden. Jedoch wirkt zusätzlich ein weiterer unabhängiger VKS-Faktor, der um das dFW bereinigt ist. Damit wird die VKS-Kompetenz zu einem Faktor, der, über das bestehende dFW hinaus, zum Lösen von VKS-Problemen befähigt. Dementsprechend bildet der VKS-Faktor, auf dem allein die kompetenzorientierten VKS-Items laden, eine „unique Kompetenz“ ab. Das Verständnis der VKS-Kompetenz ist mit Blick auf das dFW damit differentiell („über das deklarative Fachwissen hinaus“).

Konkreter gesprochen: Auch wenn Schülerinnen und Schüler aufgrund eines geringen dFW bei kompetenzorientierten VKS-Problemstellungen weniger erfolgreich sind, würden sie – gemessen an dem, was ihnen ihr dFW ermöglicht – als relativ kompetent (bezogen auf die VKS) betrachtet werden. Im umgekehrten Sinne würden Schülerinnen und Schüler mit hohem dFW als weniger kompetent betrachtet, wenn sie weniger erfolgreich VKS Problemstellungen lösen, als es ihnen ihr dFW ermöglichen würde.

Die Modellierung einer, um das dFW bereinigten, unigen VKS-Kompetenz entspricht damit einem *unabhängigen*, aber *fachwissensrelativem Kompetenzverständnis*. Es werden Aussagen darüber ermöglicht, wie hoch das „Können“ von Schülerinnen und Schülern ausgeprägt ist angesichts des dFW, über das sie verfügen. Tab. 1 fasst diese Perspektiven in einer Übersicht zusammen.

## Modellierungen in der bisherigen fachdidaktischen Forschung

In der bisherigen fachdidaktischen Forschung wurden Fachwissen und experimentelle Fähigkeiten, wie z. B. die VKS, ausschließlich mit eindimensionalen oder mit zweidimen-

sionalen Between-Item-Modellen abgebildet. So haben z. B. Klos et al. (2008) ein Between-Item-Modell aus chemischem Fachwissen und experimentellen Kompetenzen mit einem einfachen, eindimensionalen Modell verglichen, in dem es nur einen Faktor gibt, auf den sowohl Aufgaben zur experimentellen Kompetenz als auch Fachwissensaufgaben laden. Ein Vergleich der Fit-Indizes zeigt, dass das zweidimensionale Modell besser zu den Daten passt als ein eindimensionales Modell und dass beide Konstrukte nicht signifikant korrelieren. Ihre Schlussfolgerung ist folglich, dass es sich bei Fachwissen und experimentellen Kompetenzen um getrennte Konstrukte handelt. Auch Vorholzer et al. (2016) haben Fachwissen und experimentelle Denk- und Arbeitsweisen in der Physik als getrennte, aber korrelierte Faktoren in einem Between-Item-Modell abgebildet. Sie fanden eine signifikante, aber kleine Korrelation zwischen dem Gesamtfaktor zu experimentellen Denk- und Arbeitsweisen und dem Fachwissen in Physik. In der Biologie haben sowohl Kampa (2012) als auch Hammann et al. (2007) Between-Item-Modellierungen gewählt, um den Zusammenhang von Fachwissen und experimentellen Arbeitsweisen zu untersuchen. Sie fanden mittlere signifikante Korrelationen zwischen beiden Konstrukten.

Bei der Modellierung experimenteller Kompetenzen wurden bisher keine Within-Item-Modelle eingesetzt. Allerdings zeigen Beispiele aus dem Bereich sprachlicher und fachdidaktischer Kompetenzen, dass der Einsatz von Within-Item-Modellen im Vergleich zu Between-Item-Modellen andere Konsequenzen nach sich ziehen kann. Hartig und Höhler (2008) haben Aufgaben zum Hör- und Leseverstehen in der ersten Fremdsprache Englisch mit beiden Modellen abgebildet. Sie gelangen zu deutlich unterschiedlichen Befunden bezüglich von Gendereffekten und des Einflusses von weiteren Prädiktoren (u. a. Schulart, kognitive Fähigkeiten, Umgang mit englischsprachigen Medien) auf das Hör- und Leseverständnis. Generell sind die Prädiktoren für das Hör- und das Leseverstehen im

Between-Modell nahezu identisch. Bei der Within-Item Modellierung kommt es hingegen zu größeren Unterschieden zwischen Prädiktoren der beiden Faktoren. Während in der Between-Modellierung z. B. Mädchen sowohl im Hör- als auch im Leseverständnis höhere Fähigkeiten aufwiesen, hatten Mädchen in der Within-Modellierung nur bei Leseverständnis höhere Fähigkeiten. Jungen hingegen schnitten beim Hörverständnis besser ab als Mädchen. Blömeke und Suhl (2010) haben beide Modelle bei der Beschreibung des Zusammenhangs zwischen Fachwissen und fachdidaktischen Wissen von angehenden Mathematiklehrkräften in verschiedenen Länder verglichen. Sie fanden Unterschiede in der Rangfolge der Länder zwischen beiden Modellierungen. Im Falle der Within-Modellierung war die Rangfolge in beiden Bereichen mit unterschiedlicher Schwerpunktsetzung der Curricula (Fachwissen versus fachdidaktisches Wissen) erklärbar. Zusammenfassend zeigt sich, dass Within- und Between-Item Modelle unterschiedliche Implikationen nach sich ziehen, wobei es plausibel erscheint, dass Within-Item-Modelle differenziertere Aussagen über Unterschiede zwischen korrelierten Konstrukten ermöglichen.

## Forschungsfragen

Vor diesem Hintergrund besteht das Ziel dieser Analyse darin, herauszufinden, inwiefern die Anwendung von Between- oder Within-Modellen mit einer Änderung von Informationen über die VKS beim Experimentieren einhergehen kann. Folgende Fragestellungen stehen dabei im Mittelpunkt:

1. Inwiefern führt die Anwendung von Between- oder Within-Modellen zu unterschiedlichen Informationen über die Ausprägung der VKS-Fähigkeit als Teil experimenteller Kompetenz?
2. Inwiefern wirken sich diese möglichen Unterschiede auf den Vergleich von Schülergruppen aus?
3. Inwiefern lassen sich Schülerinnen und Schüler identifizieren, die von möglichen Unterschieden in besonderem Maße betroffen sind?

## Methoden

Zur Beantwortung der Forschungsfragen werden die Daten einer Querschnittstudie zu Jahrgangunterschieden in der VKS-Fähigkeit und dem deklarativen Fachwissen reanalysiert. Alle eingesetzten Testinstrumente hatten physikalische Kontexte aus dem Bereich der Wärmelehre und der Elektrizitätslehre zum Gegenstand. Da die fachlichen Inhalte aufeinander abgestimmt waren, eignet sich dieser Datensatz insbesondere für eine Re-Analyse. Die Stichprobe setzt sich aus 990 Schülerinnen und Schülern (52 % weiblich,

48 % männlich) der fünften bis dreizehnten Jahrgangsstufe dreier schleswig-holsteinischer Gymnasien zusammen. Die Verteilung der Testpersonen über die Jahrgangsstufe ist Tab. 2 zu entnehmen.

