

# Konstruktion eines Testinstruments zur Erhebung des Professionswissens von Lehramtsstudierenden im Fach Sozialwissenschaften (SoWis-L)

Dorothee Gronostay<sup>1</sup>, Sabine Manzel<sup>2</sup> und Frank Eike Zischke<sup>2</sup>

<sup>1</sup>Institut für Didaktik integrativer Fächer, Fakultät Sozialwissenschaften, Technische Universität Dortmund, Deutschland

<sup>2</sup>Institut für Politikwissenschaft / Didaktik der Sozialwissenschaften, Fakultät für Gesellschaftswissenschaften, Universität Duisburg-Essen, Deutschland

**Zusammenfassung:** In diesem Beitrag wird die Konstruktion und Validierung eines Testinstruments zur Erhebung des Professionswissens von angehenden Lehrkräften im Fach Sozialwissenschaften (SoWi) vorgestellt. Der SoWis-L („Sozialwissenschaftliches Wissen – Lehrkräfte“) stellt für das Lehramtsstudienfach SoWi das erste standardisierte Testinstrument zur ökonomischen Erhebung von Professionswissen unter Berücksichtigung des Integrationsfachcharakters dar. Anhand einer Stichprobe von  $N = 374$  Lehramtsstudierenden der Universität Duisburg-Essen (UDE) wurde das 46-Item-Instrument bezüglich seiner Testgüte untersucht. IRT-Analysen ergaben zufriedenstellende Reliabilitäts- und Itemfitwerte für alle Wissensbereiche mit Ausnahme des soziologischen Fachwissens. Strukturanalysen sprechen für eine dreidimensionale Modellierung mit einer fachdidaktischen Wissensdimension und zwei fachwissenschaftlichen Dimensionen (politisches und wirtschaftliches Wissen). Die Kriteriumsvalidität des Instruments ist mit signifikanten Gruppenunterschieden zwischen Bachelor- und Masterstudierenden, zwischen gymnasialen und nicht-gymnasialen Lehrämtern sowie einer signifikanten Korrelation mit Abiturnoten indiziert.

**Schlüsselwörter:** Didaktik der Sozialwissenschaften, Lehrerbildung, Professionswissen, Testentwicklung, Wissenstest

## Construction of an Instrument to Measure the Professional Knowledge of Teacher Students in the Social Sciences (SoWis-L)

**Abstract:** This article presents a standardized test measuring the pedagogical content knowledge (PCK) and content knowledge (CK) of social science teacher students. The SoWis-L (“Sozialwissenschaftliches Wissen – Lehrkräfte,” Engl.: Professional Knowledge of Social Science Teachers) covers declarative and conceptual knowledge from three knowledge domains. The 46-item instrument was developed and validated based on a sample of  $N = 374$  social science teacher students. We examined research questions concerning psychometric test criteria, especially reliability and criterion validity. IRT analyses showed good item fit and adequate reliability for all knowledge areas except sociological content knowledge. A structural equation model resulted in a good fit of the model with three correlated latent factors, i. e., PCK and two facets of CK (political and economic CK). A comparison of the test results in teacher students aiming at a grammar school vs. a nongrammar school teaching degree and those enrolled in master vs. bachelor studies as well as significant correlations with grades in final school exams (“Abitur”) indicate the criterion validity of the SoWis-L.

**Keywords:** didactics in the social sciences, teacher education, professional knowledge, test development, knowledge test

Vor dem Hintergrund der Evidenzbasierung von Bildung ist auch im Hochschulsektor die Erfassung von Lernergebnissen in den Fokus gerückt (vgl. Diagnostica Themenheft, 66/2; Köller, Pant & Zlatkin-Troitschanskaia, 2020). In der Lehrkraftbildung bezieht sich die Forschung auf das universitär vermittelte Professionswissen der Lehrkräfte, dessen Struktur, Genese und Entwicklung sowie

Zusammenhänge mit weiteren Facetten professioneller Kompetenz. Dabei ist der Stand der Kompetenzmodellierung und Testkonstruktion in den verschiedenen Lehramtsfächern ungleich fortgeschritten. Eine empirisch fundierte Forschung zum Professionswissen setzt die Verfügbarkeit geeigneter Testinstrumente voraus. Im Bereich des Lehramts Sozialwissenschaften (SoWi) ist ein Desi-

derat festzustellen. Die wenigen existierenden Instrumente sind entweder nicht für die Zielgruppe Lehramtsstudierender zugeschnitten (z.B. Beck, Krumm & Dubs, 1998; Trepte, Loy, Schmitt & Otto, 2017; Westle & Tausendpfund, 2019) oder berücksichtigen nur die Politikwissenschaft als Bezugsdisziplin des Integrationsfachs SoWi (z.B. Weißeno, Weschenfelder & Oberle, 2013; Weschenfelder, 2014).

Im vorliegenden Beitrag wird dieses Desiderat aufgegriffen und die Konstruktion und Validierung des Testinstruments „Sozialwissenschaftliches Wissen – Lehrkräfte“ (SoWis-L) vorgestellt. Der SoWis-L erfasst deklaratives Professionswissen im Fach SoWi in den fachdidaktischen und fachwissenschaftlichen Inhaltsbereichen auf Reproduktions- und Anwendungsniveau. Sein Einsatzbereich liegt in der Forschung zu professionellen Kompetenzen von Lehrkräften. Die Testwerte des SoWis-L sollen als Indikator für das professionelle Wissen von (angehenden) SoWi-Lehrkräften interpretiert werden. Nach einer ersten Pilotierungsstudie und Expertenvalidierung wurde das Instrument Lehramtsstudierenden im Bachelor (BA) und Master (MA) ( $N = 374$ ) der Universität Duisburg-Essen zur Bearbeitung vorgelegt. Dargestellt werden theoretische Begründungen zum Konstruktionsrational und zu Inhaltsbereichen des SoWis-L (inhaltlicher und kognitiver Validitätsaspekt) sowie empirische Befunde zu strukturellen und externalen Validitätsaspekten (Messick, 1995).

## Theoretischer Hintergrund

### Konzeptualisierung von Professionswissen in der Lehrkräftebildung SoWi

Als theoretische Grundlage wird auf das generische Modell professioneller Handlungskompetenz von Lehrpersonen nach Baumert und Kunter (2006) rekurriert. Die angenommene Dreiteilung des Professionswissens in die Bereiche Fachwissen, fachdidaktisches Wissen und pädagogisches Wissen geht auf Arbeiten von Shulman (1986, 1987) zurück und hat sich in der Lehrerverfessionforschung etabliert. Von diesen drei Wissensbereichen sind sowohl das Fachwissen als auch das fachdidaktische Wissen domänenspezifisch auszugestalten. Es lassen sich verschiedene Niveaustufen differenzieren. Diese reichen vom fachlichen Alltagswissen über schulisches Wissen bis hin zum akademischen, universitären Wissen (Baumert & Kunter, 2011).

Das fachdidaktische Wissen umfasst u.a. Elemente des Wissens über „das Verständlichmachen von Inhalten“ (Krauss et al., 2008, S. 227). In Anlehnung an Park und Oliver (2008) kann fachdidaktisches Wissen unterteilt werden in Wissen zum Curriculum, Instruktions- und Vermittlungsstrategien, fachbezogene Diagnostik sowie Lehr-Lernforschung und Schülerkognitionen. Fachdidaktisches Wissen orientiert sich, in Abgrenzung zum fachunspezifischen pädagogischen Wissen, an relevanten und konkreten (hier: sozialwissenschaftlichen) Fachinhalten. Beim fachdidaktischen Wissen übernimmt die vorliegende Arbeit die generische Wissensenteilung von Park und Oliver (2008) und konzipiert entsprechende fachdidaktische Items. Im Bereich des Fachwissens zielt der SoWis-L auf universitär erwerbbares Professionswissen auf dem Niveau eines vertieften Verständnisses schulrelevanter Fachinhalte (Blömeke & Zlatkin-Troitschanskaia, 2013; Krauss et al., 2008). Berücksichtigt werden inhaltliche Kernbereiche der universitären Lehrkräftebildung im Fach SoWi (Kultusministerkonferenz (KMK) 2008, S. 58–60) sowie der Empfehlungen einschlägiger Fachgesellschaften (Deutsche Gesellschaft für Soziologie (DGS), 2005; Deutsche Vereinigung für Politikwissenschaft (DVPW), 2003; Gesellschaft für Fachdidaktik (GFD), 2004). Gemäß der Kognitionspsychologie lässt sich Wissen zudem entlang kognitiver Anforderungen unterteilen (Anderson, Krathwohl & Bloom, 2001), wobei für den SoWis-L die Unterscheidung zwischen abrufbarem, deklarativem Wissen und angewandtem Wissen übernommen wird (Großschedl et al., 2015; Kauertz et al., 2010). Bei der theoretischen Konzeptualisierung des Fachwissens SoWi ergeben sich mehrere Herausforderungen, die im nächsten Kapitel thematisiert werden.

