

# 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).<sup>1</sup>

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“.<sup>2</sup>

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).<sup>3</sup>
- *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.<sup>4</sup>

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.“<sup>5</sup> (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).<sup>6</sup>

---

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.<sup>7</sup> 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:

---

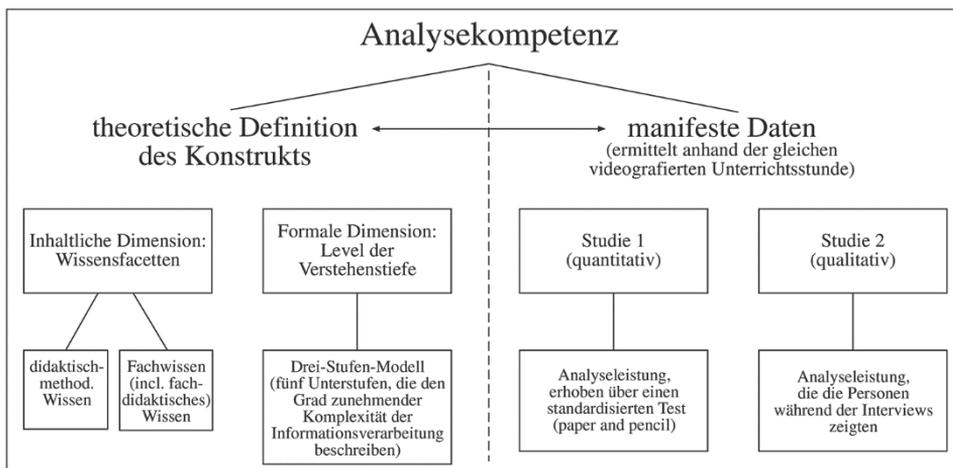
kontrovers geführten Debatten, die in ihrer „Hoch-Zeit“ mit dem Slogan „paradigm wars“ (Gage, 1989) umschrieben wurden. Diesem „paradigm war“ stellt sich die Mixed-Methods-Bewegung entgegen und erklärt ihn für beendet (Flick, 2011b, S. 76). Seit den 1990er Jahren zielt diese Bewegung darauf, eine „third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

<sup>7</sup> 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.<sup>8</sup>

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

<sup>8</sup> 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.

# 1. Theoretische Begründung des Konstrukts Analysekompetenz

In der vorangegangenen Einleitung habe ich das übergreifende Ziel meines Forschungsvorhabens umrissen, das auf die Beantwortung der Frage abzielt, ob sich die manifesten Daten, die ich mit meiner qualitativen Studie erhoben habe, als valide Belege für die faktorielle Struktur des Konstrukts Analysekompetenz interpretieren lassen.

Sollte dies der Fall sein, würde die von mir durchgeführte Qualitative Inhaltsanalyse den quantitativen Validierungsversuch von Plöger et al. (2015) stützen und damit exemplarisch das Potenzial von Mixed-Methods-Ansätzen für Fragen der (Konstrukt-)Validierung untermauern. Mit beiden methodischen Ansätzen wären dann empirische Daten auf unterschiedlichen Wegen generiert und ausgewertet worden, die kompatibel mit den theoretischen Annahmen des Konstrukts Analysekompetenz sind.

Um die zentrale Frage meines Forschungsvorhabens beantworten zu können, sind zwei aufeinander abgestimmte Schritte zu gehen: In einem ersten Schritt werden in diesem Kapitel die theoretischen Annahmen, die das Konstrukt Analysekompetenz charakterisieren, expliziert und vor dem Hintergrund der gegenwärtigen Forschungssituation legitimiert. Anschließend wird in einem zweiten Schritt (s. Kap. 2) dann zu begründen sein, inwiefern die Qualitative Inhaltsanalyse als geeignetes methodisches Instrument erscheint, adäquate empirische Daten für die Beantwortung der Fragestellung zu erzeugen und zu prüfen, ob diese manifesten Daten das theoretische Konstrukt decken.

In diesem Kapitel geht es zunächst um den ersten Schritt der theoretischen Legitimation des Konstrukts Analysekompetenz und damit um die Vorstellung der von Plöger und Scholl (2014a) sowie Plöger et al. (2015) vorgenommenen theoretischen Modellierung des Konstrukts Analysekompetenz. Anschließend werden das von ihnen gewählte Verfahren der Datenerhebung und -auswertung und die Ergebnisse ihres quantitativen Validierungsversuchs in Kürze dargestellt bzw. erläutert.

## 1.1 Lehrerkompetenzforschung – Wissen als zentrale Voraussetzung für gute Analyseleistungen

In der gegenwärtigen Forschung zur Kompetenz von Lehrpersonen wird Wissen als eine unabdingbare Disposition zur professionellen Bewältigung täglich anfallender Aufgaben und Herausforderungen verstanden (Baumert & Kunter, 2011a, 2013; König, 2010; Shulman, 1986, 1987). Das gilt nicht nur für die Planung und Durchführung von Unterricht, sondern auch für dessen Analyse (s. z.B. Blomberg, Renkl, Sherin, Borko, & Seidel, 2013; Borko, Jacobs, Eiteljorg, & Pittman, 2008; Brophy, 2004; Rosaen, Carlisle, Mihocko, Melnick, & Johnson, 2013). Denn kognitionspsychologisch gesehen lässt sich die Analyse

von Unterricht als ein Prozess der Aufnahme und Verarbeitung von Informationen verstehen, der *wissensbasiert* und *wissensgesteuert* abläuft. Demnach hängt die Ausprägung der Analysekompetenz vom Umfang und von der Organisation des verfügbaren Wissens ab. Um diese These theoretisch zu begründen und das Konstrukt Analysekompetenz zu legitimieren, greife ich in diesem Abschnitt zunächst auf die gegenwärtige Diskussion um die Bedeutung des Faktors Wissen zurück, um dann im folgenden Schritt zu erläutern, welche spezifischen Wissensfacetten für die Entwicklung und Ausprägung der Analysekompetenz eine wichtige Rolle spielen.

Die Diskussion um die Bedeutung des *Professionswissens* von Lehrpersonen beschäftigt ForscherInnen seit mehreren Jahrzehnten. Den Ausgangspunkt dieser Diskussion stellen die Klassifikationsvorschläge von Shulman (1986, 1987) dar, die zunächst im angloamerikanischen Raum intensiv rezipiert wurden. Daran schloss sich – auch im deutschsprachigen Raum – eine ausführliche und langanhaltende Diskussion zur Beschreibung der notwendigen Wissensbasis von Lehrpersonen an. Stellvertretend seien hier die Arbeiten von Bromme (1992, 2008) und der Forschungsgruppe um Weinert (Weinert, Schrader, & Helmke, 1990) genannt, die den (damaligen) wissenschaftlichen Diskurs prägten.<sup>9</sup>

Letztlich und trotz aller Disparatheit hinsichtlich der Konzeptualisierungen des professionellen Wissens von Lehrkräften hat diese Diskussion dazu geführt, dass *allgemeines pädagogisches Wissen*, *Fachwissen* und *fachdidaktisches Wissen* als unentbehrliche Aspekte professionellen Lehrerhandelns gesehen werden (s. z.B. Baumert & Kunter, 2011a, 2013; Blömeke, Suhl, & Kaiser, 2011; Borko, 2004; König, 2010; König et al., 2014; Munby, Russel, & Martin, 2002; Voss, Kunina-Habenicht, Hoehne, & Kunter, 2015; Voss, Kunter, Seiz, Hoehne, & Baumert, 2014) und dass diese Aspekte deshalb auch in entsprechenden Forschungsprojekten eine zentrale Rolle einnehmen. Das gilt z.B. für Projekte wie *COACTIV*, *TEDS-M*, *LEK* und *MT 21*.