## Testinstrumente

**VKS-Test** Der Physik-VKS-Test wurde als Online-Test konzipiert und besteht aus 48 Aufgaben, die je zur Hälfte Experimente aus den Themengebieten Wärme- und Elektrizitätslehre behandeln. Diese Themengebiete wurden ausgewählt, weil sie deutschlandweit in Curricula der Mittelstufe in sämtlichen Schulformen relevant sind. Der Test setzt sich zusammen aus jeweils zwölf Aufgaben zu jedem der vier VKS-Teilaspekte: Planung kontrollierter Experimente (PL), Identifizierung kontrollierter Experimente (ID), Interpretation kontrollierter Experimente (IN) und Verständnis der fehlenden Aussagekraft konfundierter Experimente (VER) (siehe Tab. 2). In Aufgaben zur „Planung (PL) und Identifikation (ID) kontrollierter Experimente“ wird zunächst eine Hypothese bzw. Vermutung präsentiert. Bei Planungsaufgaben werden die Testpersonen anschließend dazu aufgefordert, ein geeignetes Experiment zum Testen der Vermutung zu planen, indem sie die Ausprägung von Variablen für zwei zu vergleichende Telexperimente festlegen (siehe Abb. 1). Sobald die Ausprägung einer Variablen für ein Telexperiment durch Anklicken der entsprechenden Option festgelegt wurde, erscheint ein Bild dieser Ausprägung auf dem Bildschirm. Nach Festlegen sämtlicher Ausprägungen und somit der kontrastierten Telexperimente ist das Experiment für die Testpersonen sichtbar und kann bis zum Absenden der Antwort noch verändert werden. In Aufgaben zur „Identifikation kontrollierter Experimente (ID)“ hingegen ist aus vier grafisch dargebotenen Experimenten ein kontrolliertes Experiment, das sich zur Prüfung der Vermutung eignet, auszuwählen. Die Distraktoren zeigen konfundierte und nicht kontrastive Experimente (alle Variablen beider Telexperimente sind identisch). Aufgaben zu den Aspekten „Interpretation (IN) und Verständnis (VER)“ fordern zur Interpretation der Ergebnisse eines Experiments, bestehend aus dem Vergleich zweier Telexperimente, auf (siehe Abb. 1). Die Antwortoptionen sind für beide Aufgabentypen identisch: Variable A hat einen Einfluss auf X, Variable B einen Einfluss auf X, Variable A und B haben einen Einfluss auf X und das Experiment lässt keine sichere Schlussfolgerung zu. Abhängig davon, ob ein kontrolliertes oder ein konfundiertes Experiment gezeigt wird, ist entweder eine der ersten beiden Antwortoptionen korrekt oder die letzte. Der einzige Unterschied zwischen den Aufgabentypen besteht darin, dass Interpretationsaufgaben kontrollierte Experimente und Verständnisaufgaben konfundierte Experimente darstellen. Die VKS-Aufgaben sind in einem Matrixdesign in Viererblöcken aus jeweils

**Tab. 2** Mittleres Alter und Verteilung der Testpersonen über Jahrgangsstufen

Jahrgang	5	6	7	8	9	10	11	12	13	Total
<i>Mittleres Alter (SD)</i>	10,44 (0,57)	11,40 (0,56)	12,38 (0,65)	13,42 (0,55)	14,44 (0,65)	15,50 (0,74)	16,87 (0,63)	17,11 (0,75)	18,57 (0,80)	13,47 (2,46)
<i>Anzahl (n)</i>	161	149	138	184	93	110	23	74	58	990

einer Planungs-, Identifikations-, Interpretations- und Verständnisaufgabe aus einem Inhaltsbereich angeordnet. Jede Testperson einer Klasse wurde zufällig einem der insgesamt 12 Blöcke als Startaufgabe zugelost und hat abwechselnd Blöcke aus beiden Inhaltsbereichen bearbeitet. Durch das Testdesign wurde gewährleistet, dass fehlende Werte aufgrund unzureichender Bearbeitungszeit weder mit Inhaltsbereichen noch mit Teilaspekten korrelieren. Die Lösungswahrscheinlichkeiten der einzelnen Aufgaben liegen zwischen 14 und 88 % ( $M=60\%$ ,  $SD=24\%$ ) und decken somit ein breites Spektrum ab.

Im Kontext der Entwicklungen und Anwendungen der Instrumente konnten eine Reihe von Argumenten gesammelt werden, die für eine valide Interpretation der Testscores sprechen. Dazu gehört u. a., dass die inhaltlichen Kontexte der VKS-Aufgaben entsprechend des Physikcurriculums für Gymnasien des Landes Schleswig-Holstein gewählt (curriculare Validität) und dass sämtliche relevanten Teilkonstrukte der VKS operationalisiert wurden (ein Aspekt der Konstruktvalidität). Ein weiteres Validitätsargument liefern Befunde einer Interventionsstudie (Schwchow et al. 2016), in welcher derselbe VKS- und deklarative Fachwissenstest eingesetzt wurde. Die Ergebnisse beider Testinstrumente, die jeweils vor und nach einer VKS-Intervention im Kontext der E-Lehre eingesetzt wurden, zeigen signifikante Lernzuwächse zwischen dem Prä- und Post-Test (Sensitivität für Instruktionseffekte). Das stärkste Argument für die Validität der Ergebnisse beider Testinstrumente – im Kontext der vorliegenden Fragestellungen – stammt aus einer umfangreichen Re-Analyse des Lernzuwachses im dFW. Mittels Strukturgleichungsmodellierung untersuchten Schwchow und Nehring (2018), inwiefern der Lernzuwachs im dFW von der VKS-Fähigkeit, der Lesefähigkeit, von der allgemeinen kognitiven Fähigkeit und von der Qualität der tatsächlich von den Lernenden durchgeführten Experimente (Operationalisiert über Fotografien der Experimente) abhängt. Im Sinne der Konstruktvalidität sollte die VKS-Fähigkeit und die Qualität der Experimente den größten Einfluss auf den Lernzuwachs haben, da die VKS eine Strategie zur Erkenntnisgewinnung beim Experimentieren ist. Dies zeigen auch die Ergebnisse, da lediglich die Qualität der Experimente einen direkten Effekt auf den Lernzuwachs hat. Diese hängt wiederum von der VKS-Fähigkeit und der allgemeinen kognitiven Fähigkeiten ab. Die Lesefähigkeit hat nur auf die VKS-Fähigkeit einen direkten Effekt.

**Test des deklarativen Fachwissens** Der Test erhebt mittels 24 Einfachauswahlaufgaben (12 pro Inhaltsbereich) und sieben Mehrfachauswahlaufgaben deklaratives Wissen und konzeptionelles Verständnis in der Wärme- und Elektrizitätslehre und ist somit inhaltlich auf den VKS-Test abgestimmt. Die Einfachauswahlaufgaben (richtig/falsch) fragen nach kausalen Effekten der Variablen des VKS-Tests (z. B.: Wird ein Leiter wärmer, wenn er dünner ist?). Mittels der Multiple-Choice-Aufgaben wird deklaratives Fachwissen als Faktenwissen (z. B.: Ab welcher Temperatur schmilzt Eis?) und als konzeptionelles Wissen (z. B.: Warum frieren wir, wenn wir aus dem Wasser an die Luft kommen?) der Testpersonen erhoben. Vier der Aufgaben stellen Fragen zur Wärme- und drei zur Elektrizitätslehre. Alle Mehrfachauswahlaufgaben haben eine richtige Antwort und drei Distraktoren, die auf Basis von Schülerantworten auf offene Fragen entwickelt wurden. Der Test wurde ebenfalls als Online-Instrument umgesetzt. Die Lösungswahrscheinlichkeiten der einzelnen Aufgaben liegen zwischen 26 und 99 % ( $M=74\%$ ,  $SD=21\%$ ) und decken somit ein breites Spektrum ab. Alle Probandinnen und Probanden haben sämtliche Aufgaben des Tests in der identischen Reihenfolge vor der Beantwortung des VKS-Tests bearbeitet. Das Testinstrument wurde bei der zuvor beschriebenen Interventionsstudie eingesetzt, sodass dieselben Argumente zur Konstruktvalidität gelten wie oben beschrieben. Die Inhalte sind auf die Kontexte des VKS-Tests abgestimmt und somit ebenfalls curricular valide.

