

Forster-Heinzer, Sarah; Oser, Fritz

Wer setzt das Maß? Eine kritische Auseinandersetzung mit dem Advokatorischen Ansatz

Zeitschrift für Pädagogik 61 (2015) 3, S. 361-376



Quellenangabe/ Reference:

Forster-Heinzer, Sarah; Oser, Fritz: Wer setzt das Maß? Eine kritische Auseinandersetzung mit dem Advokatorischen Ansatz - In: Zeitschrift für Pädagogik 61 (2015) 3, S. 361-376 - URN: urn:nbn:de:0111-pedocs-153670 - DOI: 10.25656/01:15367

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-153670>

<https://doi.org/10.25656/01:15367>

in Kooperation mit / in cooperation with:

BELTZ JUVENTA

<http://www.juventa.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

ZEITSCHRIFT FÜR PÄDAGOGIK

Heft 3

Mai/Juni 2015

■ *Thementeil*

Kontextualisierte Erfassung von Lehrerkompetenzen

■ *Essay*

Bildungsforschung im Kontext gesellschaftlicher Praxis. Über (soziale) Bedingungen der Möglichkeit, Bildungspraxis durch Bildungsforschung zu beeinflussen

■ *Allgemeiner Teil*

Large Scale Assessments und die Bildung Erwachsener: Erträge, Grenzen und Potenziale der Forschung

Heimann, Schulz oder Klafki? Eine quantitative Studie zur Einschätzung der Praktikabilität allgemeindidaktischer Planungsmodelle

Inhaltsverzeichnis

Thementeil: Kontextualisierte Erfassung von Lehrerkompetenzen

Johannes König

Kontextualisierte Erfassung von Lehrerkompetenzen.

Einführung in den Thementeil 305

Sigrid Blömeke/Johannes König/Ute Suhl/Jessica Hoth/

Martina Döhrmann

Wie situationsbezogen ist die Kompetenz von Lehrkräften?

Zur Generalisierbarkeit der Ergebnisse von videobasierten Performanztests 310

Andreas Busse/Gabriele Kaiser

Wissen und Fähigkeiten in Fachdidaktik und Pädagogik:

Zur Natur der professionellen Kompetenz von Lehrkräften 328

Kathleen Stürmer/Tina Seidel/Olga Kunina-Habenicht

Unterricht wissensbasiert beobachten – Unterschiede und

erklärende Faktoren bei Referendaren zum Berufseinstieg 345

Sarah Forster-Heinzer/Fritz Oser

Wer setzt das Maß? Eine kritische Auseinandersetzung

mit dem Advokatorischen Ansatz 361

Georg Hans Neuweg

Kontextualisierte Kompetenzmessung – Eine Bilanz

zu aktuellen Konzeptionen und forschungsmethodischen Zugängen 377

Deutscher Bildungsserver

Linktipps zum Thema „Kontextualisierte Erfassung

von Lehrerkompetenzen“ 384

Essay

Helmut Heid

Bildungsforschung im Kontext gesellschaftlicher Praxis. Über (soziale) Bedingungen der Möglichkeit, Bildungspraxis durch Bildungsforschung zu beeinflussen	390
------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----

Allgemeiner Teil

Josef Schrader

Large Scale Assessments und die Bildung Erwachsener: Erträge, Grenzen und Potenziale der Forschung	410
-------------------------------------------------------------------------------------------------------------	-----

Stephan Wernke/Jochen Werner/Klaus Zierer

Heimann, Schulz oder Klafki? Eine quantitative Studie zur Einschätzung der Praktikabilität allgemeindidaktischer Planungsmodelle	429
-------------------------------------------------------------------------------------------------------------------------------------------	-----

Besprechungen

Heinz-Elmar Tenorth

Johannes Drerup: Paternalismus, Perfektionismus und die Grenzen der Freiheit	452
---------------------------------------------------------------------------------------	-----

Verónica Oelsner

Phillip Dylan Thomas Knobloch: Pädagogik in Argentinien. Eine Untersuchung im Kontext Lateinamerikas mit Methoden der Vergleichenden Erziehungswissenschaft	455
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----

Dokumentation

Pädagogische Neuerscheinungen	459
Impressum	U3

Table of Contents

Topic: Contextualized Assessment of Teacher Competences

Johannes König

Contextualized Assessment of Teacher Competences. An introduction 305

Sigrid Blömeke/Johannes König/Ute Suhl/Jessica Hoth/Martina Döhrmann

To What Extent Is Teacher Competence Situation-Related?

On the generalizability of the results of video-based performance tests 310

Andreas Busse/Gabriele Kaiser

Knowledge and Skills in Subject-Related Didactics and Pedagogics –

On the nature of the professional competence of teachers 328

Kathleen Stürmer/Tina Seidel/Olga Kunina-Habenicht

Knowledge-Based Classroom Observation – Differences between
trainee teachers at the start of their professional career

and explanatory factors 345

Sarah Forster-Heinzer/Fritz Oser

Who Sets the Benchmark? A critical discussion of the advocacy approach ... 361

Georg Hans Neuweg

Contextualized Competence Measurement – Taking stock

of current conceptions and research-methodological approaches 377

Deutscher Bildungsserver

Tips of links relating to the topic of “Contextualized Assessment

of Teacher Competences” 384

Contributions

Helmut Heid

Educational Research in the Context of Social Practice –

On the (social) conditions for the possibility of influencing

the educational practice through educational research 390

Josef Schrader

Large-Scale Assessments and the Education of Adults:

Results, limits, and potentials of research 410

Stephan Wernke/Jochen Werner/Klaus Zierer

Heimann, Schulz, or Klafki? – A quantitative study on the assessment of the practicability of general didactical planning models	429
Book Reviews	452
New Books	459
Impressum	U3

Sarah Forster-Heinzer/Fritz Oser

Wer setzt das Maß?

Eine kritische Auseinandersetzung mit dem Advokatorischen Ansatz

Zusammenfassung: In Forschung und Praxis ist das Konstrukt der professionellen Kompetenz allgegenwärtig. Allerdings bleibt oft diffus, was zu diesem Konstrukt gehört. Begünstigt durch eine einseitige Fokussierung auf Wissensaspekte wird es nicht angemessen abgebildet. Folgender Beitrag geht in einem ersten Teil kritisch auf den Kompetenzbegriff ein. Es wird für eine ganzheitliche und kontextspezifische Betrachtung von Professionskompetenzen argumentiert. Im zweiten Teil wird der Advokatorische Ansatz vorgestellt. Dieser setzt Videovignetten ein, um festzustellen, wie sensibel Lehrpersonen die Qualität einer ausgeübten professionellen Handlung einschätzen können. An einem konkreten Beispiel wird aufgezeigt, wie die Kompetenzsensibilität gemessen wird. Durchgeführte IRT-Analysen zur Überprüfung der psychometrischen Güte verweisen auf einige Schwierigkeiten, die wir kritisch diskutieren.