Wie diese drei Aspekte – allgemeines pädagogisches Wissen, Fachwissen und fachdidaktisches Wissen – näher präzisiert werden, sei exemplarisch am Projekt *COACTIV* gezeigt. Das Anliegen von *COACTIV* besteht darin, „die individuellen Merkmale zu identifizieren, die Lehrkräfte für die erfolgreiche Bewältigung ihrer beruflichen Aufgaben benötigen“ (Baumert & Kunter, 2011a, S. 29), um darauf aufbauend ein Modell professioneller Kompetenz von Lehrpersonen zu entwickeln.

Zur Identifizierung entsprechender Merkmale und zur Entwicklung des Modells werden verschiedene (theoretische und empirische) Ansätze aufgegriffen, von denen zwei von zentraler Bedeutung sind: Erstens wird ein *professionsspezifischer* Zugang gewählt, der auf den oben angegebenen Arbeiten von Shulman (und Bromme) basiert. Zweitens wird

---

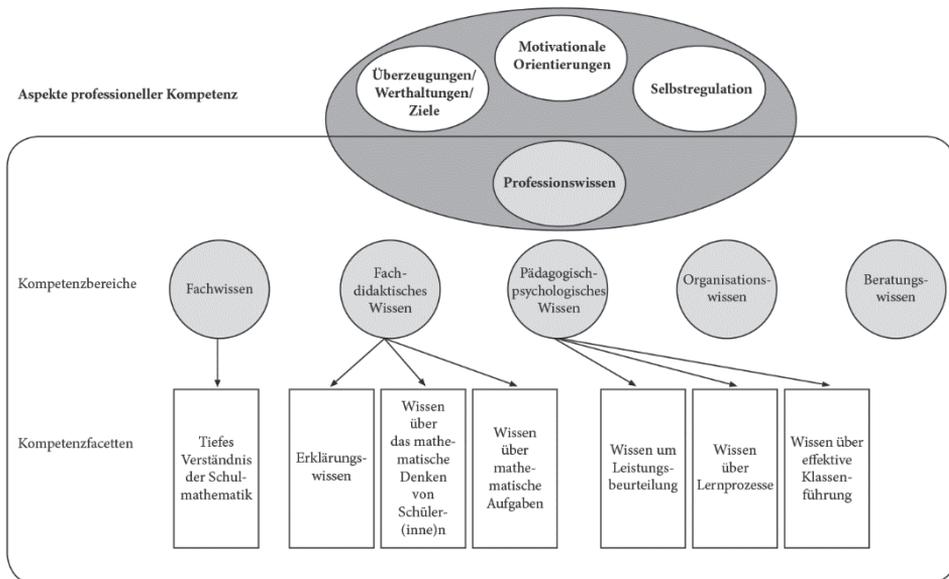
9 In den 1980er Jahren unterschied Shulman (1986) zunächst zwischen *general pedagogical knowledge* (allgemeines, fächerübergreifendes pädagogische Wissen), *subject matter content knowledge* (Fachwissen), *pedagogical content knowledge* (fachdidaktisches Wissen) und *curricular knowledge* (curriculares Wissen). Später wurde diese Klassifikation durch weitere Wissensfacetten erweitert, sodass diese nun zusätzlich *knowledge of learners and their characteristics* (Vorwissen der SchülerInnen), *knowledge of educational contexts* (Wissen über den Bildungskontext), sowie *knowledge of educational ends, purposes, and values, and their philosophical and historical grounds* (Wissen über pädagogische Ziele, Zwecke) umfasste (s. Shulman, 1987, S. 8). Bromme (1992, 1997) griff die Klassifikationsvorschläge auf, unterschied aber nur zwischen fünf inhaltlichen Kategorien: *allgemeines pädagogisches Wissen*; *Fachwissen*; *curriculares Wissen*; *Philosophie des Schulfaches*; *fachspezifisch-pädagogisches Wissen*. Weinert et al. (1990) beschränkten sich dagegen auf vier Aspekte des Wissens, über die Lehrpersonen verfügen sollten: *Klassenführungswissen*, *unterrichtsmethodisches Wissen*, *Fachwissen* sowie *diagnostisches Wissen*.

zur Beschreibung der „Bewältigung beruflicher Anforderungen“ (Baumert & Kunter, 2011a, S. 31) der *weite* Kompetenzbegriff von Weinert verwendet, der neben kognitiven Aspekten zusätzlich noch motivationale, metakognitive und selbstregulative Merkmale professionellen Handelns berücksichtigt:

The theoretical construct of action competence comprehensively combines those intellectual abilities, content-specific knowledge, cognitive skills, domain-specific strategies, routines and subroutines, motivational tendencies, volitional control systems, personal value orientations, and social behaviors into a complex system. Together, this system specifies the prerequisites required to fulfill the demands of a particular professional position.

(Weinert, 2001, S. 51)

Im Anschluss an diese Definition Weinerts wurde im COACTIV-Projekt ein Kompetenzmodell entwickelt (s. Abb. 2), in dem Baumert und Kunter vier *Kompetenzaspekte* unterscheiden, die zugleich die oberste Ebene des Modells bilden: (1) Überzeugungen/Werthaltungen/Ziele, (2) motivationale Orientierungen, (3) Selbstregulation und (4) das Professionswissen.



**Abb. 2:** Das Kompetenzmodell von COACTIV mit Spezifikation für das Professionswissen (aus: Baumert & Kunter, 2011a, S. 32)

Dieser vierte Aspekt, das Professionswissen, wird in fünf weitere Kompetenzbereiche unterteilt: Fachwissen, fachdidaktisches Wissen, pädagogisch-psychologisches Wissen, Organisationswissen und Beratungswissen. Die ersten drei Kompetenzbereiche werden dann in weitere Kompetenzfacetten untergliedert:

- *Fachwissen* wird dabei als tiefes Verständnis der Schulmathematik verstanden.
- *Fachdidaktisches Wissen* wird unterteilt in Erklärungswissen, Wissen über das mathematische Denken von Schülerinnen und Schülern und das Wissen über mathematische Aufgaben.
- *Pädagogisch-psychologisches Wissen* wird aufgefasst als Wissen über Leistungsbeurteilung, Lernprozesse sowie effektive Klassenführung.

An dieser Konzeptualisierung von Wissen orientiert sich auch das DFG-Projekt „Analysekompetenz – Diagnose und Entwicklung“ von Plöger und Scholl (2014a), und zwar in Verbindung mit der Annahme, dass die Ausprägung der Analysekompetenz von Lehrpersonen durch derartige Wissensfacetten bedingt ist. Auch die von mir durchgeführte Qualitative Inhaltsanalyse steht in der Spur dieser Annahme, wobei ich zugleich die von Plöger und Scholl aus forschungspragmatischen Gründen vollzogenen folgenden *Einschränkungen* und *Modifikationen* teile:

- Plöger und Scholl beschränken sich auf drei Facetten des Professionswissens (pädagogisch-psychologisches Wissen, Fachwissen und fachdidaktisches Wissen). Die Begriffe *Organisationswissen* und *Beratungswissen* werden *nicht berücksichtigt*, da sie für die Analyse von Unterricht keine unmittelbare Relevanz besitzen.
- Zudem wird der Begriff des pädagogisch-psychologischen Wissens bei Plöger und Scholl durch die Bezeichnung *didaktisch-methodisches* Wissen ersetzt, um den *handlungsnäheren* Charakter dieses Wissens stärker zu betonen (s. Plöger & Scholl, 2013, 2014a). Denn bei der Analyse von Unterricht stehen Situationen und Handlungen im Fokus, deren Lernwirksamkeit beurteilt werden muss.
- Darüber hinaus nehmen sie keine Differenzierung zwischen fachlichem und fachdidaktischem Wissen vor, sondern subsumieren unter dem Etikett „Fachwissen“ zugleich fachdidaktisches Wissen. Dafür sprechen insbesondere zwei Gründe: Zum einen hat die Diskussion gezeigt, dass es sich um schwer separierbare Wissensaspekte handelt, die mittels Paper-Pencil-Tests und hinreichend differenzierender Aufgabentypen zwar getrennt erhoben werden können, die aber umso eher „verschmolzen“ vorliegen (s. Shulman, 1987), je näher man an das Lehrerhandeln kommt. Denn das *Handeln* von Lehrpersonen ist nicht direkt an den jeweiligen disziplinären Wissensstrukturen orientiert, sondern erfolgt prinzipiell *kontext-* und *situationsgebunden* (s. Bromme, 1992). Zum anderen konnten im Rahmen der oben angesprochenen Forschungsprogramme auf empirischem Wege hohe Korrelationen zwischen dem fachlichen und dem fachdidaktischen Wissen nachgewiesen werden. Beispielsweise liegt bei COACTIV eine Korrelation von .79 für die Gesamtstichprobe und .91 für Gymnasiallehrkräfte vor (Krauss et al., 2011, S. 148f.). In TEDS-M beträgt sie .70 (Blömeke, Kaiser, Döhrmann, & Lehmann, 2010, S. 212), in MT21 wurde eine Korrelation von .81 (Blömeke, Seeber et al., 2008, S. 69) ermittelt. Als Konsequenz derartig hoher Zusammenhänge zwischen Fachwissen und fachdidaktischem Wissen werden in neueren Kompetenzmodellierungen beide Wissensbereiche als eine *Einheit* zusammengefasst (s. für den Bereich der Mathematikdidaktik z.B. Lindmeier, 2011, S. 105).

## 1.2 Definition der Facetten des didaktisch-methodischen Wissens

Didaktisch-methodisches Wissen ist *generisches* fächerübergreifendes Wissen, über das alle Lehrpersonen neben ihrem spezifischen fachlichen und fachdidaktischen Wissen verfügen sollten, um Unterricht professionell planen, durchführen und analysieren zu können. In der Planungsphase dient dieses didaktisch-methodische Wissen dazu, den (möglichen) Unterrichtsverlauf strukturell vorwegzunehmen, sodass er dann in der unterrichtlichen Kommunikation und Interaktion mit den SchülerInnen unter dem Gesichtspunkt der Lernwirksamkeit umgesetzt werden kann. Der Weg von der Planung bis hin zur Durchführung von Unterricht ist also dadurch bestimmt, dass künftiges Handeln *wissensgesteuert* bzw. *wissensbasiert* entworfen wird.

Im Rahmen der nachträglichen Analyse gleicht die Lehrperson ihr eigenes Handeln mit den Lernergebnissen der Schülerinnen und Schüler ab. Während also Handeln in der Planungsphase von Unterricht wissensbasiert *erzeugt* wird, muss dieses Handeln im Rahmen der nachträglichen *Analyse* wissensbasiert *rekonstruiert* werden. In diesem Prozess der Analyse wird dann das didaktisch-methodische (und das fachliche) Wissen zur Beurteilung der Lernwirksamkeit des vorangegangenen Unterrichts genutzt (s. Plöger & Scholl, 2013, 2014b).<sup>10</sup>

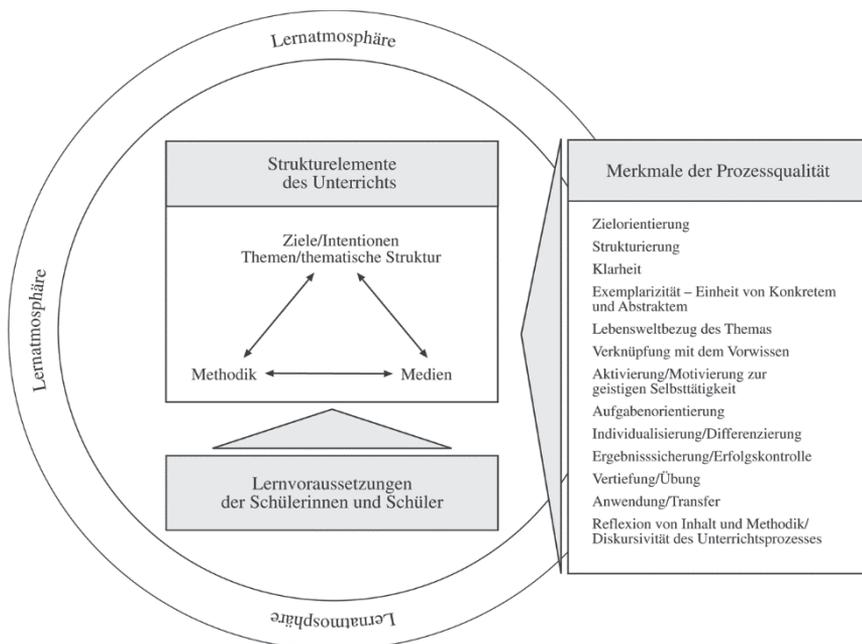
Das für die Planung und Analyse notwendige didaktisch-methodische Wissen wird von Plöger und Scholl im Rahmen eines von ihnen so bezeichneten *Modells der Struktur- und Prozessqualität* von Unterricht konzeptualisiert. In der folgenden Abbildung sind auf der linken Seite die Elemente der *Strukturqualität* und auf der rechten Seite die Merkmale der *Prozessqualität* von Unterricht abgebildet. Gerahmt wird das Modell durch den Faktor der

---

<sup>10</sup> Unter einem lernwirksamen Unterricht wird hier ein Unterricht verstanden, der das sinnstiftende Lernen seitens der SchülerInnen fördert. Dabei stellt das sinnstiftende Lernen einen mentalen Konstruktionsprozess dar, bei dem zwischen (spezifischen) Wissens-elementen neue Verknüpfungen hergestellt werden (s. Reynolds, 1992), wodurch neue zusammenhängende kognitive Strukturen entstehen (s. Collins, Greeno, & Resnick, 2001). In diesem Lernprozess sollten die neu zu generierenden Verbindungen inhaltlich spezifisch gerichtet sein, damit die SchülerInnen die benötigten Wissens-elemente aus ihrem Vorwissen abrufen, die neu aufzunehmenden Informationen damit abgleichen und somit relevante von irrelevanten Informationen trennen können (s. Mayer, 2002; Reynolds, 1992). Die Lernenden sollen also nicht willkürliche Beziehungen stiften, sondern solche, „die strukturell im Hinblick auf die Bildung der neuen kognitiven Struktur notwendig sind“ (Plöger & Scholl, 2015, S. 4). Durch das Erkennen der spezifischen Beziehungen erhalten sie somit Einsicht in den entsprechenden Sachverhalt (s. Köhler, 1969).

Abzugrenzen ist das sinnvolle Lernen vom mechanischen Lernen (rote learning) (s. Ausubel, Novak, & Hanesian, 1980; Mayer, 2002). Wenngleich das mechanische Lernen eine wichtige Rolle im Unterricht spielt, z.B. beim Lernen von Vokabeln oder historischen Fakten, und oft Grundlage für sinnstiftendes Lernen ist, beschränkt es sich auf die Speicherung von Informationen. Im Gegensatz dazu soll aber beim bedeutungsvollen Lernen, durch die Integration von Informationen, neuer Sinn gestiftet werden (s. Fiorella & Mayer, 2015). Sinnvolles Lernen kann beispielsweise durch konkrete Situationen gefördert werden, in denen den Schülerinnen und Schülern Gelegenheit gegeben wird, ihr Vorwissen zu aktivieren und die neuen Informationen damit abzugleichen. Damit das Gelernte dann wiederum auf weitere Situationen übertragen werden kann, muss es aus dem konkreten Kontext gelöst, also dekontextualisiert werden (Bransford, Franks, Vye, & Sherwood, 1989; Brown, Collins, & Duguid, 1989). Gelingender Transfer stellt daher einen wichtigen Indikator für sinnvolles Lernen dar (s. Mayer, 2002).