### Datenauswertung

Die Daten wurden mittels zweier unterschiedlicher, mehrdimensionaler IRT Modelle ausgewertet. Diese Modelle wurden im Theorieteil des Artikels beschrieben und hergeleitet. Im Falle des Between-Item-Modells laden die dFW- und VKS-Aufgaben jeweils nur auf einen dFW- bzw. VKS-Faktor (korreliert-abhängiges Verhältnis; siehe auch Abb. 1b). Der Zusammenhang wird durch eine Korrelation zwischen den beiden Faktoren berücksichtigt. Bei der IRT-Modellierung wird dies in einem Within-Item-Modell durch einen Kompetenzfaktor, auf dem sowohl die dFW- als auch VKS-Aufgaben laden, umgesetzt (siehe Abb. 1a). Zusätzlich gibt es noch einen unigen VKS-Faktor, auf dem nur die VKS-Aufgaben laden. Die VKS-Aufgaben laden folglich sowohl auf den unigen VKS- als auch auf den dFW-Faktor. Da der Zusammenhang zwischen beiden Kompetenzbereichen



bereits durch den gemeinsamen dFW-Faktor abgebildet ist, wird die Korrelation zwischen beiden Faktoren auf null festgelegt (Blömeke und Suhl 2010; Hartig und Höhler 2008). Die Ladung der Aufgaben auf die unterschiedlichen Faktoren wird in beiden Modellierungen so geschätzt, dass alle Aufgaben eines Kompetenzbereichs auf denselben Faktor gleich laden. Im Falle des Between-Modells werden somit zwei Faktorlandungen und eine Korrelation zwischen beiden Faktoren geschätzt. Für das Within-Item-Modell werden drei Faktorladungen bestimmt. Die Varianz der beiden Faktoren ist in beiden Modellen auf eins festgelegt.

Bei den Modellen handelt es sich um 2PL Modelle mit frei geschätzten, aber über Aufgaben eines Kompetenzbereichs gleichen Ladungen. Im Unterschied zu klassischen Raschmodellen (1PL-Modellen), bei den alle Aufgaben gleich stark auf die Faktoren laden, können mit den verwendeten Modellen Unterschiede zwischen den Ladungen verschiedener Konstrukte aufgedeckt werden. Gegenüber 2PL Modellen mit völlig frei geschätzten Ladungen bietet die Randbedingung der gleichen Ladungen innerhalb des gleichen Kompetenzbereichs den Vorteil einer einfachen Interpretation auf Ebene der betrachteten Konstrukte (Hartig und Höhler 2008), was auch aus Gründen der Modellsparsamkeit einer freien Schätzung der Faktorladungen vorzuziehen ist.

Um die Auswirkung der unterschiedlichen theoretischen Kompetenzverständnisse und ihrer statistischen Realisierung auf Kompetenzunterschiede zwischen Jahrgangsstufen und Geschlechtern zu untersuchen, wurden Regressionen der Jahrgangsstufe und des Geschlechts (0 = männlich; 1 = weiblich) auf die beiden Kompetenzfaktoren in den Modellen mitgeschätzt. Zur Analyse der Auswirkungen auf der Ebene einzelner Personen wurde für jede Testperson ein Expected A Posteriori (EAP)-Schätzer berechnet. Es wurden EAPs anstelle von Weighted Likelihood Estimators (WLE) bestimmt, da die im genutzten Statistikpaket verfügbaren Methoden keine reliablen WLE-Schätzer für Modelle mit Items liefern, die auf mehreren Dimensionen laden. Bei der Berechnung der Personenschätzer wurden keine Informationen aus einem Hintergrundmodell genutzt, da solche Plausibel-Value Verfahren sich nicht für die Schätzung individueller Fähigkeitsausprägungen eignen (Lüdtke und Robitzsch 2017). Die Reliabilität der EAP-Schätzer für die dFW-Faktoren beträgt für beide Modelle 0,70. Der VKS Faktor hat im Within-Modell eine Reliabilität von 0,72 und im Between-Modell von 0,87. Um die Personenschätzer besser interpretieren zu können, wurden die Werte auf die in den Bildungswissenschaften bekannte PISA-Metrik transformiert. Dementsprechend sind die Mittelwerte der Skalen auf 500 und die Standardabweichung auf 100 festgelegt. Für den Vergleich der Personenschätzer aus den Modellen mit den Summenwerten der Rohdaten wurden auch die Summenwerte auf dieselbe Metrik transformiert.

Die Summenwerte bilden die Grundlage der Skalierungen und können als naive Skalenwerte oder Rohwerte betrachtet werden, aus welchen die Personenschätzer der Modelle generiert werden. Die Berechnung der Modelle und der Personenschätzer erfolgte mit dem Paket TAM in R (Kiefer et al. 2017).

Um die Effekte beider Modellierungen weiterführend zu vergleichen, wird untersucht, inwiefern sich Geschlechterunterschiede in der Within-Item- und der Between-Item-Skalierung ergeben. Zwar ist der Fokus dieser Re-Analyse nicht spezifisch auf Fragestellungen der Genderforschung ausgerichtet, dennoch spielen derartige Analysen im Bereich des Systemmonitorings eine bedeutsame Rolle und werden regelmäßig in die Diskussion um die Teilhabe sämtlicher Schülerinnen und Schüler an MINT-Bildung eingebunden und mit Disparitäten der Berufs- und Studienwahl verknüpft (Reiss et al. 2019; Stanat et al. 2019). Sollten sich hier Hinweise für eine Abhängigkeit von Gendereffekten von den gewählten Modellen ergeben, hätte das weitreichende Implikationen für die genderspezifischen Aspekte von Large Scale Studien. Ziel unserer Analyse ist somit nicht eine Replikation von Gendereffekten. Stattdessen dienen diese bzw. deren Änderung als ein Kriterium, um den Einfluss der Modellwahl zu untersuchen. Im besten Fall ergeben sich für die Genderforschung daraus differenziertere Analysemöglichkeiten. Der bestehenden Datenlage geschuldet, wird das Geschlecht dichotom abgebildet. An dieser Stelle soll dies als Einschränkung gegenüber der mittlerweile höheren Offenheit für genderspezifische Vielfalt in derartigen Studien explizit benannt sein.

Als ein weiteres Kriterium, um den Einfluss des gewählten Analysemodells zu untersuchen, werden die Testpersonen entsprechend ihrer Ausprägung auf den VKS-Faktoren für beide Modelle in vier gleichgroße Gruppen (Quartile) eingeteilt. Anschließend wird die Gruppe der Schülerinnen und Schüler analysiert, die nach beiden Modellen zum einen in dieselbe Gruppe, zum anderen in unterschiedliche Gruppen eingeteilt werden. Auf diese Weise soll Aufschluss darüber erhalten werden, für welche Schülergruppen die Modellwahl einen besonders großen Einfluss auf die zugeschriebene Fähigkeitsausprägung bezüglich der VKS hat.

