

Matthias Krepf

Wie analysieren ExpertInnen und NovizInnen Unterricht?

Validierung des Konstrukts Analysekompetenz
mittels Qualitativer Inhaltsanalyse

Matthias Krepf

Wie analysieren ExpertInnen und NovizInnen Unterricht?

Validierung des Konstrukts Analysekompetenz
mittels Qualitativer Inhaltsanalyse

Verlag Julius Klinkhardt
Bad Heilbrunn • 2019

k

Meinen Eltern, Barbara und Ulrich, gewidmet.

Die vorliegende Arbeit wurde im Januar 2018 von der Humanwissenschaftlichen Fakultät der Universität zu Köln als Dissertation angenommen.

Gutachter: Prof. Dr. Wilfried Plöger, Prof. Dr. Petra Herzmann.

Tag der Disputation: 24.01.2018

Dieser Titel wurde in das Programm des Verlages mittels eines Peer-Review-Verfahrens aufgenommen.
Für weitere Informationen siehe www.klinkhardt.de.

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation
in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten
sind im Internet abrufbar über <http://dnb.d-nb.de>.

2019.kg. © by Julius Klinkhardt.

Das Werk ist einschließlich aller seiner Teile urheberrechtlich geschützt.

Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Druck und Bindung: AZ Druck und Datentechnik, Kempten.

Printed in Germany 2019.

Gedruckt auf chlorfrei gebleichtem alterungsbeständigem Papier.

ISBN 978-3-7815-2312-8

Danksagung

Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät der Universität zu Köln im Januar 2018 angenommen.

Zum Abschluss des Promotionsverfahrens wird meist ein letzter, abschließender Textbaustein der Arbeit hinzugefügt: die Danksagung, denn am Ende einer langen, herausfordernden und intensiven Reise lohnt sich ein Blick zurück, um den Personen zu danken, die maßgeblich an der Entstehung der Arbeit beteiligt waren.

Zuallererst gilt mein Dank meinem Doktorvater Herrn Prof. Dr. Wilfried Plöger für die intensive, motivierende Unterstützung in allen Phasen des Promotionsprozesses. Danken möchte ich ihm für seinen engagierten Einsatz, seine Geduld, die stets offene Tür für Fragen, seine konstruktiven Rückmeldungen und die gemeinsamen Gespräche, die ich als Ermutigung und Motivation verstanden habe. Ich danke ihm von Herzen für die hervorragende Betreuung und die lehrreiche Zeit an seinem Lehrstuhl!

Des Weiteren möchte ich meinem Kollegen Herrn Prof. Dr. Daniel Scholl für die erfolgreiche Zusammenarbeit danken und dass er mir mit Rat und Tat zur Seite stand.

Ferner möchte ich Alina Diederichs, Eric Berndt und Max Peltzer herzlich danken. Sie haben mir, im Rahmen ihrer Tätigkeit als studentische Hilfskräfte, tatkräftig bei der Transkription und Kodierung der Interviews geholfen.

Einen besonderen Dank möchte ich Herrn PD Dr. Rainer Wisbert für unsere zahlreichen kontroversen, anregungsreichen Debatten sowie motivierenden Gespräche aussprechen.

Darüber hinaus möchte ich mich bei meiner Familie und meinen Freunden bedanken. Ein aufrichtiges Dankeschön dafür, dass sie mich stets unterstützt haben sowie mit Geduld und Verständnis meinen Weg verfolgten.

Mein abschließender Dank gilt meinen Eltern, Barbara und Ulrich. Ich danke ihnen für ihre Liebe und ihre fortwährende, bedingungslose Unterstützung, die mir die Kraft gab, dieses Projekt in Angriff zu nehmen und erfolgreich abzuschließen. Aus diesem Grund möchte ich ihnen dieses Buch widmen!

Kurzzusammenfassung

Die vorliegende Arbeit fokussiert auf die Analysekompetenz von Lehrpersonen, definiert als jene Fähigkeit, „die Qualität ... von beobachtetem Unterricht im Hinblick auf dessen Lernwirksamkeit angemessen erfassen und bewerten zu können“ (Plöger & Scholl, 2014a, S. 88). Im Sinne dieser Definition stellen konkrete Analyseleistungen empirische Manifestationen des theoretisch postulierten Konstrukts Analysekompetenz dar.

Zwischen den Annahmen eines solchen Konstrukts und den jeweils erhobenen Daten klafft – methodologisch gesehen – allerdings eine prinzipiell unüberwindbare Lücke, weil Daten empirischer, Konstrukte dagegen rein gedanklicher bzw. theoretischer Natur sind. Aufgrund dieses prinzipiellen Unterschiedes ist man im praktischen Forschungsprozess permanent mit der Frage der Validität konfrontiert: Repräsentieren die Daten in adäquater Weise das, was das theoretische Konstrukt (hier Analysekompetenz) postuliert?

Eine Variante der evidenzbasierten Validitätsprüfung stellt der *Mixed-Methods-Ansatz* dar, dessen Potenzial ich in meiner Untersuchung genutzt habe. Ihren Ausgangspunkt bildeten zwei Validierungsbelege, die von Plöger et al. (2015) und Plöger und Scholl (2014a) durch Konfirmatorische Faktorenanalysen erbracht wurden und bei denen sich zeigte: (1) Analysekompetenz kann als ein zweidimensionales Konstrukt interpretiert werden; dabei repräsentiert die inhaltliche Dimension das verfügbare pädagogische und fachliche Wissen, während die formale Dimension für die Komplexität der Informationsverarbeitung steht. (2) Im Sinne des sogenannten *well-known-group-Vergleichs* (s. z.B. Cronbach & Meehl, 1955) konnte zudem nachgewiesen werden, dass sich Personen, denen man vorab eine unterschiedliche hohe Analysekompetenz unterstellen kann, auch tatsächlich in den erzielten Testscores bei der Messung der Analysekompetenz unterscheiden.