Schlagnorte: Kompetenzmessung, Videovignetten, Expertenrating, Advokatorischer Ansatz, IRT

1. Über die professionelle Kompetenz von Lehrpersonen

1.1 Zum Kompetenzbegriff

Wer von Kompetenz spricht, meint damit das Vorhandensein einer Fähigkeit zum richtigen und situationsadäquaten Handeln. Meist wird, wenn professionelle Kompetenzen von Lehrpersonen untersucht werden, auf die Definition von Weinert zurückgegriffen, der Kompetenz versteht als „die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösung in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ (Weinert, 2001, S. 27–28). Unwidersprochen ist diese Definition zu komplex. Moser (2014) meint gar, dass diese Definition so komplex sei, dass man ihr empirisch nicht gerecht werden könne.

Es gibt ein weiteres Problem: Instrumente zur Messung professioneller Kompetenzen fokussieren meist auf das Erfassen von Wissenskomponenten. Insbesondere seit dem wegweisenden Artikel von Baumert und Kunter (2006) zu Kompetenzfacetten gilt das Top-down-Modell. Dies ist in letzter Zeit jedoch in Kritik geraten (Nickolaus, 2011; Seeber, 2014; Zlatkin-Troitschanskaia, Förster, Brückner, Hansen & Happ, 2013). Es zeichnet sich ein Bewusstsein ab, dass mit bloßem Erfassen von professionellem Wissen mit ergänzenden Skalen zu Motivation und Überzeugungen der Komplexität des

Kompetenzbegriffs nicht entsprochen werden kann (Berger et al., 2013; Oser, 2013a; Shavelson, 2013). Insbesondere rufen kompetenznahe Formulierungen nach methodischen Zugängen, welche die Komplexität und Kontextgebundenheit von Kompetenz sichtbar machen. Ein vielversprechender Zugang wird im Einsatz von Videovignetten, die professionelles Handeln abbilden, gesehen (vgl. Reusser, 2005).

1.2 *Zum Verhältnis von Kompetenz, Wissen und Performanz*

Unter Akzeptanz der Kontextgebundenheit von Kompetenz kann eine Handlung in vergleichbaren Situationen unterschiedlich gut ausgeführt werden. Obwohl stark betont wird, dass Kompetenz nicht dasselbe wie die Performanz sei, spielt das Handeln in der Situation im Kompetenzdiskurs dennoch eine zentrale Rolle (vgl. Frey, 2006; Oser, Curcio & Düggeli, 2007; Weinert, 2002). Denn die Performanz ist ein Zeichen dafür, dass Kompetenz vorhanden ist. In Bezug auf die Notwendigkeit einer Unterscheidung zwischen Wissen, Kompetenz und Handeln (Förster & Klinke, 2013) lässt sich feststellen, dass selbst die feinsten, auf Shulman (1986) zurückgehenden professionellen Wissenstypen nicht automatisch zu Kompetenzen (Neuweg, 2011) führen. Insofern stellen verschiedene Wissensformen bloß eine Bedingung der Möglichkeit für Handeln dar (Oser, 2013a). Darüber hinaus wird oft nicht gesehen, dass nicht bloß Wissen in die Kompetenz einfließt, sondern dass die Folge von Kompetenz neues Wissen (auch Handlungswissen) sein kann. Ähnlich versteht Riegel (2013) Kompetenz als Bindeglied zwischen Wissen und Können. Auch Lindmeier (2013, S. 52) meint diesbezüglich: „Da Kompetenzen stets als auf professionelle Anforderungen bezogene komplexe Fähigkeitskonstrukte gefasst sind, umfassen Kompetenzkonstrukte professionelles Wissen und es kann angenommen werden, dass eine hohe Ausprägung von Kompetenz eine profunde Wissensbasis voraussetzt. Die Gleichsetzung von Wissen und Kompetenz in einem Bereich ist damit aber aus theoretischer Sicht nicht gerechtfertigt“. Hohe Qualität hat damit zwei Seiten; einerseits setzt sie besonderes Wissen, andererseits besonders zweckmässiges Wissen voraus.

1.3 *Analytische versus holistische Sicht*

Im Forschungsprojekt Professional Minds¹ näherten wir uns dem Kompetenzbegriff durch die Definition notwendiger Kompetenzprofile für Lehrpersonen. Verstanden wird unter einem Kompetenzprofil das Zur-Verfügung-Haben „eines in komplexen Situationen des Unterrichts abgrenzbares, zieladäquates, effektives und ethisch gerechtfertigtes

1 Das Forschungsprojekt Professional Minds gehört zum Leading House „Qualität der beruflichen Bildung“, das von 2005 bis 2010 an der Universität Fribourg/Schweiz vom Staatssekretariat für Berufsbildung, Forschung und Innovation (SBFI, ehemals Bundesamt für Berufsbildung und Technologie, BBT) finanziert wurde.

Einflusshandelns, welches das Lernen von Schülern und Schülerinnen differenziell fördert“ (Oser, 2013b, S. 39). Dahinter steckt die Annahme, dass Unterrichtsqualität nur dann sichergestellt werden kann, wenn Lehrpersonen über diese Kompetenzprofile verfügen. Konkret sind 45 Kompetenzprofile für Berufsfachschullehrpersonen mithilfe des Quasi-Delphi-Verfahrens (siehe Heinzer, Oser & Salzman, 2009) formuliert worden. Diese 45 Kompetenzprofile zeigen das professionelle Handlungsfeld einer Lehrperson auf und definieren deren gefordertes Können. Weil jedes dieser Profile aus dem Feld gewonnen wurde – und nicht theoretisch abgeleitet worden war –, verstehen wir – anders als andere Kompetenzdefinitionen – unter einem Kompetenzprofil etwas Ganzheitliches, das nicht in Einzelkomponenten wie Wissen, Motivation, Überzeugungen zerlegt, wohl aber von dort gespeist werden kann. Zudem wurde, um diesem Kompetenzverständnis gerecht zu werden, ein methodischer Ansatz gesucht, der diese Kontextspezifität in seiner Komplexität möglichst ganzheitlich erfassen kann. Die Lösung wurde in der Abbildung der Kompetenzprofile mithilfe von Videovignetten gefunden. In der Handlung scheint Kompetenz performativ auf. Der Rückschluss von dem, was sichtbar ist, zur Kompetenz wird automatisch vollzogen. Das resultierende Instrument bezeichnen wir als den Advokatorischen Ansatz (folgend als AvA abgekürzt).

2. Der Advokatorische Ansatz im Kontext videobasierter Forschung

Da der AvA an anderen Stellen (etwa Heinzer & Oser, 2013) bereits ausführlich dargestellt wurde, werden im Folgenden nur dessen Kernelemente (Abschnitt 2.1) beschrieben. In Anschluss daran wird das Verfahren mit anderen videovignettenbasierten Verfahren zur Kompetenzmessung verglichen (Abschnitt 2.2). Implizite Annahmen des AvA werden in Abschnitt 2.3 erläutert.