Die Syntax der Analysen ist als Online-Anhang des Artikels mit eingereicht. Die zugrunde liegenden Daten können bei den Autoren angefragt werden.

## Ergebnisse

In Tab. 3 sind die Faktorladungen der beiden Modelle sowie die Korrelationen zwischen den beiden Kompetenzfaktoren und die Gewichte der Regressionen der Jahrgangsstufe und des Geschlechts auf die Kompetenzfaktoren dargestellt. Es ist zu beachten, dass die Faktoren jeweils andere theore-

**Tab. 3** Faktorladungen der beiden Modelle sowie die Korrelationen zwischen den beiden Kompetenzfaktoren und die Gewichte der Regressionen der Jahrgangsstufe und des Geschlechts auf den dFW- und VKS Faktor

Within-Item Modell	$\lambda_1$ , umfassendes dFW (VKS-Items)	$\lambda_2$ , umfassendes dFW (FW-Items)	$\lambda_3$ , unique VKS	$r_{VKS,dFW}$	$b_{JG,VKS}$	$b_{JG,dFW}$	$b_{sex,VKS}$	$b_{sex,dFW}$
	***0,51 (0,004)	***0,40 (0,004)	***0,86 (0,009)	0,00	***0,20 (0,005)	***0,41 (0,006)	0,11 (0,06)	***-0,32 (0,07)
Between-Item Modell	$\lambda_4$ , wissensassoziierte VKS	$\lambda_5$ , VKS-assoziiertes dFW	–	$r_{VKS,dFW}$	$b_{JG,VKS}$	$b_{JG,dFW}$	$b_{sex,VKS}$	$b_{sex,dFE}$
	***1,01 (0,006)	***0,40 (0,005)		0,30	***0,36 (0,004)	***0,41 (0,006)	-0,06 (0,05)	***-0,33 (0,07)

$\lambda_1$  umfassendes deklaratives FW (VKS-Items): Faktorladungen der VKS-Items auf den umfassenden dFW-Faktor,  $\lambda_2$  umfassendes dFW (dFW-Items): Faktorladungen der Fachwissens-Items auf den umfassenden dFW-Faktor,  $\lambda_3$  unique VKS: Faktorladungen der VKS-Items auf den unique VKS-Faktor,  $\lambda_4$  wissensassoziierte VKS: Faktorladungen der VKS-Items auf den wissensassoziierten VKS-Faktor,  $\lambda_5$  VKS-assoziiertes dFW: Faktorladungen der Fachwissensitems auf den wissensassoziierten dFW-Faktor,  $r_{VKS,dFW}$  Korrelationen zwischen VKS- und dFW-Faktor,  $b_{VKS}$  Regressionsgewichte des Alters auf den VKS-Faktoren,  $b_{dFW}$  Regressionsgewichte des Alters auf dFW-Faktor,  $b_{sex,VKS}$ , bzw.  $b_{sex,dFW}$  Regressionsgewichte des Geschlechts (0 = männlich, 1 = weiblich) auf die VKS- bzw. dFW-Faktoren

tische Konstrukte im Sinne eines unterschiedlichen Verständnisses von Kompetenz abbilden (siehe Theorieteil). Im Between-Item-Modell beschreiben beide Faktoren getrennte, jedoch korrelierte Konstrukte, deren Ausprägung abhängig von einander ist. Im Within-Item-Modell hingegen beschreibt der zweite Faktor, auf den Items beider Kompetenzbereiche laden, einen umfassenden dFW-Faktor. Der unique VKS-Faktor hingegen beschreibt den Varianzanteil der VKS-Items, der nicht mit der Varianz deklarativer Fachwissensaufgaben zusammenfällt. Im Within-Modell wurde die Korrelation zwischen beiden Faktoren auf null festgelegt und im Between-Modell frei geschätzt.

Die Ladungen der Fachwissens-Items sind in beiden Modellen identisch (Tab. 4). Gleichzeitig ist die Ladung der VKS-Aufgaben auf den VKS-Faktor im Between-Item-Modell größer als im Within-Item-Modell. Die Höhe der Faktorladungen der VKS-Items auf den unique VKS-Faktor bestätigt, dass der unique VKS-Faktor einen eigenständigen Beitrag zur Aufklärung von Varianz auf den Items leistet. Der unique VKS-Faktor kann damit als durchaus relevante Informationsquelle für die Ausprägung der VKS-Fähigkeit angesehen werden.

Weitere Unterschiede bestehen in den Zusammenhängen der Faktoren zur Jahrgangsstufe der Testpersonen. Während die Regressionsgewichte auf die beiden Kompetenzfaktoren im Between-Item-Modell ähnlich groß sind, ist der Zusammenhang zwischen dem umfassenden Fachwissensfaktor und der Jahrgangsstufe im Within-Item-Modell deutlich größer als der Zusammenhang zwischen der Jahrgangsstufe und der VKS-Fähigkeit. Die Regressionsgewichte auf den umfassenden Fachwissensfaktor im Within-Item-Modell und das Gewicht auf den VKS Faktor im Between-Item-Modell sind nahezu identisch. Auf der Ebene der latenten Modellierungen ergeben sich keine signifikanten Unterschiede bezüglich der Abhängigkeit der Fähigkeiten vom Geschlecht der Testpersonen.

Tab. 4 zeigt Ergebnisse bezüglich Gendereffekte auf Ebene von Personenschätzer (EAPs) aus beiden Modelle und ermöglicht es, diese Ergebnisse ins Verhältnis zu den Roh-Daten zu setzen, die als Grundlage der Skalierungen dienen. Bei der Betrachtung der einfachen Roh-Daten zeigen sich geringe bis mittlere Effekte zu Ungunsten der weiblichen Testpersonen sowohl beim Fachwissen als auch bei der VKS Fähigkeit. Bei Betrachtung der Personenschät-

**Tab. 4** Vergleich der Fähigkeitsschätzung weiblicher und männlicher Testpersonen auf Basis der Personenschätzer (EAP) aus dem Between- und Within-Item-Modell und Summenwerte (naive Skalenwerte, siehe Methodenteil)

	$M$ ( $SD$ ) männliche Testpersonen	$M$ ( $SD$ ) weibliche Testpersonen	$t$ (989)	$p$	$d$
Summenwert VKS (Rohdaten)	512,25 (101,27)	488,70 (97,56)	-3,72	0,002	-0,24
Within-Item Modell (unique VKS-Faktor)	499,70 (102,57)	500,27 (97,67)	0,09	0,93	0,01
Between-Item Modell (wissensassoziiertes VKS-Faktor)	509,96 (103,41)	490,81 (95,94)	-3,01	0,003	-0,19
Summenwert Fachwissen (Rohdaten)	509,42 (110,44)	491,31 (88,53)	-2,83	0,005	-0,18
Within-Item Modell (umfassender Fachwissensfaktor)	522,45 (103,57)	479,29 (91,95)	-6,91	<0,001	-0,44
Between-Item Modell (VKS-assoziiertes Fachwissensfaktor)	522,86 (103,27)	478,91 (92,07)	-7,05	<0,001	-0,45

VKS Variablenkontrollstrategie,  $d$  Effektstärke als Cohens  $d$ ,  $t(df)$  t-Wert und Freiheitsgerade eines zweiseitigen t-test für unabhängige Stichproben,  $p$ -Wert

zer (EAPs) aus den beiden Modellen wird jedoch deutlich, dass der Effekt zu Ungunsten weiblicher Testpersonen bei der VKS-Fähigkeit im Within-Item-Modell verschwindet. Im Within-Item-Modell ist kein Vorsprung männlicher Testpersonen bei der VKS-Fähigkeit festzustellen.