### Modellierung und Erfassung des Professionswissens von angehenden Lehrkräften im Fach SoWi

Sozialwissenschaftliche Bildung wird je nach Bundesland sowie auch innerhalb der Bundesländer je nach Schulform in unterschiedlich zugeschnittenen Schulfächern<sup>1</sup> mit variierenden Bezugswissenschaften unterrichtet. Verbindliche Bildungsstandards für die schulische sozialwissenschaftliche Bildung gibt es nicht. Hinzu kommt, dass sozialwissenschaftliche Bildung regelmäßig bildungs- und parteipolitischen Akzentverschiebungen unterliegt (vgl. die Debatte<sup>2</sup> in NRW zur Einführung des Schulfachs „Wirtschaft-Politik“). Inhaltliche Kernelemente der universitä-

<sup>1</sup> Überblick über politisch-sozialwissenschaftliche Unterrichtsfächer unter: [http://library.fes.de/pdf-files/studienfoerderung/14009/polbild\\_als\\_unterrichtsfach\\_gesamt.pdf](http://library.fes.de/pdf-files/studienfoerderung/14009/polbild_als_unterrichtsfach_gesamt.pdf)

<sup>2</sup> [https://www.deutschlandfunk.de/neues-schulfach-wirtschaft-politik-in-nrw-protest-gegen.680.de.html?dram:article\\_id=492016](https://www.deutschlandfunk.de/neues-schulfach-wirtschaft-politik-in-nrw-protest-gegen.680.de.html?dram:article_id=492016)

ren Lehrkräftebildung im Fach legen nur die ländergemeinsamen inhaltlichen Anforderungen der Kultusministerkonferenz fest (KMK 2008, S. 58–60). Diese bilden daher den Rahmen für die Modellierung des Wissenstests SoWis-L.

Sozialwissenschaftliche Wissensbereiche wurden bislang stets aus der Perspektive einzelner Fachdisziplinen modelliert und operationalisiert. Im Folgenden werden zuerst vorliegende Wissenstests<sup>3</sup> zum politischen, dann auch zum wirtschaftlichen Fachwissen überblicksartig vorgestellt. In Deutschland wird in der *Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS)* in regelmäßigen Abständen politisches Quiz-Wissen in Tests zur Erfassung von Allgemeinwissen erhoben. Einzelne Unterskalen liegen im *Differentiellen Wissenstest* (Fürntratt & Jäger, 1996), im *Bochumer Wissenstest* (Hossiep & Schulte, 2007) sowie im *Studenten-Programme for International Student Assessment (Studenten-PISA)* (Trepte & Verbeet, 2010) vor. Diese Wissenstests bieten keine Anknüpfungspunkte für den SoWis-L. Subjektives und objektives politisches Wissen ist in der sozial- und medienpsychologischen, soziologischen und politikwissenschaftlichen Forschung immer wieder Gegenstand fachspezifischer Untersuchungen, oftmals in Verbindung mit politischem Interesse, Medien- und Informationsnutzung oder Partizipation (z. B. Westle & Tausendpfund, 2019). Der Fokus auf lehrrelevantes Professionswissen fehlt in diesen Erhebungen. Ein Überblick über den Stand der empirischen Forschung zum politischen Wissen sowohl international als auch national findet sich u. a. bei Landwehr (2017, S. 129 ff.). Zwei aktuellere Studien von Schübel (2018) und Trepte et al. (2017) werden exemplarisch vorgestellt, um die Desiderata für einen SoWis-L-Test aufzuzeigen.

Um politische Inhalte für einen Wissenstest zu bestimmen, nutzt der Politologe Schübel (2018) in Anlehnung an die amerikanische Tradition von Wissenstests die etablierte Einteilung von Delli Carpini und Keeter (1993, S. 118 ff.), die mittels Expertenbefragung in den USA die Themenfelder politische Institutionen und Verfahrensweisen, Sachfragen und Akteure als relevantes Kernwissen ausweisen. Er lehnte sich zwar an kognitionspsychologische Konzeptionen von Lernen und Wissen an (Anderson et al., 2001; Spitzer, 2003 zitiert nach Schübel, 2018, S. 12 ff.), untersuchte in seiner Arbeit jedoch nur Faktenwissen („knowing that“), nicht jedoch konzeptuelles Wissen (Neuman, 1986). Aktuelle Diskurse der Politikdidaktik und das Modell der Basis- und Fachkonzepte (Weißeno et al., 2010) als Mindeststandard wurden nicht berücksichtigt.

Das *Hohenheimer Inventar zum Politikwissen (HIP)* (Trepte, Loy, Smitt & Otto, 2017) setzte sich ebenfalls zum Ziel, politisches Wissen von Bürger\_innen valide zu erfassen. Dabei wurden auf Basis von theoretischen Vorarbeiten (u. a. Jennings, 1996; Eveland & Hively, 2008) sowie einer Inhaltsvalidierung durch Expertenurteile Items entlang der beiden Dimensionen Grundlagen und Aktuelles konstruiert. Die Dimension Grundlagen umfasst dabei „überdauernde nationale wie internationale politische Strukturen“ und „historisch bedeutsame Fakten“, während Aktuelles die Subdimension „aktuelle Themen“ und „aktuelle Akteure“ entlang der Arbeitsgebiete deutscher Bundesministerien beinhaltet (Trepte et al., 2017, S. 2). Auch wenn mit einem Rasch-Modell die Konstruktvalidierung nachweislich gelungen ist, so gilt als Limitation des Hohenheimer Inventars die ausschließliche Messung von Faktenwissen und nicht die Messung von vernetztem Konzeptwissen.

Die Konstrukte des HIP mögen auf das allgemeine politische Wissen von Bürger\_innen schließen lassen, greifen jedoch bzgl. der relevanten Wissensbereiche und Wissensarten für die Lehramtsausbildung in SoWi zu kurz. Die Frage der relevanten Inhaltsbereiche ist theoretisch in den Disziplinen umstritten und bislang nur in der Politikdidaktik mit dem Fachkonzeptmodell von Weißeno et al. (2010) theoretisch für die schulische politische Bildung begründet und empirisch in einigen Studien validiert. Studien zu schulischem politischem Wissen von Kindern und Jugendlichen greifen mehr (z. B. Richter et al., 2010; Landwehr, 2017) oder weniger (Abs & Hahn-Laudenberg, 2017) auf dieses Fachkonzeptmodell bei der Erfassung von Wissen zurück.

In der Ökonomie wird über Tests wirtschaftswissenschaftliches Wissen erfasst. Mit dem *wirtschaftskundlichen Bildungstest (WBT)* haben Beck, Krumm und Dubs (1998) den *US-Test of Economic Literacy (TEL4)* für Deutschland adaptiert; Wuttke & Beck (2002) haben damit die Eingangsvoraussetzungen von Studierenden der Wirtschaftswissenschaften erfasst. „Die Aufgaben lassen sich den vier Inhaltsbereichen Grundlagen, Mikroökonomie, Makroökonomie und Internationale Beziehungen zuordnen“ (Wuttke & Beck, 2002, S. 3) und orientieren sich an der Bloomschen Lerntaxonomie (Bloom, 1956). Bank und Retzmann (2012) kritisieren, dass der WBT für Lehrkräfte nur bedingt geeignet ist.

Förster, Happ und Zlatkin-Troitschanskaia (2012) haben mit der Adaption der amerikanischen Wissenstests *Test of Understanding of College Economics (TUCE4)* und *TEL4* ins Deutsche im Projekt WiWiKom I/II ein theoriebasiertes Modell der wirtschaftswissenschaftlichen Kom-

<sup>3</sup> Ein Wissenstest, der soziologisches Wissen erfasst, ist bis dato nicht bekannt.

petenz entwickelt und umfassend validiert. Sie berücksichtigen neben den Inhaltsbereichen drei kognitive Dimensionen (mentale Prozesse) und drei Wissensdimensionen (Fakten-, fallbasiertes und strategisches Wissen, Zlatkin-Troitschanskaia et al., 2017, S. 8). Diskriminante Analysen zwischen kognitiver allgemeiner Leistungsfähigkeit und dem ökonomischen Wissenstest anhand zweier Studierendengruppen (Wirtschafts- und Sozialwissenschaften) zeigen, dass sich die allgemeine Leistungsfähigkeit vom ökonomischen Wissen trennen lässt und auch beide Subgruppen leicht signifikante und zu erwartende Unterschiede im Wissenstest aufweisen (Zlatkin-Troitschanskaia et al., 2019, S. 129).