2.1 Kernelemente des Advokatorischen Ansatzes

Der AvA zeichnet sich durch die stellvertretende Art der Kompetenzmessung aus. Konkret bedeutet dies, dass eine Lehrperson X eine andere Lehrperson Y in deren Unterricht, dargestellt durch eine Videovignette, beobachtet. Die gezeigte Videovignette bildet ein ausgewähltes Kompetenzprofil ab und bettet dieses in den Kontext ein. Ein Video dauert zwischen 4 und 8 Minuten. Wichtig ist zu beachten, dass diese Vignette authentischen Unterricht zeigt. Dadurch wird ein Kompetenzprofil sichtbar. Lehrperson X erhält zusätzlich Informationen über die Lernziele von Lehrperson Y, kann ausgeteilte Unterlagen einsehen und erhält einen Überblick über die Inhalte der vorangegangenen sowie nachfolgenden Lektion. Die Vignette ist charakterisiert durch ihre Multiperspektivität. Das Unterrichtsgeschehen wird gleichzeitig aus der Perspektive des einzelnen Schülers bzw. der einzelnen Schülerin, dem Klassenverband sowie der Lehrperson abgebildet. Im Anschluss an die Vignette wird Lehrperson X gebeten, die Qualität des beobachteten

Unterrichts bzw. die Ausprägung des Kompetenzprofils bei Lehrperson Y mittels vorgegebenen, geschlossenen Items einzuschätzen.

2.2 Videovignettenbasierte Erhebungsverfahren im Vergleich

Wie Lindmeier (2013) in einem Übersichtsartikel betont, lassen sich erhebliche Unterschiede in videovignettenbasierten Erhebungsverfahren feststellen. Die Autorin identifiziert u. a. Unterschiede im Grad der Authentizität der Videos, der Darstellungsoption, dem Antwortformat sowie der Bearbeitungsdauer. Nicht nur formal unterscheiden sich die Erhebungsverfahren, sondern auch in deren Zielsetzung. So werden Videovignetten etwa zur Erfassung von Unterrichtswahrnehmung und -analyse (Pauli & Reusser, 2006), zum Erfassen kognitiver Prozesse (Brouwer & Robijns, 2013) und – in je unterschiedlicher Weise – zur Kompetenzmessung (Blömeke, 2013; Oser, Bauder, Salzmann & Heinzer, 2013) eingesetzt. Beispiele für die Kompetenzmessung sind etwa die Erfassung aktionsbezogener Kompetenz (Lindmeier, 2013), pädagogisch-psychologischer Kompetenz (LUV: Seidel & Prenzel, 2007) oder von Classroom Management Expertise (König & Lebens, 2012). In COACTIV-R kommen Videovignetten zum Einsatz, um pädagogisches Wissen über Klassenmanagement (Voss & Kunter, 2013) oder fachdidaktisches Wissen (Bruckmaier et al., 2013) abzufragen. Im Projekt CVA (Classroom-Video-Analysis) wird mittels Videovignetten versucht, „usable knowledge“ sichtbar zu machen (Kersting, Givvin, Thompson, Santagata & Stigler, 2012). Durch den Einsatz von Videovignetten verspricht man sich eine erhöhte Validität (Lindmeier, 2013) sowie die Möglichkeit, die Komplexität und Kontextgebundenheit des Lehrerhandelns adäquater abbilden zu können als mit anderen Vorgehensweisen wie Papier-Bleistift-Tests (Kersting, 2008; Oser et al., 2013).

Trotz erwähnter Vorzüge der Videovignetten hat dieser Ansatz auch Nachteile. Einer besteht darin, dass jeweilige Instrumente in vollem Umfang nur nachvollzogen werden können, wenn der Inhalt der Vignetten sowie die Einbettung mit zuzüglichen Informationen gekannt und studiert werden kann.² Bei geschlossenen Items zur Einschätzung beobachteter Qualität liegt ein Problem in der Abhängigkeit von einer Norm, welche oft mittels Experten(ratings) eruiert wird. Hinsichtlich dem Bestreben, Kompetenz mittels Instrumenten zu erfassen, die Unterrichtsgeschehen authentisch und kontextuiert abbilden, fällt auf, dass die Mehrheit der oben erwähnten videovignettenbasierten Verfahren versucht, Wissen als Teilkomponente einer Kompetenz abzubilden (Bruckmaier et al., 2013; Kersting et al., 2012; Voss & Kunter, 2013). Der AvA nimmt diesbezüglich eine andere Position ein. Ein Kompetenzprofil wird in seiner Ganzheitlichkeit begriffen und in einem Qualitätsmaß, das eine Kompetenzsensibilität voraussetzt, ausgedrückt. Mit Kompetenzsensibilität meinen wir die Fähigkeit, die Güte eines beobachteten Kompetenzprofils adäquat oder sensibel einzuschätzen.

2 An dieser Stelle möchten wir Herrn Neuweg für seinen konstruktiven Diskussionsbeitrag an der DGfE-Tagung in Berlin im März 2014 danken.

2.3 Implizite Annahmen des Advokatorischen Ansatzes

Der Grundgedanke hinter dem AvA ist der vergleichbare „frame of reference“. Es ist davon auszugehen, dass Lehrpersonen (z. B. Lehrperson X) mit einer gewissen Sensibilität ihr eigenes Kompetenzverständnis aufrufen, wenn sie die Qualität des beobachteten durch Lehrperson Y visualisierten Kompetenzprofils einschätzen müssen. Somit wird weder die Performanz der einschätzenden Lehrperson gemessen noch ihre Kompetenz. Vielmehr interessiert, wie eine Person das Kompetenzprofil einer anderen Person z. B. aufgrund beobachteten Unterrichtshandelns hinsichtlich seiner Qualität diskutiert. In diesem Sinne sprechen wir auch von der Sensibilität, die Qualität der attribuierten Kompetenz einzuschätzen, manchmal verkürzt von Kompetenzsensibilität. Es geht also um die Frage, ob und wie die beobachtende Lehrperson die Qualität eines in der Videovignette gezeigten Kompetenzprofils nach vorgegebenen Dimensionen einschätzt. Höhere Sensibilität produziert höhere positive oder negative Differenz. Da das in der Vignette abgebildete Kompetenzprofil aus authentischem Unterricht stammt, muss für den Vergleich der Kompetenzsensibilität der einschätzenden Lehrpersonen eine Norm festgelegt werden. Dies erfolgte bisher mittels eines Expertenratings. Das bedeutet, dass drei Expertinnen bzw. Experten, die sich hinsichtlich drei erziehungswissenschaftlicher Teildisziplinen wie z. B. Wirtschaftspädagogik, Lehrerbildung/Didaktik sowie Bildungs- und Erziehungswissenschaft durch ihre Expertise auszeichnen (vgl. Bauder, 2013), zum Konsensfindungsprozess zusammenkamen und im Diskurs durch Austausch von Argumenten gemeinsam eine Norm setzten. Bislang wurde die Abweichung der einschätzenden Person von der Expertennorm mittels der euklidischen Distanz ausgedrückt. Befunde zeigten, dass Lehrpersonen gegenüber Nicht-Lehrpersonen dem Expertenurteil näher kommen und daher kompetenzsensibler sind. Zwischen Lehrpersonen unterschiedlicher Alterskategorien fanden sich jedoch keine signifikanten Unterschiede. Dies deutet darauf hin, dass Erfahrungsjahre nicht automatisch zum Aufbau von Kompetenzsensibilität beitragen (Heinzer & Oser, 2013), was auch Kersting (2008) in ihrer Untersuchung bezüglich Expertise fand.