Um die Auswirkung der beiden Modellierungen auf Ebene von Einzelpersonen zu betrachten, wurden die Personenschätzer für die VKS-Fähigkeit in Quartile eingeteilt und gekreuzt (Tab. 5). Auf diese Weise kann die Anzahl der Schülerinnen und Schüler identifiziert werden, die in beiden Skalierungen einer ähnlichen (Zellen der Diagonale) oder einer höheren bzw. niedrigeren Kompetenzausprägung (Zellen über oder unter der Diagonalen) zugewiesen werden.

Dabei wird deutlich, dass 67,4% der Schülerinnen und Schüler in beiden Modellen demselben Quartil zugewiesen werden. Jedoch ergibt sich für ca. 32,6% der Schülerinnen und Schülern auch eine Abweichung, wobei die Anzahl der Testpersonen, die einem höheren Quartil zugeteilt werden, gleichmäßig auf beide Modelle verteilt ist.

Um weiterführende Informationen über diejenigen Schülerinnen und Schüler zu erhalten, bei denen es zu Abweichungen zwischen den beiden Modellen gekommen ist, stellt Tab. 6 die Mittelwerte und Korrelationen zwischen den Summenwerten auf der VKS- und Fachwissensskala dar. Der Terminus „Überschätzung“ wird dabei für Schülerinnen und Schüler verwendet, die im jeweiligen Modell einem höheren Quartil zugewiesen werden als im jeweils anderen Modell.

Es zeigt sich, dass Überschätzungen im Between-Item-Modell dann auftreten, wenn überdurchschnittliche Summenwerte auf den Fachwissens-Items vorliegen. Im Within-Item-Modell ist dies der Fall, wenn die Schülerinnen und Schüler im Fachwissen unterdurchschnittliche Summenwerte haben. Interessant ist dabei, dass die Korrelation zwischen dFW und VKS in beiden Gruppen nicht signifikant ist, während sie in der Gruppe der Schülerinnen und Schüler, die in beiden Modellen gleich eingeteilt wur-

**Tab. 5** Einteilung der Testpersonen in vier Gruppen der Ausprägung der VKS-Kompetenz in beiden Modellen

Modell einer VKS- Fähigkeit (Within-Item)	Modell wissensassoziierter VKS-Fähigkeit (Between-Item)			
	1. Quartil	2. Quartil	3. Quartil	4. Quartil
1. Quartil	202 (20,4%)	44 (4,4%)	2 (0,2%)	0 (0,0%)
2. Quartil	44 (4,4%)	137 (13,8%)	64 (6,5%)	2 (0,2%)
3. Quartil	2 (0,2%)	60 (6,1%)	135 (13,6%)	51 (5,2%)
4. Quartil	0 (0,0%)	6 (0,6%)	47 (4,7%)	194 (19,6%)

Die Einteilung erfolgte auf Grundlage der EAP Werte der Personen. Zur Gruppeneinteilung wurden Quartile gebildet, um vier gleich große Gruppen zu erhalten

den, im mittleren Bereich liegt. Dies liegt daran, dass die Summenwerte der VKS in beiden überschätzten Gruppen zwar ebenfalls leicht überdurchschnittlich bzw. unterdurchschnittlich sind, jedoch nicht so weit vom Mittelwert abweichen wie die Summenwerte des dFW.

Die hohen Korrelationen von  $r=0,92$  ( $p<0,001$ ) zwischen den VKS-Skalen bzw. von  $r=0,99$  ( $p<0,001$ ) zwischen den Fachwissensskalen legen einen hohen Zusammenhang zwischen den Personenschätzern aus den unterschiedlichen Modellen nahe. Um zu untersuchen, inwiefern die Unterschiede zwischen den Modellen sich auf die individuellen Personenfähigkeiten auswirken, haben wir zudem die relative Abweichung (mittlere Abweichung des Betrags der Differenz) zwischen den Personenschätzern aus beiden Modellen sowie deren Korrelationen berechnet. Die mittlere relative Abweichung liegt auf der Personenebene für die VKS Fähigkeit bei 30,31 (22,62) und für das dFW bei 47,18 (35,84) und somit zwischen einer viertel bzw. einer halben Standardabweichung. Erwartungskonform fallen die relativen Abweichungen zwischen den Skalen für die Schülerinnen und Schüler, die in beiden Modellen unterschiedli-

**Tab. 6** Vergleich von Gruppen, die in den Modellen gleich eingeteilt bzw. über- oder unterschätzt wurden auf Grundlage von Summenwerten

	Überschätzung im Between-Item Modell	Überschätzung im Within-Item Modell	Gleiche Einteilung in beiden Modellen
Anzahl (%)	159 (16,1%)	163 (16,5%)	668 (67,4%)
<i>M (SD) dFW</i>	572,81 (82,55)	437,17 (56,22)	498,00 (101,1)
<i>M (SD) VKS-Fähigkeit</i>	534,21 (80,26)	475,19 (83,57)	497,91 (105,70)
<i>r<sub>FW,VKS</sub></i>	0,12 ( $p=0,11$ )	0,01 ( $p=0,93$ )	0,40***
<i>M (SD) relative Abweichung dFW</i>	65,04 (33,67)	63,60 (36,45)	38,92 (33,21)
<i>M (SD) relative Abweichung VKS</i>	49,23 (23,14)	44,11 (21,96)	22,43 (17,91)

Relative Abweichung: mittlere Abweichung des Betrags der Differenz zwischen Personenschätzern aus beiden Modellen, die Mittelwerte, Standardabweichungen und Korrelationen wurde auf Basis von transformierten Summenwerten ( $M=500$  und  $SD=100$ ) berechnet  
dFW deklaratives Fachwissen

chen Quartilen zugeteilt wurden, deutlich größer als die relativen Abweichungen in den Schülergruppen, die gleich eingeteilt wurden.

## Diskussion

### Zusammenfassung und Beantwortung der Fragestellungen

Im Folgenden werden die Ergebnisse der beiden psychometrischen Modelle und die sich aus ihnen ergebenden Schlussfolgerungen zunächst getrennt für jedes Modell zusammengefasst, und anschließend erfolgt die Diskussion der Unterschiede zwischen beiden Modellen und deren Interpretationen. Schließlich werden die Ergebnisse unserer Analysen mit ähnlichen Befunden zur Kompetenzmodellierung verglichen und Konsequenzen für die Unterrichts- und Forschungspraxis abgeleitet.

Im Between-Item-Modell laden die VKS-Aufgaben stärker auf den VKS-Faktor als die dFW-Items auf den Fachwissensfaktor. Dies zeigt, dass die gemeinsame Varianz zwischen den VKS-Aufgaben deutlich größer ist als zwischen den dFW-Items. Zudem korrelieren die beiden Faktoren erwartungskonform positiv miteinander. Unter Anwendung dieses Modells entwickeln sich die VKS-Fähigkeit und das Fachwissen über die Jahrgangsstufen in ähnlicher Weise. Bezogen auf das Geschlecht zeigt sich im Between-Item-Modell ein Effekt zu Ungunsten der Mädchen auf dem Fachwissensfaktor.