Beide Validierungsbelege stellten wichtige Orientierungspunkte für meine Untersuchung dar, indem ich ebenfalls das von Plöger und Scholl definierte Verständnis von Analysekompetenz zugrunde legte und die verwendete Stichprobe neun ExpertInnen und neun NovizInnen umfasste, um starke Unterschiede in den Analyseleistungen sichtbar machen zu können. Als Stimulus wurde den ProbandInnen eine videographierte Unterrichtsstunde vorgeführt. Auf Grundlage dieser Unterrichtsstunde wurden anschließend *fokussierte, leitfadengestützte* Interviews geführt.

Die Auswertung der Interviews erfolgte in zwei Schritten: (1) Die transkribierten Interviews wurden im Sinne der Qualitativen Inhaltsanalyse nach Mayring anhand eines umfangreichen Kategoriensystems ausgewertet. (2) In einem weiteren Schritt wurden die ermittelten Aktivierungen pädagogischen und fachlichen Wissens weiter quantifiziert, um Signifikanzprüfungen, Reliabilitätsanalysen sowie Trennschärfe- und Korrelationsberechnungen durchführen zu können.

Die Ergebnisse meiner Untersuchung lieferten eindeutige Belege, mit denen die zweidimensionale Struktur des Konstrukts Analysekompetenz auch auf qualitativem Wege mittels der hier angewandten Qualitativen Inhaltsanalyse repliziert werden konnte. Darüber hinaus konnten die im Sinne des *well-known-group-Vergleichs* vermuteten Unterschiede zwischen ExpertInnen und NovizInnen bei der Analyse von Unterricht ebenfalls repliziert werden.

Abstract

The present work focuses on the analytical competence of teachers, defined as the ability of teachers to properly perceive and assess their own and others' lessons in terms of the effectiveness of learning. In the sense of this definition, concrete analytical performances represent empirical manifestations of the theoretically postulated construct analytical competence. From a methodological point of view, however, there is in principle an insurmountable gap between the assumptions of such a construct and the data collected, because data are empirical, while constructs are of pure conceptual or theoretical nature. Due to this difference in principle, the question of validity arises permanently in the practical research process: do the data adequately represent what the theoretical construct (here analytical competence) claims?

One variant of evidence-based validity checking is the *mixed-methods approach*, being used in my research project presented in this book. The starting point were two proofs of validity, provided by Plöger et al. (2015) and Plöger and Scholl (2014a) by confirmatory factor analyses showing that: (1) analytical competence can be defined as a construct having two dimensions – a content and a formal dimension. The content dimension is made up of two factors: pedagogical knowledge and content knowledge. The formal dimension of the construct is defined using three main levels and five substages of complexity with which teachers can process information relevant for teaching. (2) In the sense of the so-called *well-known-group comparison* (e.g. Cronbach & Meehl, 1955), it could also be proven that persons, assumed to have a different level of analytical competence in advance, actually did so in the test scores obtained by measuring the analytical competence. Both successful proofs of validity provided important reference points for my study, by sharing the concept of analytical competence as defined by Plöger and Scholl, and by using a sample which comprised nine experts and nine novices in order to make visible large differences in the analytical performances. As a stimulus the participants were shown a physics lesson on optics in which the Snell's law of refraction was being studied. In the subsequent interview, subjects were asked to comment on the lesson they had observed.

The analysis of the data was undertaken in two steps: (1) First, the interviews were transcribed in full and analysed using the qualitative content analysis method (Mayring, 2014). (2) In the second stage, the ascertained activations of pedagogical and content knowledge were quantitatively processed in order to calculate the reliability of the codings and the correlations of the categories between the factors of the content and formal dimensions. Furthermore, I used the Mann-Whitney Test to check whether the experts and novices differed significantly in their analytical performance.

The results of my investigation provided clear evidence to replicate the two-dimensional structure of the construct analytical competence in a qualitative way using the content analysis applied here. In addition, the well-known-group comparison confirmed the supposed differences between experts and novices in the analysis of lessons.

Inhalt

Einleitung	11
1 Theoretische Begründung des Konstrukts Analysekompetenz	19
1.1 Lehrerkompetenzforschung – Wissen als zentrale Voraussetzung für gute Analyseleistungen	19
1.2 Definition der Facetten des didaktisch-methodischen Wissens	23
1.3 Definition der Facetten des fachlichen (incl. fachdidaktischen) Wissens	27
1.4 Das Paradigma der Expertiseforschung – Komplexität der Informationsverarbeitung	32
1.4.1 Ergebnisse der Expertiseforschung	32
1.4.2 Fünfstufiges Kompetenzmodell nach Plöger und Scholl	35
1.5 Methode und Ergebnisse der Untersuchung von Plöger und Scholl (2014)	38
1.5.1 Anlage der Studie	38
1.5.2 Zentrale Ergebnisse der Studie von Plöger und Scholl (2014)	40
1.6 Quantitativer Validierungsversuch von Plöger, Scholl und Seifert (2015)	41
1.6.1 Modellprüfung mittels Konfirmatorischer Faktorenanalysen	41
1.6.2 Über CFA ermittelte faktorielle Struktur von Analysekompetenz	46
1.6.3 Relevanz der Ergebnisse für den durchzuführenden qualitativen Validierungsversuch	48
2 Methodisches Vorgehen	49
2.1 Qualitative Inhaltsanalyse als methodischer Zugriff	49
2.2 Vorüberlegungen zur Datengenerierung	52
2.3 Datengenerierung mittels Interviews	53
2.3.1 Das Interview als Datengenerierungsquelle	53
2.3.2 Fokussierte Interviews	54
2.3.3 Der Interviewleitfaden	55
2.3.4 Auswahl der TeilnehmerInnen	57
2.3.5 Durchführung der Interviews	59
2.4 Datenaufbereitung – Transkription	60
2.5 Kategorienbasierte Kodierung der Transkripte	62
2.5.1 Entwicklung eines Kategoriensystems	62
2.5.2 Bestimmung der Analyseeinheiten	63
2.5.3 Konsensuelles Kodieren zur Überprüfung der Bildung und Anwendung von Kategorien	64
2.5.4 Prüfung der Reliabilität	66
2.5.5 Prüfung der Validität	69
2.6 Ablaufmodell der Qualitativen Inhaltsanalyse	70