3. Forschungsdesign

In folgendem Beitrag steht das Kompetenzprofil Frage-Antwort-Unterricht (FAU) im Zentrum der Untersuchung. Dieses lautet wie folgt: „Die Lehrperson kann im erarbeitenden Unterricht Fragen stellen und die Antworten der Lernenden aufgreifen. Sie kann mögliche negative Folgen des fragenden Unterrichts abschätzen und verhindern“ (Heinzer et al., 2009, S. 42). Ziel der Studie war, bisher unbeantwortete Fragen wie beispielsweise die Expertenübereinstimmung, die psychometrische Güte der Items (IRT-skaliert), den Reihenfolge-Effekt sowie Stabilität der Kompetenzsensibilität genauer zu untersuchen.

3.1 Fragestellung

Das FAU-Instrument ist durch folgende Neuerungen gekennzeichnet: (1) Anstelle eines Expertenratings wurden zwei Expertenratings durchgeführt und (2) anstelle einer wurden zwei Videovignetten zum Kompetenzprofil FAU eingesetzt. Zudem wurde die Güte der Items auf dem Hintergrund der Item-Response-Theorie geprüft. Folgende Fragen standen im Zentrum:

- A) Stimmen verschiedene Expertenratinggruppen in ihrer Normsetzung überein?
- B) Weist die IRT-Skalierung eine angemessene Güte aus?
- C) Beeinflusst die Reihenfolge der Videovignetten die Kompetenzsensibilität?
- D) Ist es gelungen, mit dem AvA eine gewisse Stabilität der latenten Disposition (Kompetenzsensibilität) abzubilden?

3.2 Das Messinstrument

Das Instrument zur Messung des Kompetenzprofils FAU bestand aus zwei Vignetten. In Tabelle 1 ist eine Übersicht der beiden Versionen A und B gegeben.

FAU wurde demnach im Unterricht von angehenden Polymechanikern (Vignette A) und im Unterricht von angehenden Elektrotechnikern (Vignette B) von zwei unterschiedlichen Fachlehrpersonen gezeigt. Zu Beginn der Befragung wurden einige sozio-demografische Variablen erfasst. Danach wurde die erste Vignette in den Kontext der vor- und nachfolgenden Unterrichtsplanung eingebettet und es folgten 27 Items zur Einschätzung der Qualität des beobachteten Kompetenzprofils FAU der Lehrperson in der ersten Vignette. Als Nächstes wurden Informationen zur Einbettung der zweiten Vignette in das Unterrichtsgeschehen gegeben, und es folgten dieselben Items zur Einschätzung der Qualität bezogen auf die zweite Vignette. Beispielsweise lautete ein Item: „Die Lehrperson stellt in diesem Filmausschnitt zu viele Fragen“. Das Instrument

	Vignette A	Vignette B
Lehrperson	Männlich	Männlich
Alter	> 50 Jahre	30–40 Jahre
Lernende	Angehende Polymechaniker/innen	Angehende Elektrotechniker/innen
Dauer Vignette	5:14 Minuten	5:23 Minuten
Unterrichtsform	Frage-Antwort-Unterricht	Frage-Antwort-Unterricht
# Items zur Einschätzung	27	27
# verbleibende Items nach Expertenrating	18	14

Tab. 1: Übersicht über die eingesetzten Videovignetten

wurde so programmiert, dass es online ausgefüllt werden kann und die Reihenfolge der Vignetten (Version A/B oder B/A) automatisch variiert wird. Zum Bearbeiten des Instruments wurden durchschnittlich 43.5 Minuten benötigt, bei einer Standardabweichung von 16.5 Minuten und einem Range von 12 bis 99 Minuten (zeitunlimitiert). Es bestand die Möglichkeit, mehrmals auf die jeweiligen Vignetten zurückzugreifen. Das Antwortformat war eine 6-Punkt-Likert-Skala von „trifft überhaupt nicht zu“ bis „trifft vollkommen zu“. Für statistische Analysen wurde die Skala jedoch auf eine 4-Punkt-Likert-Skala verkürzt. Grund dafür war, dass in den Expertenratings anhand der Argumentation festgestellt werden konnte, dass die beiden äußersten Antwortkategorien nicht trennscharf genug waren. In einem weiteren Schritt wurde die Einschätzung der Lehrpersonen mit der Expertennorm verglichen. Pro Übereinstimmung wurde der Wert eines Punktes gesetzt.

3.3 Stichprobe

151 Berufsfachschullehrpersonen der Deutschschweiz füllten das Instrument aus. Fünf Personen wurden im Nachhinein ausgeschlossen, da sie über kein Lehrdiplom verfügten und gleichzeitig nie oder gegenwärtig nicht an einer Berufsfachschule arbeiteten. Von den verbleibenden 146 Personen waren 19 weiblich und 127 männlich. Dies entspricht in etwa dem Geschlechterverhältnis an gewerblichen Berufsfachschulen. Im Durchschnitt waren die Lehrpersonen 49 Jahre alt (SD = 8.9 Jahre), bei einem Range von 28 bis 66 Jahren. Die Lehrpersonen hatten durchschnittlich 16.6 Jahre (SD = 9.3 Jahre) Unterrichtserfahrung, wovon sie im Durchschnitt 14.3 Jahre (SD = 8.7 Jahre) an einer Berufsfachschule unterrichteten. Insgesamt haben 30 Lehrpersonen (20.5%) kein Lehrdiplom erworben. In Tabelle 2 wird deutlich, dass die Frauen weniger lang an einer Berufsfachschule unterrichteten (M = 9.3, SD = 9.1 Jahre) als die männlichen Teilnehmenden (M = 15, SD = 8.4 Jahre). Die Mehrheit der befragten Personen unterrichtete berufsspezifische Fächer (71%), gefolgt von Allgemeinbildung (21%) und Sport (8%).

	N	Alter	Unterrichtserfahrung allgemein	Unterrichtserfahrung an Berufsfachschulen
Weiblich	19	M = 46.9 Jahre SD = 11.2 Jahre Min: 28 – Max: 66 Jahre	M = 13.7 Jahre SD = 11.4 Jahre Min: 1 – Max: 36 Jahre	M = 9.3 Jahre SD = 9.1 Jahre Min: 1 – Max: 32 Jahre
Männlich	127	M = 49.3 Jahre SD = 8.5 Jahre Min: 28 – Max: 64 Jahre	M = 17 Jahre SD = 8.9 Jahre Min: 2 – Max: 41 Jahre	M = 15 Jahre SD = 8.4 Jahre Min: 1 – Max: 37 Jahre
Total	146	M = 49 Jahre SD = 8.9 Jahre Min: 28 – Max: 66 Jahre	M = 16.6 Jahre SD = 9.3 Jahre Min: 1 – Max: 41 Jahre	M = 14.3 Jahre SD = 8.7 Jahre Min: 1 – Max: 37 Jahre

Tab. 2: Stichprobe der befragten Berufsfachschullehrpersonen

Für die Datenerhebung wurden vorgängig die Rektoren bzw. Schulleiter der jeweiligen Berufsfachschule kontaktiert und der Link für die Befragung zur internen Streuung zugesandt. Version B/A (Lehrperson in Elektrotechnik/Lehrperson in Polymechanik) haben 115 (79%) Lehrpersonen ausgefüllt, Version A/B nur 31 Personen (21%). Leider variieren die beiden Gruppen in ihrer Größe.

4. Ergebnis

Im Folgenden werden die oben aufgelisteten Fragestellungen beantwortet.