Im Within-Item-Modell laden sowohl die Fachwissens- als auch die VKS-Items auf den umfassenden Fachwissensfaktor. Die Ladungen auf diesen Faktor können als das Fachwissen gedeutet werden, das sowohl zum Lösen der VKS- als auch der dFW-Aufgaben notwendig ist. Allerdings zeigt die Ladung auf dem unigen VKS-Faktor, dass die VKS-Aufgaben eine vergleichsweise große gemeinsame Varianz aufweisen, die nicht mit den dFW-Aufgaben zusammenfällt. Der eigenständige Beitrag, den der unique VKS-Faktor zur Erklärung von Varianz auf den VKS-Items leistet, ist im Vergleich zum dFW-Faktor damit größer, was die Gültigkeit der Annahme einer unigen VKS-Fähigkeit unterstreicht. Es wird deutlich, dass der unique VKS-Faktor ausreichend bedeutsame Information liefert, auch wenn – oder gerade weil – er um das dFW bereinigt ist. Für die Abhängigkeit der Kompetenz- und Fachwissensausprägung von der Jahrgangsstufe ergibt sich ein verändertes Bild gegenüber dem Between-Modell. Der umfassende dFW-Faktor hängt stärker von der Jahrgangsstufe ab als der unique VKS-Faktor. Damit bestätigen sich Befunde von Schwichow und Nehring (2018), die in zwei unabhängigen Studien herausfanden, dass sich vor allem das Fachwissen und

nicht die VKS-Fähigkeiten über die Jahrgangsstufen zu entwickeln scheinen.

Einen Gendereffekt auf Ebene latenter Modelle wurde nicht identifiziert. Bei Betrachtung der Personenschätzer des Between-Modells, zeigt sich jedoch ein Gendereffekt zu Ungunsten der Mädchen sowohl für die VKS als auch für das dFW. Blickt man mit dem im Theorieteil beschriebenen *unabhängigem*, aber *fachwissensrelatives Kompetenzverständnis* (Within-Modell) auf diese Ergebnisse, deuten sich ganz andere Implikationen hinsichtlich des Vergleichs dieser beiden Gruppen an: Die an dieser Studie teilnehmenden Mädchen könnten – angesichts des geringeren Fachwissens – eine höhere *unabhängige, fachwissensrelative Kompetenzausprägung* haben und bei der problemorientierten Anwendung der VKS den Jungen möglicherweise gegenüber im Vorteil sein. Dieser Vorteil würde aber darin bestehen, aus einem geringeren Fachwissen heraus vergleichbar adäquate Problemlösungen generieren zu können. Auch wenn diese Effekte nur auf Ebene der Personenschätzer identifiziert wurden, verdeutlicht unsere Interpretation die Abhängigkeit von Gendereffekten von der Modellwahl.

Die unterschiedlichen Modellierungen haben vor allem auf die Erfassung von Jahrgangsunterschieden einen Einfluss. Je nachdem, welches Modell und welche theoretische Beschreibung des Zusammenhangs zwischen Fachwissen und VKS-Fähigkeiten gewählt werden, zeigt sich entweder eine gleichmäßige Entwicklung von Fachwissen und VKS-Fähigkeit (Between-Item Modell) oder ein deutlich stärkerer Zuwachs im dFW gegenüber den Kompetenzen zur Anwendung des VKS (Within-Item-Modell). Die Konsequenzen dieser Befunde sind für die Evaluation des naturwissenschaftlichen Unterrichts erheblich. Da Videostudien zeigen, dass eine VKS-orientierte Förderung prozessorientierter Kompetenzen im naturwissenschaftlichen Unterricht eher die Ausnahme als die Regel sind (Nehring et al. 2016), würden die Ergebnisse des Between-Item-Modells nahelegen, dass durch einen klassischen, auf die Vermittlung von Fachwissen ausgerichteten Unterricht im gleichen Maße Kompetenzen aus dem experimentellen Bereich gefördert werden. Die Ergebnisse des Within-Item-Modells hingegen zeigen, dass über die Beschuldungsdauer nicht die unique Kompetenz zur Anwendung der VKS gefördert wird, sondern dass die Schülerinnen und Schüler über die Schulzeit hinweg lernen, ihr dFW und ihre VKS-Kompetenzen besser zu integrieren. In der Folge hätte eine gezielte Förderung der VKS-Kompetenzen durchaus noch zusätzlich Konsequenzen für die unique VKS-Fähigkeit und den interagierenden Kompetenzfaktor. Würden sich vergleichbare Unterschiede in der Kompetenzerfassung im Kontext von Large Scale Assessments, wie die bundesweiten Vergleichsarbeiten zeigen, könnten diese Befunde eine hohe wissenschaftliche und gewisse politische Relevanz haben (siehe unten).



Die theoretische Sichtweise auf den Zusammenhang zwischen dFW und VKS-Kompetenz und ihre psychometrische Modellierung hat darüber hinaus Auswirkung auf die Erfassung von Kompetenzausprägungen von Schülergruppen. Die Einteilung der Schülerinnen und Schüler in Quartile zeigt, dass sogar in einer vergleichsweise groben Betrachtung nur 67,4% der Schülerinnen und Schüler in beiden Modellen demselben Quartil zugewiesen werden. Dementsprechend ist der Anteil von Testpersonen, die unterschiedlichen Quartilen zugeordnet werden, mit 32,6% substantiell. Besonders Schülerinnen und Schüler, die untypische Kompetenzausprägungen aufweisen, indem sie z. B. für ihr Fachwissen überdurchschnittlich gut im VKS-Test abschnitten, ergeben sich Unterschiede zwischen den Modellen. Dabei werden Testpersonen mit eher überdurchschnittlichem deklarativen Fachwissen im Between-Item-Modell höher eingeordnet, während Personen mit unterdurchschnittlichen Ausprägungen im Within-Item-Modell höher eingeordnet werden.

Wie bereits in anderen Studien, die eine Within- und Between-Item-Modellierung verglichen haben (Blömeke und Suhl 2010; Hartig und Höhler 2008), zeigen sich auch für den Zusammenhang zwischen Fachwissen und VKS-Kompetenzen Auswirkungen der Modellierung auf die Interpretation von Kompetenzunterschieden sowohl auf Ebene der Gesamtstichprobe als auch auf subgruppenspezifischer Ebene. Ein Unterschied zu bisherigen Beispielen für die Nutzung von Within-Modellen ist, dass empirische Daten einen engen Zusammenhang zwischen beiden Konstrukten zeigen, der allerdings aus theoretischer Sicht für die Lösung der VKS-Aufgaben Fachwissen zwar nützlich, aber nicht zwingend erforderlich ist. In der Studie von Blömeke und Suhl (2010) hingegen ist ein mathematisches Wissen notwendig zur Lösung mathematik-didaktischer Aufgaben. Unsere Ergebnisse zeigen daher, dass auch im Falle von schwächeren Zusammenhängen zwischen Konstrukten eine genaue theoretische Betrachtung der Zusammenhänge sowie eine entsprechende Modellierung erforderlich sind, da diese die Ergebnisse und ihre Interpretation beeinflussen.