*Lernatmosphäre.* Darunter werden jene Aspekte gefasst, die ein förderliches und unterstützendes Lernklima charakterisieren, wie z.B. die Trennung von Lern- und Leistungssituationen, ein konstruktiver Umgang mit Schülerbeiträgen und ein freundliches Auftreten der Lehrperson (s. z.B. Helmke, 2012; Jahn, Prenzel, Stürmer, & Seidel, 2011; Klieme, Lipowsky, Rakoczy, & Ratzka, 2006; Meyer, 2013; Mujs & Reynolds, 2011; Seidel & Shavelson, 2007; Seidel, Prenzel, Rimmel, Schwindt et al., 2006).



**Abb. 3:** Merkmale der Struktur- und Prozessqualität von Unterricht (aus: Plöger & Scholl, 2014c, S. 140)

### Strukturqualität

Die linke Seite der Abbildung stellt Unterricht als *Strukturzusammenhang* dar. Mit Blick auf die beiden großen *didaktischen Theorien*, die bildungstheoretische Didaktik (s. z.B. Klafki, 1995, 2007) und die lerntheoretische Didaktik (s. z.B. Heimann, 1976a, 1976b; Heimann, Otto, & Schulz, 1972), kann die Struktur von Unterricht als Zusammenhang von mindestens vier Elementen gekennzeichnet werden, die das formale Gerüst darstellen, „in dessen Rahmen sich inhaltlich die Planung, Durchführung und *Analyse* von Unterricht bewegt“ (Plöger & Scholl, 2012, S. 241): (1) Ziele/Intentionen/Themen/thematische Struktur, (2) Methoden, (3) Medien sowie (4) Lernvoraussetzungen der Schülerinnen und Schüler. In den entsprechenden Übersichtsdarstellungen bzw. Lehrbüchern der Allgemeinen Didaktik gelten diese Elemente als *unbestritten* (s. z.B. Ditton, 2002; Kron, 2008; Peterßen, 2000; Plöger, 2008; Tulodziecki, Blömeke, & Herzig, 2009).

Gesichert wird die Strukturqualität von Unterricht durch eine *sinnvolle* Abstimmung dieser Elemente. Diese Abstimmung nimmt die Lehrperson bereits bei der Planung von Unterricht vor, indem sie bspw. hinreichend klare und stimmige Lernziele formuliert, aber

auch geeignete Sozialformen, Unterrichtsmethoden und Medien auswählt, die das Erreichen der Ziele unterstützen. In diesem Sinne ist Unterrichtsplanung stets *Strukturplanung*, da es sich um einen „*Entscheidungsprozess* im Strukturzusammenhang von Intention, Inhalt, Methode und Medium (bzw. Medien)“ (Ditton, 2002, S. 200) handelt (s. z.B. Heimann, 1976a; Schulz, 1972, 2006).

### Prozessqualität

Diese Strukturelemente bilden zunächst nur einen *statischen* Rahmen. Im späteren Unterrichtsgeschehen muss die in der Planung entworfene Struktur in einen „*prozessualen* Fluss übersetzt werden“ (Plöger & Scholl, 2014b, S. 276), um auf diese Weise konkrete Lerngelegenheiten zu realisieren, die die Lernprozesse der SchülerInnen optimal unterstützen sollen. Zur Beurteilung dieses Prozessgeschehens sind Kriterien notwendig, die in der Unterrichtsforschung als Merkmale von Unterrichtsqualität bezeichnet werden (s. z.B. Clausen, Reusser, & Klieme, 2003; Clausen, Schnabel, & Schröder, 2002; Ditton, 2002; Einsiedler, 2002; Gruehn, 2000; Helmke, 2012; Kunter et al., 2005; Lipowsky, 2006; Seidel, 2003a; Weinert & Helmke, 1996, 1997; Weinert, Schrader, & Helmke, 1989; Weinert et al., 1990).

Diese Prozessmerkmale von Unterrichtsqualität – gelegentlich werden sie auch als *didaktisch-methodische Prinzipien* oder auch als *Prinzipien effektiven Lehrens* bezeichnet<sup>11</sup> – stellen einen wichtigen *Teil* pädagogischen Wissens dar, den ich mit Plöger und Scholl (2014c) als *didaktisch-methodisches* Wissen bezeichne.<sup>12</sup> Bei der Analyse von Unterricht fungieren diese Prozessmerkmale bzw. Prinzipien effektiven Lehrens als *Kriterien* zur Beurteilung der Lernwirksamkeit des abgelaufenen Unterrichts (s. z.B. Jamil, Sabol, Hamre, & Pianta, 2015; Krull, Oras, & Pikksaar, 2010; Panasuk & Sullivan, 1998; Wiens, Hessberg, Lo-Casale-Crouch, & DeCoster, 2013).

---

11 An dieser Stelle sei auf den jeweils unterschiedlichen begrifflich-konzeptuellen Zugriff zwischen „gutem“ und „effektivem“ Unterricht hingewiesen (s. Berliner, 2005; Kunter & Ewald, 2016; Kunter & Trautwein, 2013). Während guter Unterricht (good teaching) eher als Konstruktion des „grünen Tisch“ und als idealisierte Vorstellungen von Unterricht verstanden werden kann, bezieht sich effektiver Unterricht (effektive teaching) auf wirksamen Unterricht, sodass derjenige Unterricht „gut“ ist, für den eine bestimmte (Lern-) Wirksamkeit nachweisbar ist (s. Terhart, 2010, S. 40f.). Berliner (2005) fordert daher eine Vereinigung beider Herangehensweisen zu einem „qualitätvollen Unterricht“. Kunter & Ewald meinen dazu: „Um diesen qualitätvollen Unterricht zu finden, braucht es bestimmte Wertvorstellungen, in die auch theoretische Annahmen über Lernen und die Gestaltung von pädagogischen Settings einfließen. Zur umfassenden Bewertung wird jedoch darüber hinaus die Empirie benötigt, die prüft, ob Unterricht wirklich in der Lage ist, die anvisierten Lernziele zu erreichen“ (Kunter & Ewald, 2016, S. 12).

12 Plöger und Scholl begrenzen das didaktisch-methodische Wissen auf die Prinzipien des effektiven Lehrens, um dadurch „die Handlungen von Lehrpersonen im Unterricht in sehr direkter Weise“ (Plöger & Scholl, 2015, S. 5) beschreiben zu können. Diese Eingrenzung erfüllt zwei Funktionen: Zum einen können die Bestandteile des didaktisch-methodischen Wissens in Handlungskomponenten transformiert werden, die die Lehrperson im Unterricht ein- und umsetzen sollte (s. Plöger & Scholl, 2014b). Unterstellt man nun noch die Erlernbarkeit solcher Handlungskomponenten (s. Baumert & Kunter, 2006), dann sollten sich diese als Handlungskompetenzen einer professionellen Lehrperson ausdrücken, die sie aus ihrem didaktisch-methodischen Wissensfundus heraus generieren kann (s. Plöger & Scholl, 2014b, S. 277). Zum anderen dient das didaktisch-methodische Wissen dazu, das Handeln von Lehrpersonen im Unterricht beschreiben zu können. Im Rahmen der Analysekompetenz müsste eine Lehrperson in der Lage sein, erkennen und beurteilen zu können, ob diese einzelnen Merkmale bzw. Prinzipien effektiven Lehrens umgesetzt wurden und inwieweit diese zu Lernwirkungen im Sinne eines sinnstiftenden Lernens seitens der SchülerInnen beigetragen haben (s. Plöger & Scholl, 2015).