3 Theoretische Herleitung der Kategorien	73
3.1 Kategorien für didaktisch-methodisches Wissen	75
3.1.1 Bezüge zum Modell der Struktur- und Prozessqualität von Unterricht	75
3.1.2 Prozessmerkmale von Unterricht	77
3.1.3 Kategorien für didaktisch-methodisches Wissen	85
3.2 Kategorien für fachliches (incl. fachdidaktisches) Wissen	93
3.2.1 Das Vorgehen zur Bildung von Kategorien für fachliches und fachdidaktisches Wissen	93
3.2.2 Gegenstand und Methode der Physik als indirekte Bezugsgröße der Kategorienbildung	95
3.2.3 Auswahl der Kategorien für fachliches und fachdidaktisches Wissen ...	100
3.2.4 Kategorien für fachliches (incl. fachdidaktisches) Wissen	109
3.3 Kategorien für die formale Qualität des Wissens (Verstehenstiefe)	120
3.3.1 Das Kompetenzmodell von Plöger und Scholl als Bezugsrahmen	120
3.3.2 Bestimmung der Kategorien für die formale Qualität des Wissens (Verstehenstiefe) in Entsprechung zu den fünf Stufen des Kompetenzmodells von Plöger und Scholl	123
3.3.3 Kategorien und Ankerbeispiele für die formale Qualität des Wissens ...	126
4 Hypothesen	131
5 Ergebnisdarstellung	135
5.1 Überprüfung der Inter- und Intracoderreliabilität	135
5.2 Ergebnisse der Qualitativen Inhaltsanalyse	136
5.2.1 Unterschiede zwischen ExpertInnen und NovizInnen	137
5.2.2 Unterschiede zwischen ExpertInnen und NovizInnen hinsichtlich des fachlichen Hintergrundes	140
5.3 Vorstellung beispielhafter Aussagen	146
5.3.1 Beispielhafte Aussagen für die inhaltliche Dimension	147
5.3.2 Beispielhafte Aussagen für die formale Dimension	153
5.4 Quantifizierung der Ergebnisse	156
5.4.1 Reliabilitätsanalysen der bisherigen Ergebnisse	158
5.4.2 Endgültige Ergebnisse der Qualitativen Inhaltsanalyse	164
5.4.3 Beurteilung der Unterschiede zwischen ExpertInnen und NovizInnen ...	170
6 Zusammenfassung	173
7 Ausblick	179
7.1 Das Potenzial des Mixed-Methods-Ansatzes zur Konstruktvalidierung	179
7.2 Ein Ausblick auf die „Amalgamthese“ Shulmans	181
Verzeichnisse	
Literaturverzeichnis	185
Abbildungsverzeichnis	209
Tabellenverzeichnis	209
Anhang	213

Einleitung

Die vorliegende Arbeit berichtet über eine qualitative Studie, die ich als Wissenschaftlicher Mitarbeiter im Rahmen des von der Deutschen Forschungsgemeinschaft (DFG) geförderten Forschungsprojekts „Analysekompetenz – Diagnose und Entwicklung“ (Leitung: Prof. Dr. Wilfried Plöger & Dr. Daniel Scholl; in Kooperation mit Dr. Andreas Seifert) durchgeführt habe. Unter Analysekompetenz verstehen Plöger und Scholl die Fähigkeit, „die Qualität ... von beobachtetem Unterricht im Hinblick auf dessen Lernwirksamkeit angemessen erfassen und bewerten zu können“ (Plöger & Scholl, 2014a, S. 88).

Paradigmatisch gesehen, ist dieses DFG-Projekt dem *Experten-Novizen-Paradigma* verpflichtet (s. z.B. Bromme, 1997; Bromme & Rheinberg, 2006; Gruber, 2001; Gruber & Stöger, 2011; Krauss, 2011; Krauss & Bruckmaier, 2014; König, 2010; Mulder & Gruber, 2011). Das gilt auch für meine Studie, in der ich Interviews mit ExpertInnen und NovizInnen geführt und mittels Qualitativer Inhaltsanalyse (s. z.B. Mayring, 2014, 2015) ausgewertet habe, wobei meine zentrale Forschungsfrage in Anlehnung an die Definition des Begriffs der Analysekompetenz von Plöger und Scholl lautete: Welche Dispositionen ermöglichen es ExpertInnen im Vergleich zu NovizInnen, Unterricht angemessen unter dem Aspekt der Lernwirksamkeit zu analysieren?