Wie dargestellt beschränken sich vorhandene Testinstrumente entweder auf politisches Bürgerwissen allgemein, auf politikwissenschaftliches und fachdidaktisches Professionswissen von Lehrkräften oder auf wirtschaftswissenschaftliches Wissen. Empirische Aussagen zum professionellen Wissen von Lehrkräften im Fach SoWi gemäß des generisch formulierten Modells professioneller Handlungskompetenz nach Baumert & Kunter (2006) können somit nicht getroffen werden. Eine fachspezifische Modellierung entlang des Modells wurde erstmals durch Weschenfelder (2014) im Forschungsprojekt „Professionelle Kompetenz von Politiklehrkräften“ (Weißeno, Weschenfelder & Oberle, 2013) für die Domäne der Politik vorgelegt. Ein Testinstrument, welches alle drei sozialwissenschaftlichen Fachwissensbereiche und die Fachdidaktik hinsichtlich deklarativem und konzeptuellem Wissen gleichermaßen berücksichtigt, fehlt im Kontext der Lehrkräftebildung in SoWi. Aus diesem Grund wird ein neues Testinstrument entwickelt, das diese Desiderata aufgreift und sowohl kognitionspsychologisch begründet als auch aus der Theorie des Faches heraus Professionswissen von (angehenden) Lehrkräften im Fach SoWi erfasst.

## Ziele und Fragestellungen

Das Ziel dieser Studie besteht in der Entwicklung und Validierung eines standardisierten Testinstruments zur Erhebung des Professionswissen von (angehenden) Lehrkräften im Lehramtsstudienfach SoWi. Ausgangspunkt sind Vorarbeiten im Kontext des Teilprojekts „Sozialwissenschaften“ in dem vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Projekt „ProViel“ (FKZ 01JA1910) der Universität Duisburg-Essen (Manzel & Gronostay, 2018). Der SoWis-L soll deklaratives Professionswissen im Fach SoWi in den fachdidaktischen und fachwissenschaftlichen Inhaltsbereichen auf Reproduktions- und Anwendungsniveau erfassen. Sein Einsatz-

bereich liegt in der Forschung zu professionellen Kompetenzen von Lehrkräften. Die Testwerte des SoWis-L sollen als Indikator für das professionelle Wissen von (angehenden) Lehrkräften im Fach SoWi interpretiert werden. Zudem ist eine zeitökonomische Einsetzbarkeit vorgesehen. Nicht intendiert ist die Verwendung für individualdiagnostische Zwecke oder die feste Implementation und Verwendung als Instrument der Entscheidungsfindung (Kane, 2013).

Da in der Regel und auch in diesem Beitrag nicht alle Validitätsanforderungen gleichermaßen geprüft werden können (Messick, 1989, S. 50), muss eine begründete Auswahl und Fokussierung vor dem Hintergrund der intendierten Testwertinterpretation und Nutzung des Instruments erfolgen (Kane, 2013). Es lassen sich sechs Validitätsaspekte unterscheiden (Messick, 1995). Der inhaltliche und kognitive Validitätsaspekt wird theoretisch in der Phase der Testkonstruktion durch Anlehnung an etablierte Modelle adressiert (vgl. Kapitel „Testkonstruktion“). Der Aspekt der Generalisierung wird methodisch durch eine Parameterschätzung mittels IRT-Modellen in Kombination mit einem Multi-Matrix-Testheftdesign sowie DiF-Analysen zur Testfairness berücksichtigt. Der konsequentielle Aspekt wird vor dem Hintergrund der intendierten Nutzung des SoWis-L für Forschungszwecke als nachrangig betrachtet. Vor diesem Hintergrund wird bei der empirischen Prüfung ein Fokus auf strukturelle Validitätsaspekte (Forschungsfragen F1 und F2) und auf externe Aspekte (F3, F4) gelegt.

Zunächst stellt sich die Frage, ob sich das Professionswissen im Fach SoWi mit den entwickelten Testaufgaben in den Dimensionen Fachwissen und fachdidaktisches Wissen reliabel messen lässt (F1). Zweitens soll geprüft werden, wie sich die Struktur des Professionswissens empirisch darstellt (F2). Es wird angenommen, dass sich die theoretisch hergeleitete Trennung zwischen Fachwissen und fachdidaktischem Wissen sowie zwischen den fachwissenschaftlichen Wissensbereichen zeigt. F3 bezieht sich auf die Validierungsstrategie der Methode der bekannten Gruppen (Kriteriumsvalidität): Zeigen sich die zu erwartenden Gruppenunterschiede in Bezug auf Studienfortschritt und Lehramtsstudiengänge? Angenommen wird, dass MA-Studierende aufgrund ihres Studienfortschritts, d.h. der größeren Anzahl an Lerngelegenheiten in den Modulen, gegenüber BA-Studierenden eine höhere Testleistung erbringen. Auch sollten Studierende eines Gymnasial-/Gesamtschullehramts (GyGe) eine bessere Testleistung zeigen also solche des Lehramts für Haupt-, Real-, Sekundar- oder Gesamtschulen (HRG). Ein Vergleich der Testscores von Studierenden mit praktizierenden Lehrkräften soll Hinweise bezüglich der Sensitivität des Instruments in Bezug auf Expertiseunterschiede liefern. F4 bezieht sich auf eine Validierung mittels Außen-

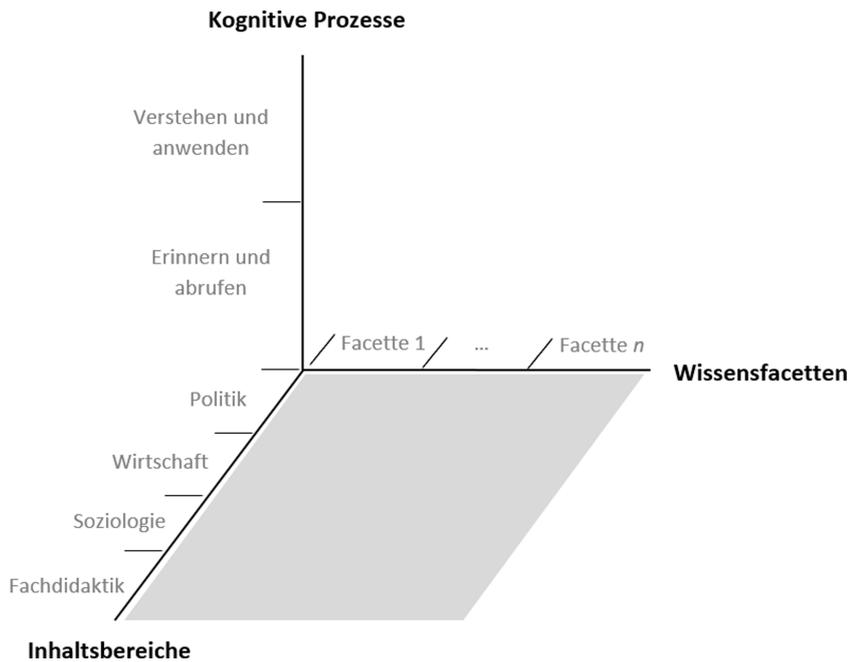


Abbildung 1. Matrix zur Aufgabenkonstruktion.

kriterien: Die Abiturnote gilt als Prädiktor für Studienerfolg und sollte daher signifikant negativ mit der Testleistung assoziiert sein (höhere Testleistung bei geringerer, d. h. besserer Abiturnote). Selbstberichtete Prüfungsnoten im Studium stellen ein weiteres Außenkriterium dar, das ebenfalls signifikant negativ mit der Testleistung korrelieren sollte.

## Methode

### Testkonstruktion

Der Testkonstruktion des SoWis-L liegt das generische Modell professioneller Handlungskompetenz von Lehrpersonen nach Baumert und Kunter (2006) sowie dessen Konkretisierung im domänenspezifischen Modell „Professionelle Kompetenz von Politiklehrkräften“ (Weißeno, Weschenfelder & Oberle, 2013; Weschenfelder, 2014) zugrunde. Zur Absicherung curricularer Validität orientierte sich die Entwicklung der Testaufgaben an den inhaltlichen Kernbereichen der universitären Lehrkräftebildung im Fach SoWi (KMK 2008, S. 58–60).