Die *Effektivität* dieser Merkmale ist in einer Vielzahl von Einzelstudien und entsprechenden Metastudien vielfach belegt worden:

- *Zielorientierung* (Bolhuis, 2003; Herweg, 2008; Kyriakides, Christoforou, & Charalambous, 2013; Lipowsky, 2006; Seidel, 2014; Seidel & Shavelson, 2007; Seidel, Blomberg, & Stürmer, 2010),
- *Strukturiertheit* (Brophy, 2000; Good & Brophy, 2000; Helmke, 2007, 2012; Muijs & Reynolds, 2011; Walberg & Paik, 2000; Weinert & Helmke, 1996, 1997),
- *Klarheit* (Helmke, 2007, 2012; Scheerens & Bosker, 1997; Seidel & Shavelson, 2007; Weinert & Helmke, 1996, 1997),
- *Exemplarizität* (Durkin & Rittle-Johnson, 2012; Kalish, Kim, & Young, 2012; Renkl & Atkinson, 2003; Renkl, Atkinson, Maier, & Stanley, 2002; Renkl, Stark, Gruber, & Mandl, 1998),
- *Lebensweltbezug des Themas* (Fey, Gräsel, Puhl, & Parchmann, 2004; Mikelskis-Seifert & Duit, 2010; Muckenfuß, 1995; Parchmann, Ralle, & Demuth, 2000),
- *Verknüpfung mit dem Vorwissen* (Ausubel, Novak, & Hanesian, 1980; Hattie, 2009; Reynolds, 1992; Walberg & Paik, 2000; Wang, Haertel, & Walberg, 1993),
- *Aktivierung zur geistigen Selbsttätigkeit* (Aebli, 1993a, 1993b; Baumert & Kunter, 2013; Kunter & Voss, 2013; Leuders & Holzäpfel, 2011; Lipowsky, 2006; Lipowsky et al., 2009),
- *Aufgabenorientierung* (Jordan et al., 2008; Kleinknecht, Maier, Metz, & Bohl, 2011; Maier, Bohl, Kleinknecht, & Metz, 2013; Neubrand, Jordan, Krauss, Blum, & Löwen, 2011),
- *Diskursivität als Reflexion von Inhalt und Methode* (Cohors-Fresenborg & Kaune, 2003, 2007; Reiss, 2002; Schmidt-Wulffen, 1989),
- *Individualisierung/Differenzierung* (Good & Brophy, 2000; Walberg & Paik, 2000; Waxman, Wang, Anderson, & Walberg, 1985),
- *Vertiefung* (Durcharbeiten) (Aebli, 1993a),
- *Übung und Transfer* (Aebli, 1993a, 1993b; Chi & VanLehn, 2012; Lipowsky, 2006; Renkl, 1998, 2001; Renkl, Hilbert, & Schworm, 2009; Schwartz, Chase, Opezzo, & Chin, 2011),
- *Ergebnissicherung/Erfolgskontrolle* (Klenck & Schneider, 2006; Voss & Kunter, 2011).

Diese im Modell von Plöger und Scholl angeführten Prozessmerkmale zeigen mehr oder weniger große Schnittmengen mit anderen Forschungsprojekten. Im Rahmen des Projektes *COACTIV* wird z.B. von einer „zieladäquaten Orchestrierung“ (Voss & Kunter, 2011, S. 195) von Unterricht unter besonderer Berücksichtigung der individuellen Lernvoraussetzungen und der Rückmeldung gesprochen (ebd., S. 196ff.). Im Projekt *TEDS-M* werden die Aspekte Strukturierung, Methodenvielfalt, Motivation, Differenzierungsmaßnahmen und Leistungsbeurteilung als zentrale Qualitätsmerkmale von Unterricht betont (s. Blömeke & König, 2010). Und in der *TIMS/III-Studie* (*Third [mittlerweile: Trends] in International Mathematics and Science Study*) wurden diese Merkmale unter „verständnisvollem Lernen“ (Baumert & Köller, 2000, S. 273f.) subsumiert und die Merkmale Vorwissen, Lebensweltbezug, multiple Repräsentationen (*Medieneinsatz*) und die Anwendung des Gelernten als zentrale Aspekte gesehen.

Äquivalente Begriffe zu den von Plöger und Scholl genannten Prozessmerkmalen findet man auch im Projekt *SINUS-Transfer* (*Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts*), welches sich selbst als ein Angebot zur Verbesserung der *Lernwirksamkeit* von naturwissenschaftlichem Unterricht versteht und im Sinne dieser Absicht auch tatsächlich „wirkt“ (s. z.B. Prenzel, Carstensen, Senkbeil, Ostermeier, & Seidel, 2005). Das Projekt besteht aus entsprechenden Modulen wie zielorientiertes und verständnisvolles Experimentieren, Aufgabenkultur, Kompetenzzuwachs sichtbar machen und rückmelden, Sicherung von Basiswissen, kumulatives Lernen, die Lernmotivation stärken usw.

Diese hier aufgelisteten Prozess- bzw. Qualitätsmerkmale von Unterricht werden im Teilkapitel 3.2 – von kleineren Modifikationen abgesehen – letztlich als Kategorien für die Durchführung meiner Qualitativen Inhaltsanalyse genutzt, um das didaktisch-methodische Wissen der ProbandInnen zu erfassen. Dort werden die Kategorien dann auch beschrieben, definiert und mit Ankerbeispielen versehen.

### 1.3 Definition der Facetten des fachlichen (incl. fachdidaktischen) Wissens

Das beschriebene didaktisch-methodische Wissen stellt nur einen Teil der für die Analyse von Unterricht notwendigen Wissensbasis dar. In Ergänzung dazu müssen Lehrpersonen für eine angemessene Analyse von Unterricht auch über entsprechendes fachliches (incl. fachdidaktisches) Wissen (s. Plöger & Scholl, 2014a, 2014b) verfügen. Was darunter zu verstehen ist, erläutere ich in diesem Abschnitt.

In der einschlägigen erziehungswissenschaftlichen wie fachdidaktischen Forschungsliteratur besteht ein breiter Konsens darüber, dass LehrerInnen über ein tiefes Verständnis der von ihnen zu unterrichtenden Fächer verfügen sollten (s. z.B. Ball & Bass, 2000; Baumert & Kunter, 2006, 2011b; Blömeke et al., 2010a, 2010b; Borowski, Neuhaus et al., 2010; Brophy, 1991; Daehler & Shinohara, 2001; Grossmann & Schoenfeld, 2005; Kind, 2009; Lipowsky, 2006; Krauss et al., 2008; Oser & Renolds, 2005; Shulman, 1986, 1987; Terhart, 2002, 2016; van Driel & Berry, 2012). Sobald aber die Frage gestellt wird, welche konkreten Wissensfacetten dazu zu zählen sind, scheint man von einem entsprechenden Konsens mehr oder weniger weit entfernt zu sein. Das gilt nicht nur für das fachliche, sondern auch für das fachdidaktische Wissen, was ich im Folgenden beispielhaft für die Unterrichtsfächer Mathematik und Physik (bzw. für den naturwissenschaftlichen Lernbereich) deutlich machen möchte.

Im Bereich der Mathematikdidaktik werden in den entsprechenden Diskussionen über das *Fachwissen* verschiedene Aspekte betont, wie etwa das Verständnis des Faches, die Berücksichtigung bestimmter Bereiche des Fachwissens oder die Unterscheidung zu beherrschender Anspruchsniveaus der Fachinhalte (s. z.B. Baumert et al., 2010). Hinsichtlich solcher Aspekte scheint ein gewisser Konsens vorzuliegen, wenn man etwa auf die entsprechenden Konzeptualisierungen in Forschungsprogrammen wie *COACTIV* und *TEDS-M* schaut. In beiden Projekten werden fachliche Kompetenzen in den inhaltlichen Bereichen *Arithmetik*, *Algebra*, *Geometrie* und *Stochastik* gemessen. Dabei orientiert man sich

an einem Anspruchsniveau, das von der Sekundarstufe I und II bis hin zur elementaren Universitätsmathematik reicht (s. Krauss et al., 2011, S. 142; Krauss et al., 2008; Blömeke et al., 2009; Döhrmann, Kaiser, & Blömeke, 2010, S. 172).