Der Ablauf der einzelnen Phasen meiner Studie lässt sich im Groben folgendermaßen skizzieren: Zu Beginn der Studie wurde den TeilnehmerInnen (9 ExpertInnen/9 NovizInnen) eine Physik-Stunde (Thema: Snelliussches Brechungsgesetz) präsentiert. Die im Anschluss an diese Videopräsentation geführten Interviews wurden transkribiert und in einem weiteren Schritt anhand von Kategorien analysiert. Diese Kategorien repräsentierten die kognitiven Dispositionen der ProbandInnen in Form ihres *pädagogischen* und *fachlichen* Wissens, sodass sich die konkrete Fragestellung ergab: Welche Facetten pädagogischen und fachlichen Wissens ermöglichen ExpertInnen und NovizInnen eine angemessene Analyse von Unterricht?

In systematischer Hinsicht ist die meiner Arbeit zugrunde liegende Forschungsfrage somit in der *Lehrerkompetenzforschung* angesiedelt (s. z.B. Baumert & Kunter, 2006; Herzmann & König, 2016; König, 2010, 2016; Mulder & Gruber, 2011; Terhart, 2007). Wenn man aus der Perspektive dieser Forschung mit Baumert und Kunter den Kern der Tätigkeit von Lehrpersonen im Unterrichten sieht (s. Baumert & Kunter, 2011a, 2013), und diese Tätigkeit permanent im Kreislauf der Planung, Durchführung und Analyse von Unterricht erfolgreich gemeistert werden muss, dann stellt die Analysekompetenz in diesem Zyklus eine zentrale Kompetenz dar, weil sie der Lehrperson „Aufschluss über die Qualität ihrer Planung und die Effektivität des durchgeführten Unterrichts“ (Plöger & Scholl, 2016, S. 181) gibt.

In methodologischer Hinsicht stellen Kompetenzen *Konstrukte* dar: Sichtbare Leistungen in Form beobachtbarer Daten (hier Analyseleistungen) werden auf latente Fähigkeiten (kognitive, motivationale Dispositionen) zurückgeführt. Ob die beobachteten Daten als Folge solcher latenten Dispositionen gedeutet werden dürfen, bedarf einer differenzierten Rechtfertigung, die letztlich auf die Kernfrage hinausläuft: Können die erhobenen Daten

mit den theoretischen Annahmen interpretiert werden, die das jeweilige Konstrukt definieren? Im Falle meiner Studie sollte also geprüft werden, ob die durch eine kategoriengeleitete Analyse der Interviews erzeugten Daten Belege für die theoretische Grundannahme darstellen, dass das Gelingen der Analyse von Unterricht vom verfügbaren pädagogischen und fachlichen Wissen der ProbandInnen abhängt.

Diese Frage zielt methodologisch gesehen auf die Frage nach der *Validität des Konstrukts Analysekompetenz*. Validität stellt, neben der Objektivität und Reliabilität, das bedeutendste und komplexeste Gütekriterium empirischer Sozialforschung dar (s. z.B. Blömeke, 2013; Hartig, Frey & Jude, 2012; Wilhelm & Kunina-Habenicht, 2015).¹

Im traditionellen Verständnis wurde und wird Validität als die Eigenschaft eines Tests aufgefasst, der dann als valide gilt, wenn er das misst, was er zu messen vorgibt bzw. „misst, was er messen soll“ (s. z.B. Eid & Schmidt, 2014; Friedrichs, 1990; Lienert & Raatz, 1998; Rost, 2004; Sedlmeier & Renkewitz, 2013).

In den vergangenen Jahren hat sich allerdings die Bedeutung von Validität geändert, so dass die ursprüngliche Auffassung von Validität (als Eigenschaft eines Tests) inzwischen abgelehnt wird: „It is incorrect to use the unqualified phrase, the validity of the test“ (AERA, APA, & NCME, 2014, S. 11). Infolgedessen wird Validität nicht mehr als Testeigenschaft aufgefasst, sondern Validität „refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests.“ (ebd., S. 11) Das bedeutet, dass nicht mehr das Testverfahren selbst überprüft wird, sondern die *Interpretationen* von Testwerten *validiert* werden (s. Frey, 2015; Furr & Bacharach, 2008; Kane, 2013). *Validierung* kann, in Anlehnung an Frey (2015), somit als „Sammlung relevanter Informationen“ verstanden werden, „die belastbare, wissenschaftlich abgesicherte Unterstützung angestrebter Testwertinterpretationen liefern“.²

Einen entscheidenden Anteil an diesem veränderten Verständnis von Validität haben die Arbeiten von Messick (1989, 1990, 1995). Nach Messick ist Validität „an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment.“ (Messick, 1989, S. 13; 1995, S. 741) Validität wird somit zum allgemeinen Bewertungsmaßstab erklärt, um Interpretationen von Testwerten zu beurteilen.

Auch im Sinne dieses neueren Verständnisses von Validität behalten „traditionelle“ Arten der Validitätsprüfung allerdings ihre Funktion. Das gilt insbesondere für diese drei Arten der Validitätsprüfung (s. AERA, APA, & NCME, 1985; Blömeke, 2013; Döring & Bortz, 2016; Bühner, 2011; Hartig et al., 2012):

-
- 1 Die Objektivität umfasst die *Unabhängigkeit* der Messung bzw. der Testergebnisse von den Rahmenbedingungen der Testung (s. z.B. Blömeke, 2013, Bühner, 2011). Mit der Reliabilität wird die *Zuverlässigkeit* der Messung bzw. der Ergebnisse erfasst. Während eine hohe Reliabilität eine objektive Erhebung erfordert, ist die Reliabilität wiederum eine Voraussetzung für eine hohe Validität (s. z.B. Bühner, 2011; Furr & Bacharach, 2008). Die Validität nimmt somit eine Sonderstellung ein, denn „Objektivität und Reliabilität sind nur notwendige Minimalanforderungen an ein Messinstrument. Das Hauptziel ist dagegen die Konstruktion möglichst valider Messinstrumente“ (Diekmann, 2009, S. 256).
 - 2 Das Zitat stammt aus einem Vortrag von A. Frey aus dem Jahre 2015 zur 80. Tagung der *AEPF* (Arbeitsgruppe für Empirische Pädagogische Forschung) in Göttingen. Ich danke ihm für die Bereitstellung des Foliensatzes.