Der SoWis-L soll zwecks testökonomischer Einsetzbarkeit in maximal 60 Minuten bearbeitbar sein und alle vier Wissensbereiche mit jeweils einem Drittel Anwendung- und Verständnisaufgaben und zwei Dritteln Reproduk-

tions- und Abrufaufgaben (Dimension: kognitive Prozesse) abdecken. Die Entwicklung der Testaufgaben erfolgte systematisch anhand der in Abbildung 1 dargestellten Matrix (wobei nicht der Anspruch erhoben wird, schwierigkeiterzeugende Aufgabenmerkmale für Testaufgaben im Bereich des Professionswissens SoWi zu definieren). Konkret wurden Testaufgaben entwickelt entlang der Dimension Wissensbereich (mit den Ausprägungen Politik, Wirtschaft, Soziologie und Fachdidaktik), Facette (drei bis vier Ausprägungen je Wissensbereich) und kognitive Prozesse (mit den Ausprägungen „erinnern und abrufen“ und „verstehen und anwenden“ (Anderson et al., 2001; Großschedl et al., 2015; Kauertz et al., 2010). Insgesamt 134 Testaufgaben im Single-Choice-Format mit je vier Antwortoptionen wurden entlang dieser Matrix konstruiert. Die entwickelten Testaufgaben wurden im Rahmen einer ersten Pilotierungsstudie mit 229 Lehramtsstudierenden des Studienfachs SoWi der Universität Duisburg-Essen erprobt. Als Kriterien der Itemselektion wurde die klassische Itemschwierigkeit ( $.10 \leq P_i \leq .90$ ), die Trennschärfe ( $\geq .30$ ), der mit der IRT-Software Conquest (Version 4.14.2) berechnete Weighted Infit Mean Square ( $.80 \leq wMNSQ \leq 1.20$  und  $-1.96 \leq T\text{-Wert} \leq 1.96$ ) sowie eine graphische Inspektion der Item Characteristic Curves berücksichtigt. Zudem wurden die fachwissenschaftlichen Items je einem von insgesamt sechs Expert\_innen der Fachwissenschaften vorgelegt und die fachdidaktischen Items je einem von drei Expert\_innen aus der Fach-

**Tabelle 1.** Wissensbereiche und Facetten des SoWis-L

Wissensbereich	Facette	Itemanzahl	Kognitive Prozesse	
			Erinnern und abrufen	Verstehen und anwenden
Politik	Politische Systemlehre	7		
	Internationale Beziehungen	3	11	5
	Politische Theorie	6		
Wirtschaft	Wirtschafts- und Sozialpolitik	6		
	Makroökonomie	5	10	5
	Mikroökonomie	4		
Soziologie	Sozialstruktur	9		
	Soziale Ungleichheit	4	10	7
	Soziologische Theorien und Konzepte	4		
Fachdidaktik	Ziele und Inhalte	3		
	Instruktion und Vermittlung	7	9	6
	Fachbezogene Diagnostik	3		
	Fachdidaktische Forschung (inkl. Schülerkognitionen)	2		

didaktik (jeweils PostDocs oder Professor\_innen), um die fachliche Korrektheit<sup>4</sup> sowie die Kategorisierung bezüglich der avisierten kognitiven Prozesse<sup>5</sup> und inhaltlichen Facetten<sup>6</sup> (vgl. Abb. 1, Matrix zur Aufgabenkonstruktion) abzusichern. Es wurden insgesamt 63 Items für die Hauptstudie ausgewählt, welche in der Expertenbefragung als fachlich eindeutig und korrekt sowie im Sinne der avisierten Aufgabenmerkmale kategorisiert wurden (vgl. Tabelle 1). Dabei wurde darauf geachtet, dass Items der intendierten Inhaltsbereiche und kognitiven Prozesse (gemäß Expertenurteil) in ausgewogenem Verhältnis und unter Berücksichtigung ihrer psychometrischen Eigenschaften im reduzierten Itempool enthalten sind. In den elektronischen Supplementen zum Beitrag sind die Items des SoWis-L im Überblick dargestellt (ESM 1) und Itembeispiele verfügbar (ESM 2).

## Stichprobe

Als Datengrundlage der Hauptstudie diente eine nicht-probabilistische Stichprobe von  $N = 374$  Lehramtsstudierenden des Fachs SoWi der UDE, die der Zielgruppe des zu entwickelnden Testinstruments entsprechen. Es wurden Studierende in BA- und MA-Studiengängen akquiriert, um erwartbare Fähigkeitsunterschiede nach Studienfort-

schrift untersuchen zu können. Das Durchschnittsalter der Befragten lag zwischen 19 und 40 Jahren ( $M = 24.79$ ,  $SD = 3.09$ ). Der Anteil weiblicher Studierender betrug 58 % und 34 % verfügten über einen akademischen Hintergrund (mindestens ein Elternteil mit (Fach-)Hochschulabschluss). Ein Anteil von 23 % der Befragten gab eine andere Muttersprache als Deutsch an. Mehr als drei Viertel der Befragten studierten für ein GyGe-Lehramt (76 %), die übrigen für ein HRG-Lehramt. Genau 50 % der Proband\_innen studierte einen MA-Studiengang, die andere Hälfte einen BA-Studiengang.

## Erhebung

Die Papier-Bleistift-Erhebung erfolgte im Zeitraum Sommersemester 2018 bis Wintersemester 2019/2020 in Pflichtveranstaltungen des Lehramtsstudiums SoWi, um eine möglichst gleiche Auswahlwahrscheinlichkeit der Studierenden der relevanten Studiengänge für die Stichprobe sicherzustellen. Geschulte Versuchsleiter\_innen führten die 60-minütige Befragung jeweils zu Beginn der Vorlesungszeit in den stets gleichen fachdidaktischen Lehrveranstaltungen im BA (5./6. Fachsemester) und MA des Lehramts SoWi durch, um eine möglichst vergleichbare Erhebungssituation zu gewährleisten. Die Testleis-

<sup>4</sup> Itemwording: „Ist das Item fachlich korrekt und eindeutig zu beantworten?“ mit den Single-Choice Antwortoptionen „ja“ und „nein“ (sowie einer Freitextoption für Kommentierungen).

<sup>5</sup> Itemwording „Erfordert das Item eine Reproduktion oder Anwendung von Wissen?“ mit den Single-Choice-Antwortoptionen „Erinnern und abrufen“ und „Verstehen und Anwenden“ (sowie einer Freitextoption für Kommentierungen).

<sup>6</sup> Itemwording: „Welchen Wissensbereich hat das Item primär zum Gegenstand?“ mit den Wissensbereichen je Domäne als Single-Choice Antwortoption (sowie einer Freitextoption für Kommentierungen).

**Tabelle 2.** Kennwerte der Skalierung des Wissenstests

Merkmal	<i>n</i>	Anzahl Items	EAP / PV	Varianz	wMNSQ	ID ( <i>M</i> )
Politik	371	16	.66	0.99	.87 – 1.10	.32
Wirtschaft	370	15	.66	1.06	.92 – 1.12	.31
Soziologie	374	17	.47	0.38	.95 – 1.08	.19
Fachdidaktik	366	15	.60	0.88	.90 – 1.08	.30

Anmerkungen: EAP / PV = Expected-A-Posteriori / Plausible Values Reliabilität; wMNSQ = Weighted Mean Square, ID (*M*) = Mittelwert der Item Discrimination.

Die Befragten im SoWis-L war zwischen den Erhebungsjahren für den Bereich Politik ( $t_{371} = 0.94, p = .346$ ) und Fachdidaktik ( $t_{368} = 0.44, p = .674$ ) vergleichbar, im Bereich Soziologie ( $t_{371} = 2.01, p = .045, d = 0.24$ ) und Wirtschaft ( $t_{371} = 2.40, p = .017, d = 0.28$ ) im Erhebungsjahr 2018 gegenüber 2019 erhöht. Über die Erfassung eines selbst generierten ID-Codes wurde sichergestellt, dass Studierende nicht doppelt in die Stichprobe dieser Auswertung eingingen. Das verwendete Testheftdesign (vgl. elektronisches Supplement (ESM) 3) kontrolliert Reihenfolge-, nicht jedoch Carry-Over-Effekte. Alle Testheftversionen enthielten eine vergleichbare Zahl von Items aller Wissensbereiche sowie jeweils Ankeritems. Um eine gleichmäßige und randomisierte Verteilung der verschiedenen Testhefte auf die Studierenden sicherzustellen, wurde das Verfahren des „Spiraling“ angewandt (Frey et al., 2009, S. 45). Die Teilnahme an der Befragung war freiwillig. Den Studierenden wurde die Möglichkeit einer individuellen Rückmeldung per Testbogen-Code eröffnet. Die Teilnahmemotivation war ausgesprochen hoch, was sich auch in dem sehr geringen Anteil fehlender Werte (0 % bis max. 3 % bei den Wissenfragen) dokumentiert.