Im Vergleich dazu scheint es weitaus schwieriger zu sein, ein gewisses Maß an Übereinstimmung hinsichtlich der Modellierung des fachdidaktischen Wissens zu erzielen. Hier herrscht eher eine gewisse Skepsis vor, wenn Krauss et al. bemerken, dass dort „eine Konzeptualisierung des fachdidaktischen Wissens ... nicht auf eindeutige Weise möglich“ (Krauss et al., 2011, S. 137) sei. Diese Situation spiegelt sich in den entsprechenden Konzeptualisierungen in COACTIV und TEDS-M, da in beiden Projekten unterschiedliche Wissensfacetten postuliert werden:

- In COACTIV sind es drei Facetten: *Erklärungswissen* (Verständlichmachen von mathematischen Inhalten), *Wissen über das mathematische Denken von Schülerinnen und Schülern* (mathematikbezogene Schülerkognition, insb. Schülerfehler und Schüler-schwierigkeiten beim Lernen) und *Wissen über mathematische Aufgaben* (kognitives Potential von Mathematikaufgaben) (s. Baumert & Kunter, 2011a, S. 32, 37f.; Krauss et al., 2008, S. 234).
- In TEDS-M werden zwei „Subdimensionen“ als Facetten definiert: *curriculares und planungsbezogenes Wissen* einerseits und *interaktionsbezogenes Wissen* andererseits (s. Döhrmann et al., 2010, S. 175).

Die Diskussion im Bereich der Didaktik der Naturwissenschaften zeigt gewisse Parallelen zu der im Bereich der Mathematikdidaktik. Auch dort wird gefordert, dass Lehrpersonen die weite Spanne zwischen Schulwissen und universitärem Fachwissen beherrschen sollten (s. z.B. Borowski et al., 2010; Borowski, Neuhaus et al., 2010; Riese, 2009; Riese & Reinhold, 2010, 2012; Tepner et al., 2012; van Driel & Verloop, 1999). Die aus dieser Perspektive heraus zu entwickelnden Konsequenzen für die Ausbildung von Lehrpersonen werden aber uneinheitlich diskutiert. So werden z.B. unterschiedliche Positionen hinsichtlich der Frage bezogen, ob dieses Wissen im Studium ausschließlich unter fachsystematischen Gesichtspunkten erworben oder ob es von Anfang an auch schon „pädagogisch überformt“ (Fischler, 2008, S. 30) sein sollte.

Noch deutlicher gehen die Meinungen über das fachdidaktische Wissen auseinander. Insgesamt entsteht der Eindruck, dass „für fachdidaktisches Wissen in den MINT-Fächern bislang weder national noch international eine einheitliche, umfassende und auf ein Rahmenmodell abgestimmte Modellierung“ (Kulgemeyer et al., 2012, S. 2; s. auch Gramzow, Riese, & Reinhold, 2013, S. 8; Riese et al., 2015, S. 57; Borowski & Riese, 2010) vorliegt: „Uneinigkeit besteht nicht nur in der Frage, welche unterrichtsrelevanten Aspekte zum fachdidaktischen Wissen gezählt werden, sondern auch darin, wie sie interpretiert werden“ (Tepner et al., 2012, S. 14; s. dazu auch Gramzow et al., 2013; van Dijk & Kattmann, 2010). Darüber hinaus wird das grundsätzliche Problem der *Eigenständigkeit* des fachdidaktischen Wissens kontrovers diskutiert (s. dazu Abell, 2007, 2008; Park, Chen, & Jung, 2011; van Driel, Verloop, & de Vos, 1998).

Die hier nur in knappen Zügen umrissene Situation hinsichtlich der ungeklärten Frage der fachlichen und fachdidaktischen Kompetenzen (s. ausführlich dazu z.B. Bernholt, 2010; Fischer, Borowski, & Tepner, 2012; Fischer, Borowski, Kauertz, & Neumann, 2010; Park & Chen, 2012; Park & Oliver, 2008) hat in den letzten Jahren zur Entstehung

einer Vielzahl von Arbeiten zur Kompetenzmodellierung und -messung im Bereich der Physikdidaktik geführt (s. z.B. den Überblick in Riese et al., 2015).

Stellvertretend für diese Anstrengungen seien hier zwei Forschungsprogramme exemplarisch angeführt: das Forschungsprogramm *ProwiN* (*Professionswissen von Lehrkräften in den Naturwissenschaften*; Borowski, Neuhaus et al., 2010) und das Programm *ProfiLe-P* (*Professionswissen in der Lehramtsausbildung Physik*, Kulgemeyer et al., 2012; Gramzow, 2015; Riese et al., 2015).

Das übergeordnete Ziel des Forschungsprogramms *ProwiN* (*Professionswissen von Lehrkräften in den Naturwissenschaften*) besteht darin, Wirkungen des Professionswissens von Lehrkräften auf Schülerleistungen zu überprüfen. Dazu möchte die Projektgruppe „das Professionswissen von Lehrkräften in den Fächern Biologie, Chemie und Physik“ (Borowski, Neuhaus et al., 2010, S. 344) untersuchen, damit entsprechende Zusammenhänge zwischen dem Fachwissen, dem fachdidaktischen und dem pädagogischen Wissen beschrieben werden können, um dann darauf aufbauend den „Einfluss dieser drei Dimensionen des Professionswissens auf für Unterrichtsqualität relevante Variablen“ (ebd.), wie z.B. Zielklarheit, kognitive Aktivierung und Lernzuwachs, nachweisen zu können.

Um dem Problem der bisher uneinheitlichen Konzeptualisierungen hinsichtlich der fachlichen Kompetenzen zu begegnen, setzt man in *ProwiN* auf einen *Minimalkonsens*, um „erstmalig aus einer den Naturwissenschaften Biologie, Chemie und Physik gemeinsamen Perspektive“ (Tepner et al., 2012, S.18) ein dreidimensionales Modell des Professionswissens zu entwickeln, wobei der Begriff der *Dimension* ein Äquivalent zum Begriff des *Kompetenzbereiches* im COACTIV-Modell darstellt (s. S. 17):

- Die Dimension *Fachwissen* steht für verschiedene inhaltliche Themenbereiche (im Fach Physik z.B. für die Bereiche Mechanik, Elektrizitätslehre und Optik); *formal* wird zusätzlich zwischen deklarativem, prozeduralem und konditionalem Wissen unterschieden.
- Die Dimension des *fachdidaktischen* Wissens wird im Rückgriff auf Shulman in zwei Facetten unterteilt: Einerseits in das *Wissen über bestimmte fachspezifische Schülerfehler und Schülervorstellungen* (bspw. Präkonzepte oder Misskonzepte [misconcepts], die Schülerinnen und Schüler mit in den Unterricht bringen) und andererseits in *fachspezifische Instruktions- und Vermittlungsstrategien* (die exponierte Stellung sowohl des *Experiments* im Rahmen naturwissenschaftlicher Erkenntnisgenerierung als auch von *Modellen* im naturwissenschaftlichen Unterricht zur Darstellung und Wissensvermittlung).
- Das *pädagogische* Wissen umfasst die Facetten Klassenführung, Unterrichtsmethoden, individuelle Lernprozesse und Leistungsbeurteilung (s. Tepner et al., 2012, S. 17ff.).

Aus diesen drei Dimensionen heraus werden dann Items entwickelt, mit denen das Professionswissen von Lehrpersonen im Bereich Naturwissenschaften erhoben werden soll. Die teilweise noch ausstehenden Ergebnisse dieser Testung werden in Zukunft nähere Aufschlüsse über die Tragfähigkeit der festgesetzten Kompetenzfacetten geben (s. z.B. Kirschner, 2013). In ersten Untersuchungen „konnte man die Unterscheidung der Dimensionen Fachwissen und fachdidaktisches Wissen in einer speziellen Modellierung in der Domäne Physik an einer Stichprobe Studierender empirisch nachweisen“ (Kulgemeyer et al., 2012, S. 5; s. auch Riese & Reinhold, 2010).