- *Inhaltsvalidität*: Dabei wird „das Verhältnis zwischen dem zu erfassenden Merkmal und den Test- bzw. Iteminhalten“ (Hartig et al., 2012, S. 146) durch *ExpertInnenurteile* geprüft.
- *Kriteriumsvalidität*: Hierbei wird das Ergebnis eines zu validierenden Konstrukts mit einem externen Kriterium (Außenkriterium) in Beziehung gesetzt (s. Balderjahn, 2003; Blömeke, 2013).³
- *Konstruktvalidität*: Der Grundgedanke ist, dass „Testergebnisse vor dem Hintergrund eines theoretischen Konstrukts“ (Hartig et al., 2012, S. 153) interpretiert werden. Es wird also geprüft, ob die „Gültigkeit der Interpretation empirischer Daten als Indikatoren für theoretische Konstrukte“ (Döring & Bortz, 2016, S. 98) vorliegen.⁴

Während die Konstruktvalidierung Mitte des 20. Jahrhunderts zunächst als Ergänzung zur Inhalts- und Kriteriumsvalidität gesehen wurde, stellt sie seit den 1970er Jahren einen „übergeordneten Zugang zur Bestimmung der Validität“ (Hartig et al., 2012, S. 153; Blömeke, 2013; Cronbach & Meehl, 1955; Frey, 2015) dar: „Ein Ziel bei der Untersuchung der Konstruktvalidität besteht vereinfacht ausgedrückt in der Überprüfung theoretischer Annahmen über Zusammenhangsstrukturen latenter Konstrukte anhand empirischer Daten.“⁵ (Hartig et al., 2012, S. 156) Eine solche Überprüfung der Konstruktvalidität kann aber nicht hinreichend im Rahmen einer einzelnen empirischen Studie beantwortet werden. Vielmehr gilt: „A sound validity argument integrates various strands of evidence into a coherent account of the degree to which existing evidence and theory support the intended interpretation“ (AERA et al., 2014, S. 21) von Daten für die spezifische Verwendung.

In dieser Forderung, verschiedene Arten der Evidenzprüfung zu nutzen, sehe ich eine große Chance zur Beantwortung meiner Forschungsfrage und gehe deshalb mit Leech und Dellinger davon aus, „that construct validity and the construct validation process are relevant to mixed methods research“ (Dellinger & Leech, 2007, S. 321).

Grundlegendes Ziel einer Mixed-Methods-Untersuchung ist es, dass „the researcher might put quantitative and qualitative findings side-by-side to check how congruent they are“ (Bryman, 2008, S. 163). Dabei „geht es vor allem darum, eine pragmatische Verknüpfung von qualitativer und quantitativer Forschung zu ermöglichen“ (Flick, 2011b, S. 76, s. auch Johnson & Onwuegbuzie, 2004; Tashakkori & Teddlie, 1998, 2003b).⁶

3 Externe Kriterien können dabei *Ergebniskriterien* (z.B. Schulnoten), *Verhaltenskriterien* (z.B. Feedbackverhalten von Lehrpersonen) oder *Eigenschaftskriterien* (z.B. motivationale Einstellungen von Lehrpersonen) sein (s. z.B. Blömeke, 2013).

4 Eine bedeutende Arbeit in diesem Zusammenhang haben Cronbach und Meehl (1955) vorgelegt. Unter einem Konstrukt verstehen sie Folgendes: „A construct is some postulated attribute of people, assumed to be reflected in test performance. In test validation the attribute about which we make statements in interpreting a test is a construct.“ (Cronbach & Meehl, 1955, S. 283).

5 Weitere Möglichkeiten zur Überprüfung der Konstruktvalidität könnten, neben der Prüfung der faktoriellen Struktur eines Konstrukts, die „Prüfung der Konstruktrepräsentation über die Vorhersage von Itemschwierigkeiten und die Prüfung der Verortung des Konstrukts in einem nomologischen Netzwerk“ (Blömeke, 2013, S. 7) sein.

6 Die Entwicklung des Mixed-Methods-Ansatzes kann nur im Hinblick auf die historische Entwicklung in der empirischen Sozialforschung erklärt werden (s. Kelle, 2008, 2014). Die dort seit den 1920er Jahren stattfindende Entwicklung hat zu zwei unabhängigen Methodentraditionen (qualitativ/quantitativ) geführt, die sich „im Rahmen eigener methodologischer Traditionen und Denkschulen entwickelt“ (Kelle, 2014, S. 154) haben. Doch die damit verbundenen unterschiedlichen Forschungsziele und Gütekriterien führten zu teilweise

Eine einheitliche Definition von Mixed-Methods liegt jedoch bisher nicht vor (s. Johnson, Onwuegbuzie, & Turner, 2007). Es existiert aber ein grundlegend geteiltes Verständnis, was man unter Mixed-Methods-Ansätzen verstehen kann: „Mixed methods research is the type of research in which a researcher or team of researchers combine elements of qualitative and quantitative research approaches (e.g. use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purpose of breadth and depth of understanding and corroboration ...” (Johnson et al., S. 123).