## Ergebnisse

### Ergebnisse der IRT-Skalierung

Die Schätzung der Itemschwierigkeiten, Itemdiskriminationswerte und Expected-A-Posteriori / Plausible Values (EAP / PV)-Reliabilitätswerte erfolgte IRT-basiert mit der Software Conquest (vgl. Tabelle 2). Die Items wurden als Gesamttest und entlang der vier Wissensbereiche (Politik, Wirtschaft, Soziologie, Fachdidaktik) getrennt voneinander IRT-skaliert (da die Berechnung eines vierdimensionalen IRT-Modells aufgrund der Komplexität des Schätzverfahrens nicht möglich war). Der Weighted Mean Square (wMNSQ) lag für alle Items in einem guten Bereich von  $0.87 < wMNSQ < 1.12$ . Für die übrigen drei Kennwerte unterschieden sich die Befunde je nach Wissensbereich: Während für die Wissensbereiche Politik, Wirtschaft und Fachdidaktik mit mittleren Itemdiskriminationswerten

$\geq .30$  (Bond & Fox, 2007), Varianzen  $\geq .88$  und EAP-Reliabilität im Bereich  $.60 \leq \text{Rel (EAP)} \leq .66$  zufriedenstellende Eigenschaften gefunden wurden, waren die Kennwerte der Soziologieskala nicht akzeptabel. In Bezug auf F1 lässt sich feststellen, dass die Soziologieskala eine unzureichende Reliabilität aufweist und daher in diesem Beitrag von den weiteren Analysen ausgenommen wird. Die Wirtschafts-, Politik und Fachdidaktik-Skala ergeben befriedigende Reliabilitätswerte und sind entsprechend Gegenstand der weiteren Analysen.

Die zwei fachwissenschaftlichen Skalen (Politik, Wirtschaft) sowie die Fachdidaktikskala wurden einer Differential Item Functioning (DIF) Analyse unterzogen, um Testfairness zu prüfen. Die DIF-Analysen erfolgten mit Conquest auf Basis der einzelnen Skalen. Als Kriterium wurde ein Unterschied in den Itemparametern für die verglichenen Teilgruppen von mindestens 0.5 Logits angelegt (Adams & Carstensen, 2002). Überprüft wurde DIF nach Geschlecht, Schulform und Deutsch als Muttersprache. Es konnte kein systematischer DIF zu Lasten einer bestimmten Teilgruppe der Stichprobe festgestellt werden.

### Strukturanalysen

Zur Prüfung der Struktur des sozialwissenschaftlichen Professionswissens (F2) wurden mit der Software Mplus 8 (Muthén & Muthén, 1998–2017) mittels Weighted Least Squares Mean Variance- (WLSMV) Schätzern Strukturgleichungsmodelle berechnet (Kline, 2015) und drei Modelle geprüft (Tabelle 3). Modell M1 stellt das Generalfaktormodell dar, bei dem fachwissenschaftliches und fachdidaktisches Wissen eine gemeinsame Dimension bilden. Modell M2 dagegen differenziert zwischen Fachwissen und fachdidaktischem Wissen, M3 zusätzlich zwischen den beiden fachwissenschaftlichen Wissensbereichen Politik und Wirtschaft. Das Modell M3 entspricht damit den theoretischen Annahmen zur Struktur des Professionswissens im Lehramt SoWi. Der Modellvergleich erfolgte anhand der  $\chi^2$ -verteilten Differenz der Final Deviance sowie der Reliabilitätswerte. Der Abweichungsindex (Final Deviance) dient dabei als Indikator für die glo-

Tabelle 3. Modellvergleich

	Final Deviance	df	EAP/PV-Reliabilität
Generalfaktor M1	13 867.57	47	.80
Zweifaktorielles Modell M2 (Fachwissen, Fachdidaktik)	13 821.14	49	.79 .70
Dreifaktorielles Modell M3 (Politik, Wirtschaft, Fachdidaktik)	13 774.19	52	.73 .74 .70

Anmerkungen: EAP / PV = Expected-A-Posteriori / Plausible Values Reliabilität.

bale Modellgüte. Das zweifaktorielle Modell M2 zeigte gegenüber dem Generalfaktormodell M1 eine statistisch signifikante Modellverbesserung ( $\Delta\chi^2 = 46.43$ ,  $df = 2$ ,  $p < .001$ ). Das dreifaktorielle Modell M3 führte sowohl gegenüber dem Generalfaktormodell ( $\Delta\chi^2 = 93.38$ ,  $df = 5$ ,  $p < .001$ ) als auch gegenüber dem zweifaktoriellen Modell M2 ( $\Delta\chi^2 = 46.95$ ,  $df = 3$ ,  $p < .001$ ) zu einer signifikanten Modellverbesserung. Die EAP-Reliabilitätswerte aller drei Modelle liegen mit Werten  $\geq .70$  in einem zufriedenstellenden Bereich. Die Ergebnisse sprechen insgesamt für eine dreifaktorielle Skalierung.

Zusätzlich wurden die Strukturannahmen im Rahmen einer konfirmatorischen Faktorenanalyse mit der Software Mplus 8 geprüft. Zur Beurteilung der Modellpassung wurden die Modellfit-Indizes Root Mean Square Error of Approximation (RMSEA), Confirmatory Fit Index (CFI) und Standardized Root Mean Squared Residual (SRMR) herangezogen, wobei RMSEA-Werte kleiner als .05 und CFI-Werte größer als .95 eine gute Modellpassung und RMSEA-Werte kleiner .08 und CFI-Werte größer als .90 eine akzeptable Modellpassung anzeigen und der SRMR Werte kleiner als .08 annehmen sollte (Hu & Bentler, 1999; Xia & Yang, 2018). Auch hier wurde ein Modellvergleich auf Basis von Modellfitwerten vorgenommen und das dreidimensionale Modell bestätigt (vgl. ESM 4). Die Modellfitwerte des dreifaktoriellen SoWis-L Modells sind vor dem Hintergrund der von Hu & Bentler (1999) angeführten Cut-Off-Werte als zufriedenstellend zu interpretieren,  $\chi^2 = 854.34$ ,  $df = 762$ ;  $p = .011$ ; CFI = .909; RMSEA = 0.019, SRMR = .064 (vgl. ESM 5). Auch die latenten Korrelationen  $.628 \geq r \leq .708$  deuten auf empirisch unterscheidbare Wissensbereiche hin. Erwartungsgemäß sind die beiden fachwissenschaftlichen Wissensbereiche mit  $r = .708$ ,  $p < .001$  etwas stärker assoziiert als die beiden Fachwissensdimensionen mit dem fachdidaktischen Wissen. Die Faktorladungen sollten Werte von  $\lambda \geq .50$  annehmen, um eine Varianzaufklärung von 25% oder höher zu erreichen. Beim politikwissenschaftlichen Wissen liegen die Ladungen bei  $M = .404$  (min = .290, max = .694), beim wirtschaftswissenschaftlichen Wissen bei  $M = .385$  (min = .193, max = .533) und beim fachdi-

daktischen Wissen bei  $M = .377$  (min = .209, max = .533). Die Faktorladungen sind insgesamt als gering anzusehen.

ESM 6 zeigt die Verteilung der Personenfähigkeiten und Itemschwierigkeiten in einer Person-Item Map. Die Verteilung der Itemschwierigkeiten deckt die Personenfähigkeiten insgesamt gut ab. Dies gilt vor allem für den mittleren Fähigkeitsbereich, während nur wenige sehr schwierige oder sehr leichte Items vorliegen. Die Lösungsquote  $p$  der Items variiert zwischen  $21\% \leq p \leq 95\%$ , der Itemschwierigkeitsparameter  $\sigma$  zwischen  $-2.772 \leq \sigma \leq 1.864$ . Die mittlere geschätzte Personenfähigkeit  $j$  beträgt beim Professionswissen Politik  $j = .236$ , in der Fachdidaktik  $j = .347$  und beim Professionswissen Wirtschaft  $j = .544$ .

### Zusammenhänge mit weiteren Variablen und Mittelwertvergleiche für bekannte Gruppen

F3 bezieht sich auf die Validierung des Testinstruments durch die Methode der bekannten Gruppen. Untersucht wird der Wissenszuwachs zwischen BA- und MA-Studium (im Querschnitt) sowie Wissensunterschiede zwischen GyGe- und HRG-Studierenden. Zunächst gilt: Die GyGe-Studierenden erreichten in allen Wissensbereichen durchschnittlich bessere Testleistungen als die HRG-Studierenden. Dies gilt für die Bereiche Politik ( $t_{347} = 4.42$ ,  $p < .001$ ,  $d = 0.57$ ), Wirtschaft ( $t_{347} = 3.68$ ,  $p < .001$ ,  $d = 0.48$ ) und Fachdidaktik ( $t_{347} = 1.69$ ,  $p = .046$ ,  $d = 0.22$ ). Diese Ergebnisse sind in hohem Maße erwartungskonform. Tabelle 4 zeigt Gruppenunterschiede in der Testleistung (normierte WLE-Scores) zwischen BA- und MA-Studierenden, getrennt nach Wissensbereich und Lehramtsstudiengang (GyGe vs. HRG). Ein deutlicher Wissensunterschied zeigt sich im Bereich der Fachdidaktik, sowohl bei GyGe-Studierenden ( $t_{266} = 5.26$ ,  $p < .001$ ,  $d = 0.64$ ) als auch bei HRG-Studierenden ( $t_{74} = 3.45$ ,  $p < .001$ ,  $d = 0.80$ ). Ebenso erreichten die Befragten im MA eine höhere Testleistung Politik als solche im BA. Für das Lehramt HRG liegt ein mittelgroßer signifikanter Effekt