4.1 Die Übereinstimmung der Experten und Expertinnen (Frage A)

Insgesamt sechs Expertinnen bzw. Experten sind in zwei voneinander unabhängigen Gruppen zum Konsensfindungsprozess zusammengekommen und haben im Diskurs durch Austausch von Argumenten je eine Norm gesetzt. Im Anschluss daran wurde die gefundene Norm der beiden Gruppen verglichen. In Vignette A wurde bei 18 der 27 Items eine Übereinstimmung erzielt (66.6%), in Vignette B bloss bei 14 der 27 Items (51.8%). Ursprünglich war die Idee, nur mit jenen Items weiterzurechnen, die sich für beide Vignetten als valide erwiesen. Daher wurde auch derselbe Fragenkatalog eingesetzt. Diese Idee erwies sich allerdings als nicht haltbar und auch nicht zwingend sinnvoll. Es zeigte sich, dass, obwohl in beiden Vignetten FAU durchgeführt wird, dieser dennoch zu unterschiedlich gestaltet ist und nicht alle Items für beide Vignetten gleich gut zu beantworten waren (z. B. durch verminderte Sichtbarkeit). Daher wurden nur jene Items mit Doppelkonsens (Vignette A = 18 Items, Vignette B = 14 Items, siehe Tabelle 1) zur Messung der Kompetenzsensibilität berücksichtigt.

4.2 IRT-Skalierung (Frage B)

Mit 146 Personen und verbleibenden 32 Items ist die Datengrundlage eher dünn. Da keine Zufallsstichprobe vorliegt, verstehen sich folgende Ausführungen als explorativ, und von Verallgemeinerungen ist abzusehen. Die Itemanalyse und die Schätzung der Personenfähigkeit wurden mithilfe des R Packages TAM (Kiefer, Robitzsch & Wu, 2014) durchgeführt.

Da die Beantwortung der Items videovignettenbezogen erfolgte, wurde in einem ersten Schritt eine zwei-dimensionale 1-PL-Analyse durchgeführt. Die 18 binär kodierten Items der Vignette A wurden auf die Dimension 1 geladen, die 14 binär kodierten Items der Vignette B auf Dimension 2. Es zeigte sich, dass ein Problem mit der zweiten Dimension besteht, da die EAP-Reliabilität sowie die Varianz nicht geschätzt werden konnten. Ein Blick auf die Item-Fits (INFIT) zeigte, dass alle INFITS der Items der zweiten Dimension auf 1 liegen. Um mehr Informationen bezüglich des Missfits zu er-

halten, wurde die Restriktion gleicher Trennschärfen in einem zweiten Schritt aufgegeben und es wurde ein 2-PL-Modell berechnet. Auf beiden Dimensionen fanden sich negative Item-Trennschärfen, was auf weitere mögliche Dimensionen hinweist (Hartig, 2012). Die betroffenen Items wurden nun in einem nächsten Schritt theoretisch analysiert. Für Vignette A konnte eine weitere Dimension inhaltlich nicht begründet werden. Daher wurden die Items mit negativer Trennschärfe und schlechtem FIT ($\text{INFIT} > 1.3$ sowie $\text{INFIT}_T > 2.00$; Hartig, 2012; Wright, Linacre, Gustafsson & Martin-Löf, 1994) ausgeschlossen. Es verblieben 11 Items auf der ersten Dimension. Für die Items der Vignette B wurde eine Unterteilung in zwei Dimensionen als legitim erachtet. Inhaltlich gesehen erfasst die eine Dimension eine allgemeine Einschätzung der Qualität von FAU, während die andere Dimension stärker auf Regeln wie „Wartezeit nach einer Frage“ und „Grad der Steuerung des Unterrichts“ abzielt (diese Items fielen bei Vignette A größtenteils bereits nach dem Expertenrating weg). Die 11 Items der ersten Dimension gehen inhaltlich ebenfalls auf eine allgemeine Einschätzung der Qualität von FAU (bezogen auf Vignette A) ein. In einem weiteren Schritt wurde daher ein drei-dimensionales 2PL-Modell gerechnet. Es verblieben folgende Anzahl Items in der Berechnung:

- 1. Dimension: 11 Items zur Vignette A gehörend (allgemeine Einschätzung der Qualität des beobachteten FAU)
- 2. Dimension: 6 Items zur Vignette B gehörend (allgemeine Einschätzung der Qualität des beobachteten FAU)
- 3. Dimension: 4 Items zur Vignette B gehörend (Einschätzung bezüglich Regeln von FAU)

Wie in Tabelle 3 dargestellt (Modell 2), erreichte die EAP-Reliabilität zufriedenstellende Werte (Hartig, 2012; Wright et al., 1994). Die WLE-Reliabilität fiel allerdings für die Dimensionen zwei und drei (Items bezogen auf Vignette B) sehr niedrig aus. Da mittels des 2-PL-Modells die Items mit schlechtem Fit ausgeschlossen werden konnten, wurde in einem nächsten Schritt wiederum geprüft, ob nun die Bedingung gleicher Trennschärfen aufrechterhalten werden konnte. Folglich wurde ein drei-dimensionales Raschmodell berechnet. Die Werte, abgetragen in Tabelle 3, zeigen, dass das Modell 3 höhere Varianzen und vergleichbare EAP-Reliabilität aufweist. Hinsichtlich der WLE-Schätzungen sind diese für die Dimensionen 2 und 3 allerdings noch etwas niedriger. Der BIC hingegen ist deutlich niedriger beim Modell 3 als beim Modell 2. Ein Modellvergleich bestätigt eine signifikante Abweichung ($\chi^2(18) = 37.82$; $p < .001$), weshalb dem restriktiveren Modell 3 der Vorzug gegeben wurde.

Eine deskriptive Analyse der geschätzten Itemschwierigkeiten ergab einen Mittelwert von 0.54 bei einer Standardabweichung von 1.08. Das schwierigste Item wurde von 15% der Befragten in Übereinstimmung mit der Expertennorm beantwortet, das einfachste von 70%. Die Verteilung der Itemschwierigkeiten zeigte, dass der mittlere Schwierigkeitsbereich eher untervertreten ist. Bei zusätzlicher Itementwicklung müsste dieser Bereich besser abgedeckt werden.

	Modell 2	Modell 3	Modell 4
Schätzung	3-dimensionale 2-PL-Lösung	3-dimensionale 1-PL-Lösung	1-dimensionale 1-PL-Lösung
# Items	21	21	9
Varianz			2.4
Dimension 1	1	2.44	
Dimension 2	1	1.27	
Dimension 3	1	2.83	
EAP			.78
Dimension 1	.82	.80	
Dimension 2	.65	.60	
Dimension 3	.71	.69	
WLE			.69
Dimension 1	.74	.72	
Dimension 2	.34	.23	
Dimension 3	.33	.15	
BIC	3224.93	3174.72	1502.39

Tab. 3: Kennwerte der geschätzten Modelle

4.3 Über den Effekt der Reihenfolge (Frage C)

Wie anfänglich erwähnt, wurde das Befragungsinstrument in zwei Versionen mit variierender Reihenfolge der Vignetten ausgefüllt. Es konnten keine signifikanten Unterschiede der beiden Gruppen (A/B bzw. B/A) bezüglich Alter, Geschlecht, Erfahrung und Unterrichtsfach ermittelt werden. Insofern gab es keinen Grund anzunehmen, dass die eine Gruppe stärker mit der Expertennorm übereinstimmt als die andere. Um sicherzustellen, dass sich die Aufgaben für beide Gruppen gleich eignen, wurde zunächst eine grafische Differential-Item-Functioning-Analyse (DIF; Strobl, 2010) mittels des R-Packages eRm durchgeführt.