Im Vergleich zu seinem zentralen Stellenwert „the concept of validity has been addressed sparingly in the mixed methods literature” (Dellinger & Leech, 2007, S. 314) seit längerer Zeit. Mittlerweile liegen allerdings sowohl methodologische Konzepte von Mixed-Methods-Ansätzen zur Validitätsprüfung (z.B. Dellinger & Leech, 2007; Leech, Dellinger, Brannagan, & Tanaka, 2010; Onwuegbuzie, Bustamente, & Nelson, 2010) als auch empirische Studien vor, in denen Mixed-Methods-Ansätze zur Beurteilung der Validität eingesetzt wird (s. z.B. Daigneault & Jacob, 2014; Latcheva, 2009; Luyt, 2011; Morell & Tan, 2009). Trotzdem wird immer noch konstatiert, dass „more contributions about why and how to mix methods for validation purposes are needed” (Daigneault & Jacob, 2014, S. 6). In dieser Arbeit nutze ich das Potenzial des Mixed-Methods-Ansatzes, indem ich empirische Daten auf *qualitativem* Wege sammle und prüfe, ob sie die Validität des Konstrukts Analysekompetenz stützen. Die Ergebnisse dieses qualitativen Vorgehens setzte ich dann in Beziehung zu einem *quantitativen* Validierungsversuch, der von Plöger, Scholl und Seifert (2015) durchgeführt wurde, um über den Vergleich der Ergebnisse festzustellen, ob beide Ansätze, der qualitative und der quantitative, die zugrunde liegenden theoretischen Annahmen des Konstrukts Analysekompetenz valide stützen.

Beide Studien sind in einem *sequential explanatory design* (Ivankova, Creswell & Stick, 2006) miteinander verknüpft.⁷ Bei einem sequentiellen Design sind grundsätzlich zwei Varianten denkbar: Entweder es wird mit einer qualitativen Studie oder es wird mit einer quantitativen Studie begonnen. Die Wahl des Beginns bestimmt, ob man von einem „qualitativ-vertiefenden Design“ oder einem „quantitativ-verallgemeinernden Design“ sprechen kann (Kuckartz, 2014, S. 77). Im vorliegenden Fall wurde mit dem quantitativen Validierungsversuch begonnen (s. Plöger, Scholl & Seifert, 2015), an den sich die hier vorliegende qualitativ-vertiefende Untersuchung anschließt.

Die Frage der Konstruktvalidierung steht in diesem Fall in engem Zusammenhang mit dem Konzept der *Kompetenz* (s. z.B. Gruber, 2004; Klieme & Leutner, 2006; Oser & Renold, 2005; Shavelson, 2013; Terhart, 2007; Zlatkin-Troitschanskaia & Seidel, 2011). Dieses Konzept lässt sich mit Blömeke, Gustafsson und Shavelson (2015) durch zwei Pole charakterisieren:

kontrovers geführten Debatten, die in ihrer „Hoch-Zeit“ mit dem Slogan „paradigm wars“ (Gage, 1989) umschrieben wurden. Diesem „paradigm war“ stellt sich die Mixed-Methods-Bewegung entgegen und erklärt ihn für beendet (Flick, 2011b, S. 76). Seit den 1990er Jahren zielt diese Bewegung darauf, eine „third methodological movement“ (Tashakkori & Teddlie, 2003a) zu etablieren.

⁷ Bisher liegt keine einheitliche Systematik vor, wie Mixed-Methods-Studien durchgeführt werden sollten, sondern es konkurrieren eine Vielzahl unterschiedlicher Taxonomien miteinander (s. z.B. Kelle, 2014; Kuckartz, 2014; Morse, 1991; Schreier & Echterhoff, 2013). Einen Systematisierungsvorschlag haben unter anderem Creswell und Plano Clark (2011) unternommen. Sie identifizieren sechs zentrale Untersuchungsanliegen für Mixed-Methods-Designs, von denen das explanatory (sequential) design eines ist. Weiterhin unterscheiden sie zwischen einem *convergent*, einem *exploratory*, einem *embedded*, einem *transformative* und einem *multiphase design*.

Den einen Pol bilden *latente Dispositionen* (kognitive und affektiv-motivationale Einstellungen), den anderen *manifeste Leistungen* (beobachtbares Verhalten), die als Indikatoren für die latenten Dispositionen betrachtet werden können.⁸

Prinzipiell klafft eine unüberwindbare Lücke zwischen diesen beiden Polen: Latente Dispositionen sind gedanklicher bzw. theoretischer und manifeste Leistungen dagegen empirischer Natur. Infolge dieses grundlegenden Unterschieds ist die Frage der Validität im praktischen Forschungsprozess permanent virulent: Repräsentieren die Daten in adäquater Weise das, was die theoretische Definition des Konstrukts postuliert?

Zur Veranschaulichung des Zusammenhangs zwischen diesem Verständnis von Kompetenz, dem realisierten sequential explanatory design und dem oben beschriebenen Verständnis des Konstrukts Analysekompetenz zu veranschaulichen, greife ich auf die folgende Grafik zurück:

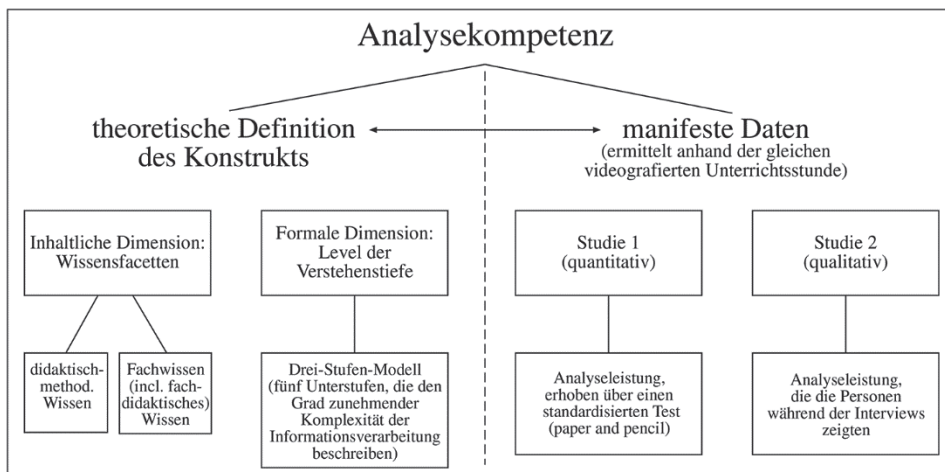


Abb. 1: Definition des Konstrukts Analysekompetenz und zeitliche Abfolge der beiden Validierungsstudien