**Tabelle 4.** Gruppenunterschiede in der Testleistung (normierte WLE-Scores) zwischen BA- und MA-Studierenden, getrennt nach Wissensbereich und Lehramtsstudiengang (GyGe vs. HRG)

	<i>n</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t (df)</i>	<i>p</i>	<i>d</i>
<b>Testleistung Politik</b>							
Lehramt HRG							
Bachelor	44	43.24	10.34	1.56	2.22 (74)	.015	0.52
Master	32	48.18	8.36	1.48			
Lehramt GyGe							
Bachelor	128	50.15	8.74	0.77	1.82 (266)	.035	0.22
Master	140	52.21	9.72	0.82			
<b>Testleistung Wirtschaft</b>							
Lehramt HRG							
Bachelor	44	46.35	8.51	1.28	0.29 (74)	.387	0.07
Master	32	45.79	8.29	1.47			
Lehramt GyGe							
Bachelor	128	51.17	9.58	0.85	0.02 (266)	.491	0.00
Master	140	51.15	10.58	0.89			
<b>Testleistung Fachdidaktik</b>							
Lehramt HRG							
Bachelor	44	45.36	8.03	1.21	3.45 (74)	< .001	0.80
Master	32	51.95	8.51	1.50			
Lehramt GyGe							
Bachelor	128	47.35	10.63	0.94	5.26 (266)	< .001	0.64
Master	140	53.63	8.91	0.75			

Anmerkungen: Normierte WLE (Weighted Likelihood Estimates)-Scores ( $M = 50$ ,  $SD = 10$ );  $d$  = Effektstärke; Lehramt HRG = Lehramt für Haupt-, Real- und Gesamtschule; Lehramt GyGe = Lehramt für Gymnasien und Gesamtschulen;  $p$ -Werte auf Basis einseitiger  $t$ -Tests ermittelt.

vor ( $t_{74} = 2.22$ ,  $p = .015$ ,  $d = 0.52$ ), für GyGe ein kleiner signifikanter Effekt ( $t_{266} = 1.82$ ,  $p = .035$ ,  $d = 0.22$ ). Einzig bei der Testleistung Wirtschaft ist keine Differenz zwischen BA- und MA-Studierenden feststellbar. Insgesamt stützen diese Ergebnisse die angenommenen Gruppenunterschiede (Kriteriumsvalidität). Eine Kontraststichprobe von 12 befragten Lehrkräften im Schuldienst (davon 3 mit Gesamtschul- und 7 mit Gymnasiallehramt) erreichte darüber hinaus eine signifikant höhere Testleistung als die Studierendenstichprobe in Politik ( $t_{383} = 5.37$ ,  $p < .001$ ,  $d = 1.575$ ), Wirtschaft ( $t_{13,11} = 9.82$ ,  $p < .001$ ,  $d = 1.442$ ) und Fachdidaktik ( $t_{380} = 6.59$ ,  $p < .001$ ,  $d = 1.932$ ), was als Hinweis auf eine Sensitivität des Testinstruments für Expertiseunterschiede verstanden werden kann.

F4 bezieht sich auf Zusammenhänge der Testleistung im SoWis-L mit Außenkriterien, hier den selbstberichteten durchschnittlichen Prüfungsnoten je Wissensbereich sowie der Abiturnote. Die Prüfungsnoten lagen im Wissensbereich Politik bei  $M = 2.64$  ( $SD = 0.78$ ), in Wirtschaft bei  $M = 2.85$  ( $SD = 0.96$ ) und in der Fachdidaktik bei  $M = 3.16$  ( $SD = 1.10$ ). Die Abiturnote betrug  $M = 2.28$  ( $SD = 0.46$ ). Die Testleistungen im SoWis-L sind in allen drei Wissensbereichen signifikant negativ mit der Abiturnote verknüpft, am stärksten im Bereich des wirtschaftli-

chen Wissens ( $r = -.27$ ,  $p < .001$ ), gefolgt vom politischen Wissen ( $r = -.23$ ,  $p < .001$ ) und am geringsten im Bereich des fachdidaktischen Wissens ( $r = -.16$ ,  $p = .004$ ). Das bedeutet, umso niedriger (besser) die Abiturnote, desto höher das jeweilige Testergebnis. Dieses Ergebnis ist hypothesenkonform.

Die Zusammenhänge zwischen Abiturnote, selbstberichteten Prüfungsnoten und Testleistung im SoWis-L stellen sich anders dar als erwartet. Zunächst ist festzustellen, dass die Prüfungsnoten nicht mit der Abiturnote korrelieren. Auch zeigen sich in den Fachwissenschaften Politik und Wirtschaft keine Korrelationen zwischen Prüfungsnote und Testleistung. Lediglich die Testleistung in der Fachdidaktik ist schwach negativ mit der selbstberichteten durchschnittlichen Prüfungsnote in der Fachdidaktik assoziiert ( $r = -.17$ ,  $p < .01$ ).

## Diskussion

In diesem Beitrag wurden die Konstruktion und Validierung eines Testinstruments zur Erfassung des Professionswissens von Lehrkräften im Fach SoWi vorgestellt.

Das Testinstrument SoWis-L wurde an 374 SoWi-Lehramtsstudierenden der UDE erprobt. Vorgestellt wurden das Konstruktionsrational und Inhaltsbereiche des SoWis-L sowie empirische Befunde zu strukturellen und externalen Validitätsaspekten. Der SoWis-L ist nach unserer Kenntnis das bislang einzig standardisierte Testinstrument für den Bereich des Professionswissens im Lehramt SoWi, das den Integrationscharakter des Fachs abbildet. Als innovativ kann gelten, dass der politische und wirtschaftliche Wissensbereich sowie die Fachdidaktik getrennt modellierbar sind und so Forschungsfragen zum Professionswissen von Lehrkräften mit höherem Differenzierungsgrad untersuchbar sind als bislang. Eben dieser Integrationsfachcharakter sowie die geringe Kanonisierung von Studieninhalten in diesem Lehramtsstudienfach stellen auch eine Herausforderung bei der Testentwicklung dar. Dieser Beitrag präsentiert erste Evidenzen zur Validierung des SoWis-L. Es sind jedoch weitere Analysen an größeren Stichproben erforderlich, um die Validität des Instruments bezüglich der intendierten Interpretation und Nutzung umfänglich abschätzen zu können (vgl. American Educational Research Association (AERA) et al., 2014). Im Folgenden werden die wesentlichen Ergebnisse bezüglich der Forschungsfragen und damit verbundene Limitationen des Instruments diskutiert.

F1 bezog sich auf die Reliabilität des neuen Testinstruments. Die beiden fachwissenschaftlichen Skalen des Professionswissens Politik und Wirtschaft sowie die Fachdidaktik zeigten bei durchschnittlich geringen Faktorladungen insgesamt zufriedenstellende Reliabilitätswerte. Dies ist besonders vor dem Hintergrund der vergleichsweise geringen Strukturierung des sozialwissenschaftlichen Wissens positiv zu bewerten. Demgegenüber ist es für den Bereich des soziologischen Fachwissens nicht hinreichend gelungen, eine reliable Skala zu entwickeln. Die konstruierten Wissensitems sollten zur weiteren Überprüfung zukünftig an einer erweiterten bzw. zusätzlichen Stichprobe erprobt werden und ggf. für den weiteren Einsatz überarbeitet werden. Im Kontext dieses Beitrags können nur Vermutungen über Ursachen angestellt werden. So kann der Befund durch Besonderheiten der soziologischen Domäne und der vergleichsweise geringen Kanonisierung der entsprechenden Studieninhalte (z. B. Kühl, 2003) durch eine inhaltlich nicht repräsentative Schwerpunktsetzung bei der Itemkonstruktion oder eine Kombination dieser Faktoren bedingt sein. Festzustellen bleibt weiterhin ein Desiderat an Testinstrumenten zur Erfassung soziologischen Wissens (nicht nur in der Lehrkräftebildung; auch allgemein ist den Autor\_innen der Studie kein solches Instrument bekannt).