Abbildung 1 (oben) verdeutlicht, dass neun Aufgaben DIF zeigen (Strobl, 2010). Eine dieser Aufgaben gehört der ersten Dimension an. Fünf Items der zweiten Dimension und drei Items der dritten Dimension (Modell 3). Nach Eliminierung der problematischen Items besteht die erste Dimension aus zehn Items. Für die Dimensionen 2 und 3 hingegen verbleiben nur noch je ein Item. Die oben bereits festgestellte Problematik der Vignette B (niedrige WLE-Schätzungen) hat sich bestätigt. Daher wurde in einem nächsten Schritt zur Schätzung der Personenfähigkeit ein revidiertes Modell berechnet, das nur noch die zehn Items der Vignette A beinhaltet (1-dimensionales 1-PL-Modell). Die Kennwerte sind in Tabelle 3 (Modell 4) dargestellt. Abbildung 1 (unten) zeigt zudem, dass sich die verbleibenden Aufgaben nicht auffällig hinsichtlich DIF verhalten. Zur Beantwortung der Fragestellung C wurden die Mittelwerte der geschätzten Personenfähigkeit (Modell 4) der beiden Gruppen mittels eines T-Tests für unabhängige Stichproben getestet. Entgegen der Erwartung fand sich jedoch ein signifikanter Unterschied in der mittleren Personenfähigkeit (siehe Tabelle 4). Die Effektstärke wurde ge-

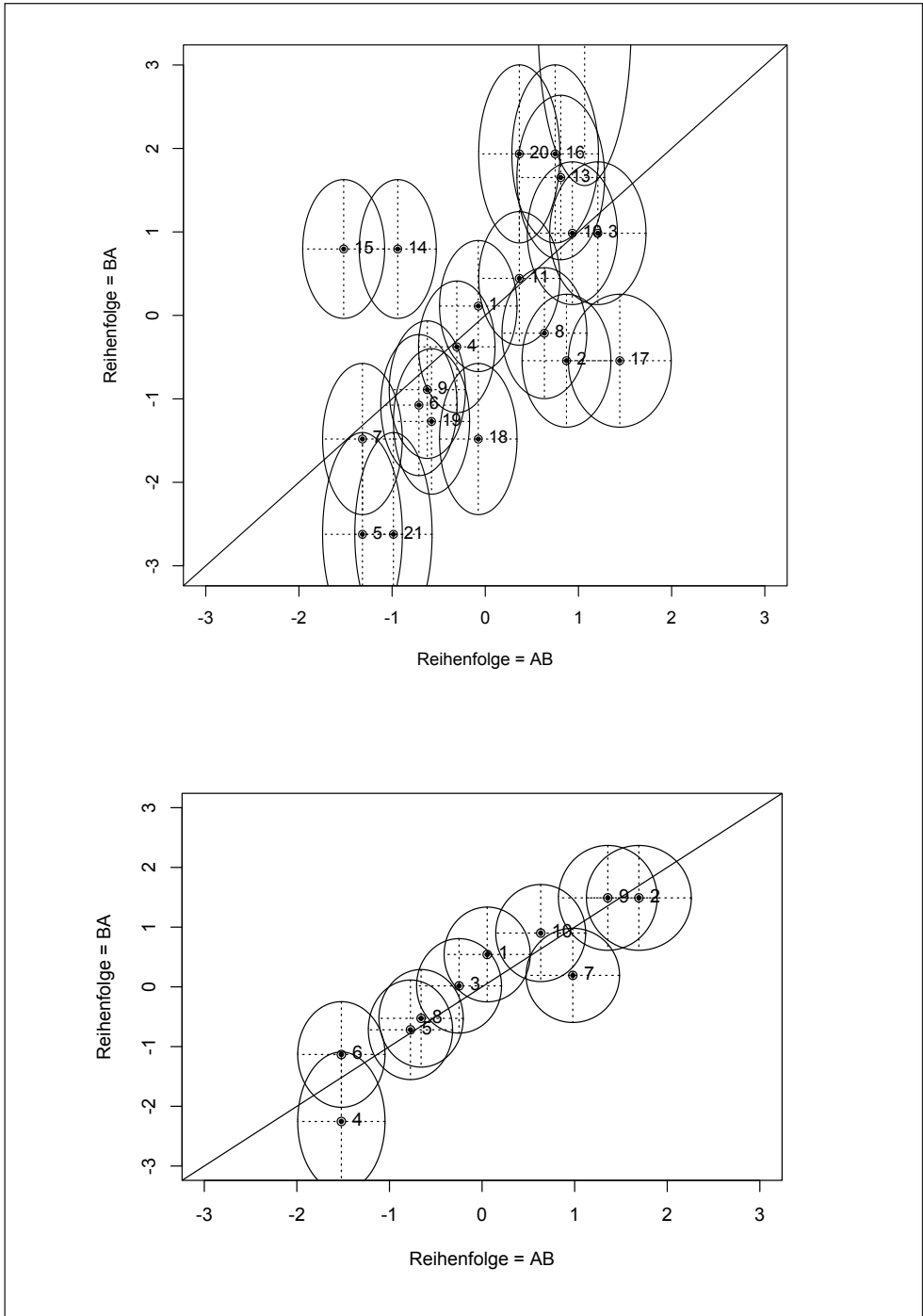


Abb. 1: Grafischer DIF-Test nach Reihenfolge für Modell 3 (oben), für Modell 4 (unten) mit 95%-Konfidenz-Regionen für die geschätzten Aufgaben-Parameter

Personenfähigkeit	Gruppe	N	Mittelwert	SD	t-Wert	Cohen's d
Theta	A/B	29	.44	.84	2.528 (131); $p < .05$.53
	B/A	108	-.09	1.04		

Tab. 4: Mittelwertunterschied der Personenfähigkeit

mäß der Formel von Cohen's d (bzw. Hedges' g; Lenhard & Lenhard, 2014) unter Berücksichtigung ungleicher Gruppengrößen berechnet. Es zeigte sich ein mittlerer Effekt.

Lehrpersonen, die zuerst Vignette A, dann B gesehen haben, erzielten im Durchschnitt einen höheren Fähigkeitsparameter als Lehrpersonen, die zuerst Vignette B sahen. Ungeklärt bleibt, ob die Tatsache, dass Gruppe B/A zuerst Vignette B sah, die Qualitätseinschätzung beeinflusst hat, die zweite Vignette weniger seriös bearbeitet wurde oder ob sich die beiden Gruppen durch weitere, nicht erhobene Charakteristiken signifikant unterscheiden.

4.4 Über die Stabilität der Personenfähigkeit (Frage D)

Fragestellung D zur Stabilität der Messung zielte auf eine Analyse der Paralleltest-Reliabilität ab. Wenn mit dem AvA Kompetenzsensibilität abgebildet wird, dann ist anzunehmen, dass die geschätzte Fähigkeit über die beiden Vignetten (A und B) hinweg stabil bleibt. Leider ist diese Fragestellung aufgrund der schlechten Itemfits der Vignette B nicht abschließend zu beantworten.