⁸ Blömeke et al. (2015) unterscheiden in diesem Kontinuum drei Ebenen von Kompetenzkonstrukten: disposition; situation specific skills und performance. Dabei beziehen sich *dispositions* auf affektive, motivationale und kognitive Konstrukte. Diese Konstrukte werden meistens mittels Tests (Wissen) oder Persönlichkeits-tests (affektive und motivationale Einstellungen/Überzeugungen) in Studien untersucht. *Situation specific skills* sind dahingegen mehr aktionsorientiert, und die Art der kognitiven Prozesse ist in diesem Bereich deutlich komplexer, da Prozesse wie Wahrnehmung, Interpretation und Entscheidungsfindung untersucht werden. Infolge der Komplexität dieser situationsspezifischen Fähigkeiten wird auch die Messung komplizierter. In vielen Fällen wird ein Video als Impulsgeber eingesetzt, um diese Fähigkeiten zu testen. In der vorliegenden Arbeit wird das Konstrukt Analysekompetenz als situation specific skill interpretiert. Als dritten Bereich benennen Blömeke et al. (2015) die *performance*. Diese ist zwar (mehr oder weniger) direkt beobachtbar, aber auch hier liegt ein latentes Konstrukt zugrunde, das definiert, was es bedeutet, ob eine Person in der Lage ist, die Anforderungen in einem bestimmten Bereich zu meistern (hier: die Fähigkeit, Unterricht angemessen analysieren zu können).

Diese drei Ebenen bauen hierarchisch aufeinander auf: Dispositions sind Voraussetzungen für situation-specific skills, welche wiederum Voraussetzungen für die performance sind. Somit wird die Struktur des Wissens (dispositions) in der Struktur der Analysekompetenz (situation specific skill) gespiegelt.

Auf der rechten Seite der Grafik ist die Abfolge der beiden Studien dokumentiert, durch die die entsprechenden manifesten Daten erhoben wurden. Auf der linken Seite sind die theoretischen Annahmen skizziert, durch die Analysekompetenz als zweidimensionales Konstrukt definiert wird. Dabei repräsentiert die erste Dimension die *inhaltliche* Qualität des Wissens, die zwei Wissensfacetten als Faktoren umfasst: das didaktisch-methodische Wissen einerseits und das *fachliche* (incl. *fachdidaktische*) Wissen andererseits. Die zweite Dimension stellt die *formale* Qualität des Wissens dar, die den Grad der für die Analyseleistungen notwendigen Komplexität der Informationsverarbeitung abbildet und sich in drei Faktoren (Analytische Kompetenz, Synthetische Kompetenz und Prozesskompetenz) mit insgesamt fünf Substufen unterteilen lässt. Im Sinne dieses Verständnisses von Analysekompetenz sind nur solche Personen in der Lage, Unterricht angemessen zu analysieren, die sowohl in *inhaltlicher* als auch in *formaler* Hinsicht über eine hinreichend hohe Qualität des Wissens verfügen.

Diese theoretischen Annahmen wurden von Plöger et al. (2015) auf quantitativem Wege mittels Konfirmatorischer Faktorenanalysen (*confirmatory factor analysis* – CFA) geprüft (Studie 1). Die dadurch ausgewiesene und mit entsprechenden Fit-Werten belegte faktorielle Struktur des Konstrukts Analysekompetenz dient nun in meiner Studie als hypothetische Basis, an das das erkenntnisleitende Interesse meiner Studie anknüpft (Studie 2). Mittels Qualitativer Inhaltsanalyse soll geprüft werden, ob sich die über CFA ermittelte Faktorenstruktur auch auf qualitativem Wege replizieren lässt. Inhaltlich gesehen orientieren sich demnach beide Verfahren (Qualitative Inhaltsanalyse, CFA) an einem *identischen* Verständnis von Analysekompetenz. In methodischer Hinsicht stellen sie dagegen im Sinne eines Mixed-Methods-Designs zwei *unterschiedliche* Wege zur Validierung dieses Konstrukts dar.

Aufbau der Arbeit

Im Folgenden skizziere ich den Aufbau der vorliegenden Arbeit, die sich insgesamt in sieben Kapitel gliedert.

In Kapitel 1 werde ich – ausführlicher als in dieser Einleitung – den theoretischen und empirischen Hintergrund meiner Untersuchung darstellen, sodass dieses Kapitel inhaltlich zweigeteilt ist: Im ersten Teil werden die theoretischen Annahmen, die dem Konstrukt Analysekompetenz zugrunde liegen, präsentiert. Daran anschließend werden in einem zweiten Teil die empirischen Studien von Plöger und Scholl (2014a, 2014b) und Plöger et al. (2015) vorgestellt, um die Relevanz der Ergebnisse dieser Untersuchungen für meine Arbeit zu erörtern.