Zweitens sollte geprüft werden, wie sich die Struktur des fachbezogenen Professionswissens empirisch darstellt (F2). Es wurde angenommen, dass sich die theoretisch

hergeleitete Trennung zwischen Fachwissen und fachdidaktischem Wissen sowie die disziplinäre Trennung innerhalb der fachwissenschaftlichen Bereiche empirisch abbilden. Erwartungskonform ließ sich das Professionswissen als mehrdimensionales Konstrukt darstellen, wobei die Ausdifferenzierung der beiden Fachwissensbereiche Politik und Wirtschaft zu einer besseren Passung führte. Auch die berichteten Korrelationen zwischen den drei Skalen sprechen für eine getrennte Modellierung. Erwartungskonform korrelierten die beiden fachwissenschaftlichen Wissensskalen stärker miteinander als mit der fachdidaktischen Skala. SoWis-L bietet gegenüber vorliegenden Testinstrumenten den Mehrwert, das Professionswissen im Fach SoWi getrennt nach politischen, wirtschaftlichen und fachdidaktischen Wissensbereichen erheben zu können. Für Erhebungen in Bundesländern mit anderem Fachzuschnitt besteht die Möglichkeit, bedarfsgerecht nur einzelne Skalen des SoWis-L einzusetzen.

Die Validierungsstrategie der Methode der bekannten Gruppen ergab weitgehend erwartungskonforme Resultate (F3). Die Studierenden eines GyGe-Lehramts erreichten in den beiden fachwissenschaftlichen Wissensbereichen signifikant bessere Testleistungen als Studierende eines HRG-Lehramts. Weiter zeigten MA-Studierende der verschiedenen Lehramtsstudiengänge höhere Testleistungen in der Fachdidaktik und im politischen Wissen. Erwartungswidrig galt dies nicht für den wirtschaftlichen Bereich. Hier ist zu berücksichtigen, dass sich in der Stichprobe eine höhere Testleistung im Erhebungsjahr 2018 gegenüber 2019 für den wirtschaftlichen Bereich zeigte, so dass von Leistungsunterschieden zwischen den Kohorten in diesem Inhaltsbereich ausgegangen werden muss. Der SoWis-L Test erfragt grundlegendes Professionswissen, welches überwiegend bereits in den Einführungsmodulen des BA-Studiums erworben werden kann. Dies bietet eine Erklärung für die insgesamt eher mäßige Wissenszunahme zwischen BA- und MA-Studierenden. Eine Leistungssteigerung würde sich demnach weniger in einer Zunahme des grundlegenden Wissens, sondern in der Qualität der Anwendung des erworbenen Wissens bspw. in Urteils- oder Problemlösungsaufgaben niederschlagen. Auch ist denkbar, dass ein Teil des Professionswissen außerhalb des Studiums, bspw. durch regelmäßigen Nachrichtenkonsum, erworben wird. Dies könnte in zukünftigen Erhebungen durch die Berücksichtigung entsprechender Fragebogenskalen zum Informationsverhalten weiter untersucht werden.

F4 bezog sich auf eine Validierung mittels der Außenkriterien Abiturnote und selbstberichteten Prüfungsnoten. Erwartungskonforme Zusammenhänge zeigten sich zwischen allen Bereichen des SoWis-L-Tests und der Abiturnote. Jedoch stand die fachwissenschaftliche Testleis-

tung in keinem systematischen Zusammenhang zu den Prüfungsnoten. Dass die Prüfungsnoten nicht mit der Abiturnote assoziiert waren, kann darauf hinweisen, dass die Testleistung den valideren Indikator für grundlegendes Wissen in dem jeweiligen Wissensbereich darstellt. Zu berücksichtigen ist ferner, dass die Prüfungsnoten gegenüber dem SoWis-L eine deutlich geringere Varianz aufwiesen und Selbstauskünfte der Studierenden darstellen.

Abschließend sind noch folgende Limitationen und damit Desiderate für zukünftige Forschungen festzuhalten. Die Orientierung an Kerninhalten der entsprechenden Lehramtsstudiengänge auf Basis einer umfangreichen Sichtung curricularer Dokumente spricht für eine Passung des Instruments auch auf andere Universitätsstandorte. Aufgrund möglicher Unterschiede im implementierten Curriculum der Bundesländer und Hochschulen kann dies jedoch nur durch empirischen Prüfung geklärt werden (z. B. Expertenrating zur Relevanz und Repräsentativität der Testinhalte des SoWis-L für das zu erwerbende Professionswissen im Lehramt SoWi). Zum zweiten gilt: Die in allen Wissensbereichen signifikant höheren Testscores der befragten Lehrkräfte vs. Studierende des Lehramts SoWi deuten auf eine Sensitivität des Instruments für Expertise-Unterschiede hin. Jedoch sollte der SoWis-L durch Kontrastgruppenvergleiche, speziell mit einer größeren Stichprobe von Lehrkräften, künftig noch umfassender untersucht werden. Schließlich ist zu berücksichtigen, dass der SoWis-L die inhaltliche Breite des sozialwissenschaftlichen Professionswissens nicht erschöpfend abdeckt. Das Instrument ist nicht für die Individualdiagnostik konstruiert, sondern ermöglicht eine zeitökonomische Messung des Professionswissens im Fach SoWi (bspw. für Gruppenvergleiche) unter Berücksichtigung des Integrationsfachcharakters.

## Elektronische Supplemente (ESM)

Die elektronischen Supplemente sind mit der Online-Version dieses Artikels verfügbar unter <https://doi.org/10.1026/0012-1924/a000313>

**ESM 1.** Items nach Konstruktionsmerkmalen

**ESM 2.** Itembeispiele

**ESM 3.** Testheftdesign

**ESM 4.** Modellvergleich

**ESM 5.** Dreifaktorielles SoWis-L Modell

**ESM 6.** Wright-Map

## Literatur

- Abs, H. J. & Hahn-Laudenberg, K. (Hrsg. 2017). *Das politische Mindset von 14-Jährigen. Ergebnisse der International Civic and Citizenship Education Study 2016*. Münster: Waxmann.
- Adams, R. J. & Carstensen, C. H. (2002). Scaling Outcomes. In R. Adams & M. Wu (Eds.), *PISA 2000 Technical Report* (pp. 149–162). Paris: OECD.
- AERA, APA & NCME (2014). *Standards for educational and psychological testing*. Washington: American Educational Research Association.
- Anderson, L.W., Krathwohl, D.R. & Bloom, B.S. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Bank, V. & Retzmann, T. (2012). *Fachkompetenz von Wirtschaftslehrern. Grundlagen und Befunde einer Weiterbildungsbedarfsanalyse*. Schwalbach/Taunus: Wochenschau.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9, 469–520. <https://doi.org/10.1007/s11618-006-0165-2>
- Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29–53). Münster: Waxmann.
- Beck, K., Krumm, V. & Dubs, R. (1998). *Wirtschaftskundlicher Bildungs-Test (WBT)*. Göttingen: Hogrefe.
- Blömeke, S. & Zlatkin-Troitschanskaia, O. (2013). *Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor: Ziele, theoretischer Rahmen, Design und Herausforderungen des BMBF-Forschungsprogramms KoKoHs* (KoKoHs Working Papers, 1). Berlin/Mainz: Humboldt-Universität/Johannes Gutenberg-Universität.
- Bloom, B. (1956). *Taxonomy of Educational Objectives. The Classification of Educational Goals, Handbook I: Cognitive Domain*. New York: David McKay Co Inc.
- Bond, T. & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Delli Carpini, M. X. & Keeter, S. (1993). Measuring political knowledge: Putting first things first. *American Journal of Political Science* 37(4), 1 179–1 206. <https://doi.org/10.2307/2111549>
- Deutsche Gesellschaft für Soziologie (2005). *Empfehlungen der Deutschen Gesellschaft für Soziologie (DGS) zur Ausgestaltung soziologischer Bachelor- und Masterstudiengänge*. Verfügbar unter <http://www.soziologie.de/uploads/media/BA-MA-Studienempfehlungen-DRUCKF-051212.pdf> [13.01.2022]
- Deutsche Vereinigung für Politische Wissenschaft (DVPW) (2003). *Gestufte Studiengänge in der Politikwissenschaft – Empfehlungen zu einem Kerncurriculum von Vorstand und Beirat der DVPW*. Verfügbar unter <https://www.dvpw.de/fileadmin/docs/Kerncurriculum.pdf> [13.01.2022].
- Eveland, W. P. & Hively, M. H. (2008). Political knowledge. In W. Donsbach (Ed.), *The International encyclopedia of communication* (pp. 3 715–3 719). Malden, MA: Blackwell.
- Förster, M., Happ, R. & Zlatkin-Troitschanskaia, O. (2012). Valide Erfassung des volkswirtschaftlichen Fachwissens von Studierenden der Wirtschaftswissenschaften und der Wirtschaftspädagogik – Eine Untersuchung der diagnostischen Eignung des Wirtschaftskundlichen Bildungstests (WBT). *bwp@Berufs- und Wirtschaftspädagogik – online*, 22, 1–21.
- Frey, A., Hartig, J. & Rupp, A. A. (2009). An NCME instructional module on booklet designs in large-scale assessments of student achievement: Theory and practice. *Educational Measure-*