5. Diskussion

Ein Spezifikum des AvA ist, dass eine Norm zur Beurteilung der Qualitätseinschätzungen gesetzt werden muss. Dies geschieht unter Rückgriff auf Experten. Dieser Rückgriff ist jedoch nicht unproblematisch. Durch das Fehlen eines klaren Richtig oder Falsch wird die Leistung (hier Kompetenzsensibilität) der einschätzenden Person von einer Expertennorm abhängig gemacht. Im Falle des AvA wurde dies zu kontrollieren versucht, indem die Norm über einen Austausch- und Konsensfindungsprozess intersubjektiv gefunden wurde. Dennoch zeigte ein Vergleich der beiden Expertenratings, dass sich zwei unabhängige Expertengruppen nur bezüglich 50–60% der Items einig waren. Im weiteren Verlauf wurden nur jene Items weiterverwendet, die einen Zweifach-Konsens erzielten. Die Frage bleibt offen, ob dies genügt oder ob noch weitere unabhängige Gruppen einbezogen werden müssten. Zudem stellt sich die Frage nach der Auswahl der Experten. Ein möglicher Lösungsansatz wäre, von den einschätzenden Lehrpersonen Begründungen einzufordern und diese dann mit denen der Experten zu vergleichen. Weiter zeigte sich, dass die Items der Vignette B problematisch waren und

DIF bezüglich der Reihenfolge aufwiesen. Gründe für die Nicht-Eignung der Vignette B können bei der Videovignette selbst, der Passung der Items oder in deren Kombination liegen. Es müssen auf jeden Fall genauere Analysen der beiden Videos sowie der Itembasis stattfinden, um mögliche Erklärungen dafür zu finden. Des Weiteren muss das überarbeitete Instrument an einer größeren Stichprobe mit gleicher Gruppengröße getestet werden. Diesbezüglich ist auch das Resultat hinsichtlich des Reihenfolge-Effekts zentral. Es ist anzunehmen, dass nach Beobachtung der zweiten Vignette automatisch ein Vergleich der beiden beobachteten Lehrpersonen stattfindet und das neue Urteil am vorgängigen Urteil angepasst wird. Dies ist allerdings nur eine Vermutung, die geprüft werden müsste. Denkbar wäre auch, dass sich die einschätzenden Lehrpersonen an die Kriterien, Vorgaben und Aufgaben gewöhnen und die zweite Vignette schneller, spontaner und mit weniger Bedenken bearbeiten. Zusammenfassend hat sich gezeigt, dass der AvA optimiert werden kann und zentrale Fragen noch unbeantwortet bleiben. Es fällt auf, dass es kaum Untersuchungen dazu gibt, was während des Betrachtens von Videovignetten passiert und welche Faktoren das Urteil der Einschätzenden beeinflussen. Dieses Wissen wäre nicht nur für den AvA, sondern für alle Forschenden, die mit Videovignetten arbeiten, relevant.

Literatur

- Bauder, T. (2013). Mit Expertenrating zur Einschätzungsnorm: Unterrichtshandeln beobachten und diskutieren. Qualität feststellen und begründen. In F. Oser, T. Bauder, P. Salzmann & S. Heinzer (Hrsg.), *Ohne Kompetenz keine Qualität. Entwickeln und Einschätzen von Kompetenzen bei Lehrpersonen und Berufsbildungsverantwortlichen* (S. 191–213). Bad Heilbrunn: Klinkhardt.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 143–157.
- Berger, S., Fritsch, S., Seifried, J., Bouley, F., Mindnich, A., Wuttke, E., Schnick-Vollmer, K., & Schmitz, B. (2013). Entwicklung eines Testinstruments zur Erfassung des fachlichen und fachdidaktischen Wissens von Studierenden der Wirtschaftspädagogik – Erste Erfahrungen und Befunde. In O. Zlatkin-Troitschanskaia, R. Nickolaus & K. Beck (Hrsg.), *Kompetenzmodellierung und Kompetenzmessung bei Studierenden der Wirtschaftswissenschaften und der Ingenieurwissenschaften* (Lehrerbildung auf dem Prüfstand, Sonderheft, S. 93–107). Landau: Verlag Empirische Pädagogik.
- Blömeke, S. (2013). Moving to a higher state of confusion. Der Beitrag der Videoforschung zur Kompetenzforschung. In U. Riegel & K. Macha (Hrsg.), *Videobasierte Kompetenzforschung in den Fachdidaktiken* (S. 25–43). Münster: Waxmann.
- Brouwer, N., & Robijns, F. (2013). Fokussierte Auswertung von Videoaufzeichnungen als Methode in der Lehrerausbildung. In U. Riegel & K. Macha (Hrsg.), *Videobasierte Kompetenzforschung in den Fachdidaktiken* (S. 303–317). Münster: Waxmann.
- Bruckmaier, G., Krauss, S., Leiss, D., Blum, W., Neubrand, M., & Brunner, M. (2013). COACTIV-Video: Eine unterrichtsnahe Erfassung fachdidaktischen Wissens mittels Videovignetten. In *Beiträge zum Mathematikunterricht 2013* (S. 212–215). Münster: WTM.
- Förster, M., & Klinke, S. (2013). Prüfung auf nicht-lineare Zusammenhänge und deren Modellierung in der Kompetenzforschung. In O. Zlatkin-Troitschanskaia, R. Nickolaus & K. Beck (Hrsg.), *Kompetenzmodellierung und Kompetenzmessung bei Studierenden der Wirtschafts-*