### Konsequenzen für die Kompetenzerfassung und Unterrichtsentwicklung

Die Ergebnisse unserer Studien könnten weit reichende Konsequenzen für die Kompetenzmodellierung in den Fachdidaktiken und das Bildungsmonitoring haben. Anders als die strikte Trennung der Kompetenzbereiche in den Bildungsstandards suggeriert (KMK 2005a, 2005b, 2005c), scheint es zwischen dem Kompetenzbereich Fachwissen und den prozessorientierten Kompetenzen aus dem Bereich Erkenntnisgewinnung nicht zu vernachlässigende Zusammenhänge zu geben. Die Auswirkungen eines solchen Zusammenhangs wurden anhand eines Datensatzes exem-

plarisch gezeigt. Da die Kompetenzbereiche auch Teil der bundesweiten IQB-Vergleichsarbeiten (Stanat et al. 2019) sind, liegt es nahe, dass die Wahl der theoretischen Sichtweise und der psychometrischen Modellierung Einfluss auf die Ergebnisse und ihre Interpretation hat. Denkbar wäre zum Beispiel, dass Stärken einiger Bundesländer im Bereich Erkenntnisgewinnung im Between-Item-Modell nicht hinreichend aufgedeckt werden. Da bisher die Daten ausschließlich mit Between-Item-Modellen ausgewertet wurden, wären differenziertere Ergebnisse auf Grundlage von Within-Item-Modellen auch für den bildungspolitischen und fachdidaktischen Diskurs relevant.

Plausibel erscheint die Annahme eines alternativen Bundesländerrankings auch, da in der Studie Blömeke und Suhl (2010) die Modellwahl Konsequenzen für den Vergleich verschiedener Länder hatte. So wurde die Stichprobe der USA in der Within-Item-Modellierung zu einer der leistungsstärksten Subgruppen der Gesamtstichprobe auf der Skala der mathematik-didaktischen Kompetenz. Den Teilnehmenden der USA gelang es offenbar, aus ihrer, im internationalen Vergleich unterdurchschnittlichen mathematischen Kompetenz die stärkste „unique“ mathematik-didaktische Kompetenz zu ziehen.

Es bleibt auch zu prüfen, inwiefern die geschlechterspezifischen Analysen im Bildungsmonitoring von einer solchen Verzerrung betroffen sein könnten oder ein *unabhängiges*, aber *fachwissensrelatives Kompetenzverständnis* zu anderen Ergebnissen führen könnte. Die Prüfung der Effekte einer solchen Modellierung trägt Perspektiven für die weitere Forschung in sich. So wäre für den IQB-Ländervergleich (Stanat et al. 2019) der naturwissenschaftlichen Fächer denkbar, den als Kompetenz abgebildete Bereich des Fachwissens als umfassenden Faktor und einen unigen Kompetenzfaktor der Erkenntnisgewinnung zu modellieren und die dort berichteten Geschlechtereffekte einer Re-Analyse zu unterziehen. Dazu müsste jedoch der Zugang zu den Daten auf Item-Ebene gegeben werden, was zum jetzigen Zeitpunkt nicht der Fall ist.

Es sollte jedoch an weiteren Datensätzen geprüft werden, inwiefern sich die Ergebnisse der Studie replizieren lassen. Ebenso könnte die Frage, inwiefern ähnliche Effekte zwischen dem Fachwissen und anderen Kompetenzbereichen, wie der Kommunikation oder der Bewertung, bestehen, in weiteren Analysen untersucht werden. In einem ersten Schritt könnten dazu die Datensätze bisheriger Studien vergleichend re-analysiert werden. Eignen würden sich beispielsweise die Studien von Klos et al. (2008), von Vorholzer et al. (2016), von Hammann et al. (2007) sowie auch von Schwichow et al. (2016).

Eine interessante Erweiterung dieser Studie wäre die Analyse von längsschnittlichen Datensätzen oder Daten aus Interventionsstudien in einem unabhängigen, fachwissensrelativen Kompetenzverständnis und Within-Item-Model-



len, da diese Datensätze kausale Interpretationen zulassen und somit dazu beitragen können, das Verhältnis von Kompetenz und Fachwissen aufzuklären. Dabei wäre es notwendig, über die Betrachtung von Geschlechtereffekten (siehe auch Einschränkungen im Methodenteil) hinauszugehen und weitere Merkmale, wie z. B. die Schulart oder auch sprachliche Fähigkeiten, in den Blick zu nehmen. Auch hier könnte in einem ersten Schritt auf die Re-Analyse bestehender Studien zurückgegriffen werden. Nötig wären hierfür Interventionsstudien bzw. Längsschnittstudien, in denen sowohl Kompetenz- als auch Fachwissenstests als Vor- und Nachtest bzw. mindestens zweimal eingesetzt wurden. Dies ist z. B. bei Kalthoff et al. (2018) sowie bei Schwichow, et al. (2016) der Fall, sodass weitere Analysen mit Blick auf eine Verstärkung der hier gefundenen Effekte möglich und zu erwarten sind.

**Danksagung** Die Autoren bedanken sich herzlich für das äußerst präzise und konstruktive Feedback von Prof. Dr. Alexander Kauertz während der Manuskripterstellung sowie der beiden, uns unbekannt, Gutachterinnen oder Gutachter im Kontext der Begutachtung.

**Funding** Open Access funding provided by Projekt DEAL.

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz 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.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