In dem *physikspezifischen* Forschungsvorhaben *ProfiLe-P (Professionswissen in der Lehramtsausbildung Physik)* haben sich die beteiligten ForscherInnen das Ziel gesetzt, „ein Rahmenmodell des Professionswissens angehender Physiklehrerinnen und -lehrer sowie von Übungsleiterinnen und -leitern der Physik zu entwickeln“ (Kulgemeyer et al., 2012, S. 1). Ein besonderer Fokus liegt dabei auf der „Modellierung domänenspezifischer und generischer Wissensbereiche, die Lehramtsstudierende der Physik im Hochschulstudium als Teil ihrer professionellen Kompetenz erwerben sollten“ (Riese et al., 2015, S. 56). Für die Realisierung dieser Intention sind drei Teilvorhaben angesetzt, die gleichzeitig eine erste Strukturierungsebene im Rahmenmodell des Professionswissens angehender Physiklehrkräfte darstellen (s. ebd., S. 59):

- 1) DaWis (Fachdidaktisches Wissen von Lehramtsstudierenden und Übungsleitern im Fach Physik): Innerhalb dieses Teilvorhabens „soll ein möglichst umfassendes Modell physikdidaktischer Kompetenz entwickelt werden“ (Kulgemeyer et al., 2012, S. 3), um deklaratives und analytisches fachdidaktisches Wissen modellieren und diagnostizieren zu können.
- 2) EWis (Erklärungswissen als prozedurales fachdidaktisches Wissen): Anknüpfend an das erste Teilvorhaben sollen mit EWis „kompetenzdiagnostische Methoden für die Fähigkeit des Verstehbarmachens physikalischer Sachverhalte bei Physikstudierenden entwickelt werden, um eine diesbezügliche Komponente des professionellen Wissens angehender Physiklehrkräfte zu modellieren“ (ebd.). Eine zentrale Stellung in diesem Vorhaben nimmt die „Standardsituation“ des Erklärens im Unterricht ein, wobei man hinsichtlich dieser Facette prozeduralen fachdidaktischen Wissens davon ausgeht, dass es nicht ausreichend sei, lediglich hohes Fachwissen zu besitzen; man müsse auch in der Lage sein, den entsprechenden Sachverhalt adressatengemäß und sachgerecht erklären zu können.
- 3) FaWis (Fachwissen im Studium zum Lehramt der Physik): Grundlegendes Ziel des dritten Teilvorhabens ist es, „ein Modell für universitäres fachliches Professionswissen und ein Testinstrument zur modellkonformen Erfassung des Fachwissens im Fach Physik“ (ebd., S. 5) zu entwickeln, um dadurch die entsprechend festgelegten Ziele für die Studiengänge Gymnasium/Gesamtschule und Haupt-/Real-/Gesamtschule curricular valide beschreiben zu können.

Diese drei Teilvorhaben bilden gleichzeitig das dem Projekt zugrunde liegende Modell des Professionswissens angehender Lehrpersonen im Fach Physik ab. Die Autorengruppe teilt das Gesamtmodell (Professionswissen) in drei einzelne Modelle auf, die jeweils noch einmal ausdifferenziert werden:

- Das Modell des fachdidaktischen Wissens umfasst vier Facetten: Instruktionsstrategien, Schülervorstellungen, Experimente und Vermittlung eines angemessenen Wissenschaftsverständnisses und fachdidaktische Konzepte.
- Das Modell Erklärungswissen wird in zwei Anforderungsbereiche gegliedert: Adressatengerechtheit und Sachgemäßheit einer Erklärung. Es wird angenommen, „dass eine Erklärung ein sprachlicher Akt ist und Erklären den Prozess der adressatengemäßen und sachgerechten Erstellung und Modifikation von Erklärungen bedeutet“ (Riese et al., 2015, S. 61).
- Das Modell Fachwissen wird hinsichtlich dreier Fachstufen unterteilt: Schulwissen, vertieftes Schulwissen und universitäres Wissen.

Die hier skizzenhaft dargestellte Forschungs- und Diskussionslage hat gewisse Schnittmengen in der Diskussion um fachliches und fachdidaktisches Wissen offenbart, aber auch exemplarisch die Schwierigkeiten deutlich gemacht, die im Rahmen der ausgewählten fachdidaktischen Bereiche (Mathematik, Physik) hinsichtlich der Begründung fachlicher und fachdidaktischer Wissensfacetten bestehen.

Für die vorliegende Arbeit sind diese Überlegungen nützlich, um einen gewissen *Minimalkonsens* hinsichtlich fachlichen und fachdidaktischen Wissens in den entsprechenden Projekten offenzulegen. Diesen Minimalkonsens werde ich nutzen, um in Teilkapitel 3.3 die entsprechenden Kategorien für die Qualitative Inhaltsanalyse herzuleiten. Da ich dort dezidiert auf die Herleitung der Wissensfacetten eingehe und sie durch entsprechende Kategorien (= Kompetenzfacetten) operationalisiere, beschränke ich mich an dieser Stelle auf die bloße Nennung dieser Facetten für fachliches (incl. fachdidaktisches) Wissen:

- Physikalische Begriffsbildung
- Durcharbeiten von Fachbegriffen
- Experimentieren als zentrale Fachmethode
- (weitere) fachliche Arbeitsweisen: Phänomene beobachten und genau analysieren; Messdaten grafisch darstellen; kritisches Beurteilen von Messwerten; qualitative Zusammenhänge (ansatzweise) mathematisieren
- Historische Genese physikalischen Wissens
- Produktiver Umgang mit Fehlern
- Lernen als konstruktive Tätigkeit von SchülerInnen

## 1.4 Das Paradigma der Expertiseforschung – Komplexität der Informationsverarbeitung

Die bisherigen Ausführungen haben gezeigt, dass für eine angemessene Analyse von Unterricht didaktisch-methodisches und fachliches (incl. fachdidaktisches) Wissen notwendig ist. Nach Plöger und Scholl (2014a) spielt darüber hinaus aber auch die Fähigkeit eine Rolle, das mehr oder weniger komplexe Zusammenspiel von Situationen und Handlungen erkennen und beurteilen zu können. Zur Begründung dieses Aspekts greifen Plöger und Scholl (2014a) auf die Ergebnisse der Expertiseforschung zurück, die gezeigt hat, dass ExpertInnen in der Lage sind, nicht nur einzelne relevante Situationen und Handlungen zu erkennen, sondern diese auch zu größeren didaktischen Einheiten zu verbinden und darauf aufbauend schließlich den gesamten Unterrichtsprozess in den Blick zu nehmen. Das gelingt ExpertInnen offensichtlich dadurch, dass sie die von ihnen bei der Beobachtung von Unterricht aufgenommenen Informationen in unterschiedlichen Komplexitätsgraden verarbeiten können.

Diese Ergebnisse der Expertiseforschung werde ich im Folgenden in gebotener Kürze referieren, um daran anschließend ein von Plöger und Scholl (2014a) entwickeltes fünfstufiges Modell vorzustellen, mit dem unterschiedliche Grade der Komplexität der Informationsverarbeitung im Rahmen der Analyse von Unterricht klassifiziert werden können. In Anlehnung an dieses Stufenmodell sind dann zur Durchführung meiner Qualitativen Inhaltsanalyse (in Ergänzung zu den erforderlichen Kategorien für didaktisch-methodisches und fachliches Wissen) auch entsprechende Kategorien zur Erfassung unterschiedlicher Grade der Komplexität der Informationsverarbeitung operationalisiert worden.