Kapitel 2 stellt den methodischen Zugriff der Arbeit dar, der sich in Kürze so zusammenfassen lässt: Auf Grundlage einer videographierten Unterrichtsstunde wurden *fokussierte, leitfadengestützte* Interviews geführt, per Tonbandgerät aufgezeichnet und anschließend *vollständig* transkribiert. Um die Interviews auswerten und die faktorielle Struktur des Konstrukts Analysekompetenz abbilden zu können, habe ich mich für die Qualitative Inhaltsanalyse (s. z.B. Mayring, 2010, 2014, 2015) entschieden. Mit ihr werden große Textmengen durch Anwendung von Kategorien auf zentrale Aussagen reduziert. Die Qualitative Inhaltsanalyse ist also insofern eine äquivalente Methode zur CFA, weil sie letztlich ein *datenreduzierendes* Verfahren darstellt.

Im Anschluss an den methodischen Teil werde ich in Kapitel 3 die Kategorien meiner Untersuchung theoretisch herleiten. Dazu ist es notwendig, auf die Forschung zur professionellen Handlungskompetenz von Lehrpersonen einzugehen und einen Überblick über aktuelle Konzeptualisierungen der professionellen Kompetenz von Lehrpersonen zu geben, wie sie z.B. in Studien wie *COACTIV (Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung mathematischer Kompetenz;* Blum, Krauss, & Neubrand, 2011; Krauss et al., 2004; Kunter, Klusmann, & Baumert, 2009; Kunter et al., 2011), *TEDS-M (Teacher Education and Development Study: Learning to Teach Mathematics,* Blömeke, Kaiser, & Lehmann, 2010a, 2010b), *LEK (Längsschnittliche Erhebung pädagogischer Kompetenzen von Lehramtsstudierenden,* König & Seifert, 2012) und *MT 21 (Mathematics Teaching in the 21st Century,* Blömeke, Kaiser, & Lehmann, 2008) vorgenommen wurden. Die einzelnen Teilkapitel sind an der faktoriellen Struktur des Konstrukts Analysekompetenz orientiert. In den einzelnen Teilkapiteln werde ich Bezüge zur empirischen Lehr-Lern-Forschung, zur Allgemeinen Didaktik und zur Kognitionspsychologie herstellen, um die Bildung meiner Kategorien sowohl empirisch als auch theoretisch zu begründen. Zum Abschluss der jeweiligen Teilkapitel werde ich die Kategorien vorstellen, die letztlich das „Herzstück“ der Untersuchung darstellen. Die Arbeit ist als Validierungsversuch geplant. Validierungen können aber nur vorgenommen werden, wenn entsprechende Hypothesen vorliegen, die geprüft werden können. Daher werde ich in Kapitel 4 die der Arbeit zugrunde liegenden Hypothesen vorstellen.

In Kapitel 5 werden die Untersuchungsergebnisse präsentiert. Dieses Kapitel untergliedert sich in einen qualitativen und einen quantitativen Teil. Im qualitativen Teil werde ich entsprechende Textbelege liefern, die die Annahmen meiner Arbeit stützen. Diese qualitativen Daten werden anschließend *quantitativ* weiterverarbeitet. Dabei wird jedes einzelne Kodierereignis als Indiz für die spezifische Wissensaktivierung einer Wissensfacette ver-

standen, sodass jedes Kodierereignis eine richtige, von den KodiererInnen akzeptierte Anwendung von Wissen repräsentiert. Unterschiede, die sich bei den Zahlen bzw. Summen ergeben, werden dann als entsprechende Unterschiede hinsichtlich der Ausprägung der Wissensaktivierung und somit der Analyseleistung ausgelegt.

In Kapitel 6 werden die zentralen Ergebnisse entlang der Fragestellung zusammengefasst und vor dem Hintergrund der Hypothesen meiner Untersuchung beurteilt.

Zum Schluss der Arbeit gebe ich in Kapitel 7 einen kurzen Ausblick auf weiterführende Forschungsfragen, die sich aus der theoretischen und forschungsmethodologischen Anlage der Untersuchung ergeben könnten.

Die vorliegende qualitative Studie „Wie analysieren ExpertInnen und NovizInnen Unterricht?“ untersucht die Analysekompetenz von Lehrpersonen – verstanden als jene Fähigkeit, die Qualität von Unterricht unter dem Aspekt der Lernwirksamkeit angemessen beurteilen zu können. Im Fokus stand dabei die Frage, welche Faktoren die Ausprägung dieser Kompetenz bestimmen. Die notwendigen Daten wurden durch leitfadengestützte Interviews generiert, in denen jeweils neun ExpertInnen und NovizInnen eine videografierte Unterrichtsstunde analysierten. Die inhaltsanalytische Auswertung der auf diese Weise gesammelten Aussagen zeigte, dass die ExpertInnen – im Gegensatz zu den NovizInnen – über umfangreiches pädagogisches sowie fachliches (incl. fachdidaktisches) Wissen verfügten und mit diesem Wissen nicht nur einzelne Situationen, sondern auch den Unterrichtsprozess als komplexes Ganzes erfassen konnten.



Der Autor

Matthias Krepf, Jahrgang 1985, promovierte 2018 an der Humanwissenschaftlichen Fakultät der Universität zu Köln. Im Anschluss an sein Lehramtsstudium arbeitete er an der Universität zu Köln zunächst am Lehrstuhl für Allgemei-

ne Didaktik und Schulpädagogik und ist seit 2018 als Wissenschaftlicher Mitarbeiter und Lehrkraft für besondere Aufgaben am Lehrstuhl für Empirische Schulforschung (Schwerpunkt quantitative Methoden) angestellt. Seine Forschungsinteressen liegen in der empirischen Lehrer*innenbildungsforschung sowie der Schul- und Unterrichtsforschung.

978-3-7815-2312-8



9 783781 523128