- ment: *Issues and Practice*, 28(3), 39–53. <https://doi.org/10.1111/j.1745-3992.2009.00154.x>
- Fürntratt, E. & Jäger, A. O. (1996). *Differentieller Wissenstest*. Göttingen: Hogrefe.
- GFD (2004). *Kerncurriculum Fachdidaktik. Orientierungsrahmen für alle Fachdidaktiken*. Verfügbar unter <https://www.fachdidaktik.org/veroeffentlichungen/positionspapiere-der-gfd/> [13.01.2022].
- Großschedl, J., Neubrand, C., Kirchner, A., Oppermann, L., Basel, N. & Gantner, S. (2015). Entwicklung und Validierung eines Testinstruments zur Erfassung des evolutionsbezogenen Professionswissens von Lehramtsstudierenden (ProWiE). *Zeitschrift für Didaktik der Naturwissenschaften*, 21, 173–185. <https://doi.org/10.1007/s40573-015-0036-6>
- Hossiep, R. & Schulte, M. (2007). *BOWIT – Bochumer Wissenstest*. Göttingen: Hogrefe.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Jennings, M. K. (1996). Political knowledge over time and across generations. *Public Opinion Quarterly*, 60, 228–252. <https://doi.org/10.1086/297749>
- Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Kauertz, A., Fischer, H. E., Mayer, J., Sumfleth, E. & Walpuski, M. (2010). Standardbezogenen Kompetenzmodellierung in den Naturwissenschaften der Sekundarstufe I. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 135–153.
- Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4rd ed.). New York, NY: The Guilford Press.
- Köller, O., Pant, H. A. & Zlatkin-Troitschanskaia, O. (2020). Diagnostik von studentischen Kompetenzen im Hochschulsektor [Themenheft]. *Diagnostica*, 66(2). <https://econtent.hogrefe.com/toc/dia/66/2>
- Krauss, S., Neubrand, M., Blum, W., Baumert, J., Brunner, M., Kunter, M. et al. (2008). Die Untersuchung des professionellen Wissens deutscher Mathematik-Lehrerinnen und -Lehrer im Rahmen der COACTIV-Studie. *Journal für Mathematik-Didaktik*, 29(3/4), 223–258.
- Kultusministerkonferenz (2008). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerinnen- und Lehrerbildung*. Beschluss der Kultusministerkonferenz vom 16.10.2008 i.d.F. vom 10.09.2015. Verfügbar unter [http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2008/2008\\_10\\_16-Fachprofile-Lehrerbildung.pdf](http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf) [20.07.2016]
- Kühl, S. (2003). Wie verwendet man Wissen, das sich gegen die Verwendung sträubt? Eine professionssoziologische Neubetrachtung der Theorie-Praxis-Diskussion in der Soziologie. In J. Howaldt (Hrsg.), *Neue Formen sozialwissenschaftlicher Wissensproduktion*. Berlin: Sigma.
- Landwehr, B. (2017). *Partizipation, Wissen und Motivation im Politikunterricht. Eine Interventionsstudie*. Wiesbaden: Springer. <https://doi.org/10.1007/978-3-658-16507-9>.
- Manzel, S. & Gronostay, D. (2018). Weiblich, Realschule, pädagogisches Interesse – männlich, Gymnasium, höheres Fachinteresse? In S. Manzel & M. Oberle (Hrsg.), *Kompetenzorientierung. Potentiale zur Professionalisierung der politischen Bildung* (S. 73–83). Wiesbaden: Springer VS. [https://doi.org/10.1007/978-3-658-16889-6\\_7](https://doi.org/10.1007/978-3-658-16889-6_7)
- Messick, S. (1989). Validity. In R. Linn (Hrsg.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Muthén, L. K. & Muthén, B. (1998–2017). *Mplus* (Version 8). Los Angeles, CA: Muthén & Muthén.
- Neuman, R. W. (1986). *The Paradox of Mass Politics. Knowledge and Opinion in the American Electorate*. Cambridge: Harvard University Press.
- 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, 261–284. <https://doi.org/10.1007/s11165-007-9049-6>
- Richter, D., Goll, T., Weißeno, G. & Eck, V. (2010). Gibt es Unterschiede im politischen Wissen zwischen Schülerinnen und Schülern mit und ohne Migrationshintergrund? (POWIS-Studie). In I. Juchler (Hrsg.), *Kompetenzen in der politischen Bildung* (S. 207–217). Schwalbach/Taunus: Wochenschau.
- Schübel, T. (2018). *Die soziale Verteilung politischen Wissens in Deutschland. Wissensunterschiede und deren Ursachen*. Wiesbaden: Springer VS.
- Shulman, L. S. (1986). Those who understand. Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–22. <https://doi.org/10.17763/haer.57.1.j463w79r56455411>
- Trepte, S., Loy, L. S., Schmitt, J. B. & Otto, S. (2017). Hohenheimer Inventar zum Politikwissen (HIP): Konstruktion und Skalierung. *Diagnostica*, 63, 206–218. <https://doi.org/10.1026/0012-1924/a000180>
- Trepte, S. & Verbeet, M. (2010). *Allgemeinbildung in Deutschland*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Weißeno, G., Weschenfelder, E. & Oberle, M. (2013). Empirische Ergebnisse zur professionellen Kompetenz von Politiklehrer/-innen. In K. P. Hufer & D. Richter (Hrsg.), *Politische Bildung als Profession. Verständnisse und Forschungen* (S. 187–202). Bonn: Bundeszentrale für politische Bildung.
- Weißeno, G., Detjen, J., Juchler, I., Massing, P. & Richter, D. (2010). *Konzepte der Politik – ein Kompetenzmodell*. Bonn: Bundeszentrale für politische Bildung.
- Weschenfelder, E. (2014). *Professionelle Kompetenz von Politiklehrkräften. Eine Studie zu Wissen und Überzeugungen*. Wiesbaden: Springer VS. <https://doi.org/10.1007/978-3-658-04193-9>
- Westle, B. & Tausendpfund, M. (Hrsg.) (2019). *Politisches Wissen. Relevanz, Messung und Befunde*. Wiesbaden: Springer VS. <https://doi.org/10.1007/978-3-658-23787-5>
- Wuttke, E. & Beck, K. (2002). Eingangsbedingungen von Studienanfängern – Die Prognostische Validität wirtschaftskundlichen Wissens für das Vordiplom bei Studierenden der Wirtschaftswissenschaften. *Arbeitspapiere WP 41*. Universität Mainz, [https://download.uni-mainz.de/fb03-wipaed/ArbeitspapiereWP/gr\\_Nr.41.pdf](https://download.uni-mainz.de/fb03-wipaed/ArbeitspapiereWP/gr_Nr.41.pdf)
- Xia, Y. & Yang, Y. (2018). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, 51, 409–428. <https://doi.org/10.3758/s13428-018-1055-2>
- Zlatkin-Troitschanskaia, O., Jitomirski, J., Happ, R., Molerov, D., Schlax, J., Kühling-Thees, C., Förster, M. & Brückner, S. (2019). Validating a test for measuring knowledge and understanding of economics among university students. *Zeitschrift für Pädagogische Psychologie*, 33(2), 119–133. <https://doi.org/10.1024/1010-0652/a000239>
- Zlatkin-Troitschanskaia, O., Pant, H. A., Toepper, M., Lautenbach, C. & Molerov, D. (2017). Valid competency assessment in higher

education – Framework, results, and further perspectives of the German research program KoKoHs. *AERA Open*, 3(1), 1 – 22. <https://doi.org/10.1177/2332858416686739>

### Historie

Onlineveröffentlichung: 14.02.2023

### Danksagung

Wir bedanken uns herzlich bei den Gutachterinnen und Gutachtern für die kritisch-konstruktiven Rückmeldungen zu einer früheren Version dieses Manuskripts.

### Förderung

ProViel wird im Rahmen der gemeinsamen „Qualitätsoffensive Lehrerbildung“ von Bund und Ländern aus Mitteln des Bundesministeriums für Bildung und Forschung gefördert (FKZ 01JA1910). Open Access-Veröffentlichung ermöglicht durch die Technische Universität Dortmund.

### ORCID

Dorothee Gronostay

 <https://orcid.org/0000-0002-1873-5123>

Frank Eike Zischke

 <https://orcid.org/0000-0002-2847-9119>

### Jun.-Prof. Dr. Dorothee Gronostay

Institut für Didaktik integrativer Fächer (IDIF)

Technische Universität Dortmund

August-Schmidt-Str. 6 (GB I, Raum 306)

44227 Dortmund

Deutschland

[dorothee.gronostay@tu-dortmund.de](mailto:dorothee.gronostay@tu-dortmund.de)