- wissenschaften und der Ingenieurwissenschaften (Lehrerbildung auf dem Prüfstand, Sonderheft, S. 49–68). Landau: Verlag Empirische Pädagogik.
- Frey, A. (2006). Methoden und Instrumente zur Diagnose beruflicher Kompetenzen von Lehrkräften – eine erste Standortbestimmung zu bereits publizierten Instrumenten. *Zeitschrift für Pädagogik*, 51. Beiheft, 30–46.
- Hartig, J. (2012). *IRT Skalierung mit ConQuest*. Workshop an der Universität Zürich (5. & 6. September 2012).
- Heinzer, S., & Oser, F. (2013). Das Advokatorische Messverfahren: Die stellvertretende Art Kompetenzen zu messen. In F. Oser, T. Bauder, P. Salzmann & S. Heinzer (Hrsg.), *Ohne Kompetenz keine Qualität. Entwickeln und Einschätzen von Kompetenzen bei Lehrpersonen und Berufsbildungsverantwortlichen* (S. 139–168). Bad Heilbrunn: Klinkhardt.
- Heinzer, S., Oser, F., & Salzmann, P. (2009). Zur Genese von Kompetenzprofilen. *Lehrerbildung auf dem Prüfstand*, 2(1), 28–55.
- Kersting, N. B. (2008). Using Video Clips of Mathematics Classroom Instruction as Item Prompts to Measure Teachers' Knowledge of Teaching Mathematics. *Educational and Psychological Measurement*, 68(5), 845–861.
- Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring Usable Knowledge: Teachers' Analyses of Mathematics Classroom Video Predict Teaching Quality and Student Learning. *American Educational Research Journal*, 49(3), 568–589.
- Kiefer, T., Robitzsch, A., & Wu, M. (2014). *Package 'tam'*. <http://cran.r-project.org/web/packages/TAM/TAM.pdf> [03.09.2014].
- König, J., & Lebens, M. (2012). Classroom Management Expertise (CME) von Lehrkräften messen: Überlegungen zur Testung mithilfe von Videovignetten und erste empirische Befunde. *Lehrerbildung auf dem Prüfstand*, 5(1), 3–28.
- Lenhard, W., & Lenhard, A. (2014). *Berechnung von Effektstärken*. <http://www.psychometrica.de/effektstaerke.html> [05.12.2014].
- Lindmeier, A. (2013). Video-vignettenbasierte standardisierte Erhebung von Lehrerkognitionen. In U. Riegel & K. Macha (Hrsg.), *Videobasierte Kompetenzforschung in den Fachdidaktiken* (S. 45–62). Münster: Waxmann.
- Moser, U. (2014). *Kompetenzorientierung im Spannungsfeld von Fachdidaktik, Schulpraxis und Testtheorie*. Vortrag am Kongress Schweizer Gesellschaft für Bildungsforschung in Luzern (25. September).
- Neuweg, G. H. (2011). Das Wissen der Wissensvermittler. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrberuf* (S. 451–477). Münster: Waxmann.
- Nickolaus, R. (2011). Die Erfassung fachlicher Kompetenz und ihre Entwicklung in der beruflichen Bildung – Forschungsstand und Perspektiven. In O. Zlatkin-Troitschanskaia (Hrsg.), *Stationen empirischer Bildungsforschung* (S. 331–351). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Oser, F. (2013a). I know how to do it, but I can't do it – modeling competence profiles for future teachers and trainers. In S. Blömeke, O. Zlatkin-Troitschanskaia, C. Kuhn & J. Fege (Hrsg.), *Modeling and measuring competencies in higher education* (S. 45–60). Rotterdam: Sense Publishers.
- Oser, F. (2013b). Kompetenzen der Lehrenden und Auszubildenden: Ein Ressourcenmodell. In F. Oser, T. Bauder, P. Salzmann & S. Heinzer (Hrsg.), *Ohne Kompetenz keine Qualität. Entwickeln und Einschätzen von Kompetenzen bei Lehrpersonen und Berufsbildungsverantwortlichen* (S. 29–65). Bad Heilbrunn: Klinkhardt.
- Oser, F., Bauder, T., Salzmann, P., & Heinzer, S. (Hrsg.) (2013). *Ohne Kompetenz keine Qualität. Entwickeln und Einschätzen von Kompetenzen bei Lehrpersonen und Berufsbildungsverantwortlichen*. Bad Heilbrunn: Klinkhardt.

- Oser, F., Curcio, G.-P., & Düggele, A. (2007). Kompetenzmessung in der Lehrerbildung als Notwendigkeit – Fragen und Zugänge. *Beiträge zur Lehrerbildung. Zeitschrift zu Theorie und Praxis der Aus- und Weiterbildung von Lehrerinnen und Lehrern*, 25(1), 14–26.
- Pauli, C., & Reusser, K. (2006). Von international vergleichenden Video Surveys zur videobasierten Unterrichtsforschung und -entwicklung. *Zeitschrift für Pädagogik*, 52, 744–798.
- Reusser, K. (2005). Situiertes Lernen mit Unterrichtsvideos. *Journal für Lehrerinnen- und Lehrerbildung*, 2, 8–18.
- Riegel, U. (2013). Videobasierte Kompetenzforschung in den Fachdidaktiken. Einleitung. In U. Riegel & K. Macha (Hrsg.), *Videobasierte Kompetenzforschung in den Fachdidaktiken* (S. 9–24). Münster: Waxmann.
- Seeber, S. (2014). Struktur und kognitive Voraussetzungen beruflicher Fachkompetenz: Am Beispiel Medizinischer und Zahnmedizinischer Fachangestellter. *Zeitschrift für Erziehungswissenschaft*, 17(Sonderheft 22), 59–80.
- Seidel, T., & Prenzel, M. (2007). Wie Lehrpersonen Unterricht wahrnehmen und einschätzen – Erfassung pädagogisch-psychologischer Kompetenzen mit Videosequenzen. *Zeitschrift für Erziehungswissenschaft*, 10(Sonderheft 8), 201–216.
- Shavelson, R. J. (2013). On an approach to testing and modelling competence. *Educational Psychology*, 48(2), 73–86.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Leadership*, 15, 4–14.
- Strobl, C. (2010). *Das Rasch-Modell. Eine verständliche Einführung für Studium und Praxis*. München: Rainer Hampp.
- Voss, T., & Kunter, M. (2013). Teachers' General Pedagogical/Psychological Knowledge. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers. Results from the COACTIV Project* (Vol. 8, S. 207–227). New York: Springer.
- Weinert, F. E. (2001). *Leistungsmessung in Schulen*. Weinheim/Basel: Beltz.
- Weinert, F. E. (2002). Concept of Competence: A Conceptual Clarification. In D. S. Rychen & L. H. Salganik (Hrsg.), *Defining and selecting key competencies* (S. 46–65). Bern/Göttingen: Hogrefe & Huber.
- Wright, B. D., Linacre, M., Gustafsson, J.-E., & Martin-Löf, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370. <http://www.rasch.org/rmt/rmt83b.htm> [08.09.2014].
- Zlatkin-Troitschanskaia, O., Förster, M., Brückner, S., Hansen, S., & Happ, R. (2013). Modellierung und Erfassung der wirtschaftswissenschaftlichen Fachkompetenz bei Studierenden im deutschen Hochschulbereich. In O. Zlatkin-Troitschanskaia, R. Nickolaus & K. Beck (Hrsg.), *Kompetenzmodellierung und Kompetenzmessung bei Studierenden der Wirtschaftswissenschaften und der Ingenieurwissenschaften* (Lehrerbildung auf dem Prüfstand, Sonderheft, S. 108–133). Landau: Verlag Empirische Pädagogik.

Abstract: In both research and practice, the construct of professional competence is omnipresent. The components of this construct, however, mostly remain diffuse. Due to a lopsided focus on knowledge aspects its representation often remains inadequate. Therefore, the first part of this contribution is devoted to a critical discussion of the concept of competence. The authors argue in favor of a holistic and context-specific approach to professional competences. In the second part, the advocacy approach is presented. This approach uses video-vignettes to investigate in how far teachers are able to assess in a sensitive way the quality of a performed professional action. A concrete example is given to illustrate how competence sensitivity is measured. IRT-analyses carried out to examine the psychometrical quality reveal a few difficulties that are then critically discussed.

Keywords: Competence Measurement, Video-Vignettes, Expert Rating, Advocatory Approach, IRT

Anschrift der Autorin/des Autors

Dr. Sarah Forster-Heinzer, Universität Zürich, Institut für Erziehungswissenschaft,
Lehrstuhl Allgemeine Erziehungswissenschaft, Freiestrasse 36, 8032 Zürich, Schweiz
E-Mail: sarah.forster-heinzer@ife.uzh.ch

Prof. Dr. Dr. h. c. mult. Fritz Oser, Emeritus, Universität Freiburg, Departement
Erziehungswissenschaften, Rue P.-A. de Faucigny 2, 1700 Freiburg, Schweiz
E-Mail: fritz.oser@unifr.ch