## Literatur

- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing. A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Arnold, J. C., Kremer, K., & Mayer, J. (2014). Understanding students' experiments—what kind of support do they need in inquiry tasks? *International Journal of Science Education*, 36(16), 2719–2749. <https://doi.org/10.1080/09500693.2014.930209>.
- Blömeke, S., & Suhl, U. (2010). Modellierung von Lehrerkompetenzen. *Zeitschrift für Erziehungswissenschaft*, 13(3), 473–505. <https://doi.org/10.1007/s11618-010-0130-y>.
- Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, 41(2), 189–225. [https://doi.org/10.1207/s15327906mbr4102\\_5](https://doi.org/10.1207/s15327906mbr4102_5).
- Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, 70(5), 1098–1120. <https://doi.org/10.1111/1467-8624.00081>.
- Gesellschaft für Didaktik der Chemie und Physik (2019). Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen 09. bis 12. September 2019, GDCP-Jahrestagung in Wien. <http://www.gdcp.de/index.php/gdcp/Jahrestagung>. Zugegriffen: 5. Aug. 2019.
- Grünkorn, J., Upmeyer zu Belzen, A., & Krüger, D. (2014). Assessing students' understandings of biological models and their use in science to evaluate a theoretical framework. *International Journal of Science Education*, 36(10), 1651–1684. <https://doi.org/10.1080/09500693.2013.873155>.
- Hammann, M., Phan, T. H., & Bayrhuber, H. (2007). Experimentieren als Problemlösen: Lässt sich das SDDS-Modell nutzen, um unterschiedliche Dimensionen beim Experimentieren zu messen? *Zeitschrift für Erziehungswissenschaft (ZfE)*, 8, 33–49. [https://doi.org/10.1007/978-3-531-90865-6\\_3](https://doi.org/10.1007/978-3-531-90865-6_3).
- Hartig, J., & Höhler, J. (2008). Representation of competencies in multidimensional IRT models with within-item and between-item Multidimensionality. *Zeitschrift für Psychologie/Journal of Psychology*, 216(2), 89–101. <https://doi.org/10.1027/0044-3409.216.2.89>.
- Henke, C. (2007). *Experimentell-naturwissenschaftliche Arbeitsweisen in der Oberstufe. Untersuchung am Beispiel des HIGHSEA-Projekts in Bremerhaven*. Studien zum Physik- und Chemielernen, Bd. 59. Berlin: Logos-Verlag.
- Hodson, D. (2014). Learning science, learning about science, doing science: different goals demand different learning methods. *International Journal of Science Education*. <https://doi.org/10.1080/09500693.2014.899722>.
- Kalthoff, B., Theyßen, H., & Schreiber, N. (2018). Explicit promotion of experimental skills. And what about the content-related skills? *International Journal of Science Education*. <https://doi.org/10.1080/09500693.2018.1477262>.
- Kampa, N. (2012). *Aspekte der Validierung eines Tests zur Kompetenz in Biologie*. Dissertation. Berlin: Humboldt-Universität zu Berlin, Philosophische Fakultät IV. <https://doi.org/10.18452/16640>.
- Kampa, N., & Köller, O. (2016). German national proficiency scales in biology: internal structure, relations to general cognitive abilities and verbal skills. *Science education*. <https://doi.org/10.1002/sci.21227>.
- Kiefer, T., Robitzsch, A., & Wu, M. (2017). TAM: test analysis modules. R package version 1.99999-31. <https://CRAN.R-project.org/package=TAM>. Zugegriffen: 28. Mai 2020.
- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktsprogramms der DFG. *Zeitschrift für Pädagogik*, 52(6), 876–903.
- Klos, S., Henke, C., Kieren, C., Walpuski, M., & Sumfleth, E. (2008). Naturwissenschaftliches Experimentieren und chemisches Fachwissen – zwei verschiedene Kompetenzen. Paralleltitel: Scientific experiments and chemical knowledge—two different competences. *Zeitschrift für Pädagogik*, 54(3), 304–321. <http://www.pedocs.de/volltexte/2011/4353>; <http://nbn-resolving.de/urn:nbn:de:0111-opus-43535>.
- KMK (2005a). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss. Beschluss vom 16.12. 2004*. München: Wolters Kluwer.
- KMK (2005b). *Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss. Beschluss vom 16.12. 2004*. München: Wolters Kluwer.
- KMK (2005c). *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss. Beschluss vom 16.12. 2004*. München: Wolters Kluwer.
- KMK (2010). *Konzeption der Kultusministerkonferenz zur Nutzung der Bildungsstandards für die Unterrichtsentwicklung*. Köln: Wolters Kluwer. [http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2010/2010\\_00\\_00-Konzeption-Bildungsstandards.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2010/2010_00_00-Konzeption-Bildungsstandards.pdf).

- Lüdtke, O., & Robitzsch, A. (2017). Eine Einführung in die Plausible-Values-Technik für die psychologische Forschung. *Diagnostica*, 63(3), 193–205.
- Nehring, A. (2014). *Wissenschaftliche Denk- und Arbeitsweisen im Fach Chemie*. Berlin: Logos Verlag.
- Nehring, A., Nowak, K. H., Belzen, A. U. zu, & Tiemann, R. (2015). Predicting students' skills in the context of scientific inquiry with cognitive, motivational, and sociodemographic Variables. *International Journal of Science Education*, 37(9), 1343–1363. <https://doi.org/10.1080/09500693.2015.1035358>.
- Nehring, A., Stiller, J., Nowak, K. H., Upmeyer zu Belzen, A. & Tiemann, R. (2016). Naturwissenschaftliche Denk- und Arbeitsweisen im Chemieunterricht – eine modellbasierte Videostudie zu Lerngelegenheiten für den Kompetenzbereich der Erkenntnisgewinnung. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 77–96. <https://doi.org/10.1007/s40573-016-0043-2>.
- Neumann, I., & Kremer, K. (2013). Nature of Science and epistemologische Überzeugungen – Ähnlichkeiten und Unterschiede. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 209–232.
- Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T., & Pöhlmann, C. (Eds.). (2013). *IQB-Ländervergleich 2012 – Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I*. Münster: Waxmann.
- Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16(Suppl 1), 19–31. <https://doi.org/10.1007/s11136-007-9183-7>.
- Reiss, K., Sälzer, C., Schiepe-Tiska, A., Klieme, E., & Köller, O. (2016). *PISA 2015 – Eine Studie zwischen Kontinuität und Innovation*. Münster: Waxmann.
- Reiss, K., Weis, M., Klieme, E., & Köller, O. (Eds.). (2019). *PISA 2018. Grundbildung im internationalen Vergleich*. Münster u.a.: Waxmann.
- Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991). Causal models and experimentation strategies in scientific reasoning. *Journal of the Learning Sciences*, 1(2), 201–238. [https://doi.org/10.1207/s15327809jls0102\\_3](https://doi.org/10.1207/s15327809jls0102_3).
- Schecker, H., & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften (ZfDN)*, 12, 45–66.
- Scherer, R. (2012). *Analyse der Struktur, Messinvarianz und Ausprägung komplexer Problemlösekompetenz im Fach Chemie. Eine Querschnittstudie in der Sekundarstufe I und am Übergang zur Sekundarstufe II*. Studien zum Physik- und Chemielernen, Bd. 141. Berlin: Logos-Verlag.
- Schwichow, M., & Nehring, A. (2018). Variablenkontrolle beim Experimentieren in Biologie, Chemie und Physik: Höhere Kompetenzausprägungen bei der Anwendung der Variablenkontrollstrategie durch höheres Fachwissen? Empirische Belege aus zwei Studien. *Zeitschrift Für Didaktik Der Naturwissenschaften*, 70(5), 1098. <https://doi.org/10.1007/s40573-018-0085-8>.
- Schwichow, M., Zimmerman, C., Croker, S., & Härtig, H. (2016). What students learn from hands-on activities. *Journal of Research in Science Teaching*, 53(7), 980–1002. <https://doi.org/10.1002/tea.21320>.
- Stanat, P., Schipolowski, S., Mahler, N., Weirich, S., & Hensche, S. (Eds.). (2019). *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich*. Münster u.a.: Waxmann.
- Vorholzer, A., von Aufschnaiter, C., & Boone, W. J. (2018). Fostering upper secondary students' ability to engage in practices of scientific investigation: a comparative analysis of an explicit and an implicit instructional approach. *Research in Science Education*. <https://doi.org/10.1007/s11165-018-9691-1>.
- Vorholzer, A., von Aufschnaiter, C., & Kirschner, S. (2016). Entwicklung und Erprobung eines Tests zur Erfassung des Verständnisses experimenteller Denk- und Arbeitsweisen. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 25–41. <https://doi.org/10.1007/s40573-015-0039-3>.
- Weinert, F. E. (2001). Concept of competence—A conceptual clarification. In D. S. Rychen & L.-H. Salyanik (Hrsg.), *Defining and selecting key competencies* (S. 45–65). Göttingen: Hogrefe, Huber.
- Wellnitz, N., Fischer, H. E., Kauertz, A., Mayer, J., Neumann, I., Pant, H. A., Sumfleth, E., & Walpuski, M. (2012). Evaluation der Bildungsstandards – eine fächerübergreifende Testkonzeption für den Kompetenzbereich Erkenntnisgewinnung. *Zeitschrift für Didaktik der Naturwissenschaften (ZfDN)*, 18, 261–291.
- Woodward, J. (2003). *Making things happen. A theory of causal explanation*. Oxford studies in philosophy of science. New York: Oxford University Press.