### 1.4.1 Ergebnisse der Expertiseforschung

In der Expertiseforschung existieren schon seit längerem Beschreibungen des ExpertInnenwissens unter formalen Gesichtspunkten.<sup>13</sup> Die dort stattfindende Diskussion kann hier nur kursorisch wiedergegeben werden und fokussiert vor allem auf jene Aspekte, die für die Frage der Analysekompetenz relevant sind. Die Ergebnisse zeigen, dass sich ExpertInnen von NovizInnen vor allem hinsichtlich ihrer *kognitiven* Tätigkeiten und des *Komplexitätsgrades* zu verarbeitender Informationen unterscheiden (s. z.B. Berliner, 1987, 1992, 2004; Bromme, 2001; Carter, Cushing, Sabers, Stein, & Berliner, 1988; Carter, Sabers, Cushing, Pinnegar, & Berliner, 1987; Gruber, 1994; Krauss, 2011; Palmer, Stough, Burdinski, & Conzales, 2005).

---

13 In der Fachliteratur wird Wissen in formaler Hinsicht in deklaratives und prozedurales Wissen unterschieden. Mit deklarativem Wissen wird meistens konzeptuelles (begriffliches) Wissen bezeichnet. Unter prozeduralem Wissen wird (automatisiertes) Handlungswissen verstanden. Damit werden (teilweise unbewusste) Fertigkeiten wie auch kognitive Fähigkeiten beschrieben, die man benötigt, um entsprechende Handlungen ausführen zu können. Diese Unterscheidung ist für die folgenden Überlegungen keineswegs unwichtig, steht aber eher im Hintergrund, denn mit Gruber muss zunächst einmal konstatiert werden: „Die bloße Unterscheidung zwischen deklarativem und prozeduralem Wissen wird dem Stand der Forschung nicht gerecht, in der längst viel feinere Differenzierungen vorliegen“ (Gruber, 2008, S. 101).

- 1) Die unterschiedlichen kognitiven Tätigkeiten zwischen ExpertInnen und NovizInnen zeigen sich in der Wahrnehmung und Beurteilung von (relevanten) Situationen und Handlungen. NovizInnen sind allenfalls ansatzweise in der Lage, „to selectively attend to the most relevant events ... because they lack sufficient knowledge about the teaching profession and specific subject areas to identify these events“ (Gaudin & Chaliès, 2015, S. 46), weshalb sie meistens einen „inaccurate sense of the effectiveness of the lesson“ (Barnhart & van Es, 2015, S. 85) besitzen.<sup>14</sup> Im Gegensatz dazu können ExpertInnen relevante Situationen identifizieren, die einzelnen Situationen miteinander verknüpfen und so zu größeren Sinneinheiten verdichten. Sie „reconstruct and anticipate the context of the instruction“ (König et al., 2014, S. 78) und können dadurch zwischen zielführenden, d.h. lernwirksamen und weniger lernwirksamen Situationen und Handlungen unterscheiden. Um eine solche Unterscheidung treffen zu können, bedarf es einer *gezielten* Suche nach entsprechenden Informationen und einer *Bewertung* dieser Informationen im Zusammenhang mit der Formulierung von möglichen *Handlungsoptionen*, also alternativen, effektiveren Lernhandlungen (s. Barnhart & van Es, 2015, S. 83; Kersting, 2008, S. 847; König et al., 2014, S. 78; Plöger et al., 2015, S. 167; Sabers, Cushing, & Berliner, 1991).<sup>15</sup>
- 2) Des Weiteren gibt es zwischen ExpertInnen und NovizInnen Unterschiede hinsichtlich des Komplexitätsgrades, mit dem Informationen verarbeitet werden. Bei der Analyse von Unterricht gehen NovizInnen „step by step“ (König et al., 2014, S. 78), d.h. chronologisch (sequentiell) und additiv vor, sodass eingehende Informationen nacheinander abgearbeitet werden. Dabei neigen sie dazu, „to focus on superficial matters such as teacher and student characteristics, fleeting classroom management issues and global judgements of lesson effectiveness“ (Castro, Clarke, Jacobs, & Givvin, 2005, S. 11). Dadurch, dass sich die Wahrnehmung von NovizInnen meistens auf einzelne Situationen beschränkt, sind sie auch kaum in der Lage, den Stellenwert von einzelnen Situationen und/oder Handlungen vor dem Hintergrund des gesamten Unterrichtsverlaufes zu beurteilen. Diese Leistung erbringen aber ExpertInnen: Sie können den tieferliegenden didaktisch-methodischen Sinn (Tiefenstrukturen) des unterrichtlichen Handelns interpretativ erschließen. Das gelingt ihnen, weil sie, im Vergleich zu NovizInnen, eine Vielzahl kohärenter Beziehungstiftungen in Form von Schemata besitzen, die zu semantischen Netzwerken bzw. mentalen Modellen verknüpft sind (Baumert & Kunter, 2006, S. 483; de Jong & Ferguson-Hessler, 1996, S. 111; Gruber & Gallenberger, 2006; S. 26; Shuell, 1990, S. 542; Strahan, 1989, S. 54; Elio & Scharf, 1990). Dadurch sind ExpertInnen in der Lage, eingehende Informationen differenzierter mit dem verfügbaren Wissensnetz zu enkodieren (Berliner 1992, S. 243; Livingston

---

14 Im deutschsprachigen Raum wird für die Wahrnehmung sichtbarer Situationen und Handlungen der Begriff der *Sichtstrukturen* verwendet (Aebli, 1971; Klieme, 2006; Kunter & Trautwein, 2013; Oser & Baeriswyl, 2002; Oser & Patry, 1994; Reyer, 2004; Seidel, 2003a, 2006b).

15 Bei dieser Suche zeichnen sich ExpertInnen durch ihre hohe Problemlösekompetenz, ihre Entscheidungsstärke und ihr Urteilsvermögen aus (s. z.B. Gruber, Harteis, & Rehrl, 2006, S. 195; Hattie, 2003, S. 6; Shavelson, 1973; Shavelson & Stern, 1981). Damit diese Fähigkeiten hinreichend ausgeprägt sind, bedarf es umfangreicher Erfahrungen und Übung, denn nur so können auf lange Zeit gesehen entsprechende *prozedurale Routinen* ausgebildet werden (Bromme, 1985, S. 184; Bromme & Haag, 2008, S. 807). Diese Routinen sind kognitiv und emotional für das Arbeitsgedächtnis entlastend (Chi, 2006, S. 24; Leinhardt & Greeno, 1986, S. 95).

& Borko, 1990) und tiefer zu verarbeiten, wie dies auch bei der Analyse von Unterricht der Fall ist.<sup>16</sup>

Diese beiden zentralen Erkenntnisse der Expertiseforschung im Hinblick auf die Analyse von Unterricht werden auch in gegenwärtigen Modellierungen zur Analysekompetenz berücksichtigt, wie ich anhand von drei Untersuchungen zeigen möchte.

In den Untersuchungen von van Es und Sherin wird zur Unterscheidung der Analysefähigkeit von Personen ein Dreischritt, bestehend aus den kognitiven Tätigkeiten *noticing*, *interpreting* und *informing decision*, beschrieben (van Es & Sherin, 2002, 2008).<sup>17</sup> *Noticing* meint, Situationen und Handlungen wahrzunehmen und zu beschreiben. Unter *interpreting* verstehen die Autorinnen, dass Personen in der Lage sind, einzelne Situationen miteinander in Beziehung zu setzen und deren Lernwirksamkeit zu bewerten. *Informing decision* heißt, dass vor dem Hintergrund der festgestellten Lernwirksamkeit alternative Handlungsoptionen (pedagogical solutions) entwickelt werden (s. van Es & Sherin, 2002, S. 582).

Die Forschungsgruppe um Seidel (s. z.B. Seidel & Stürmer, 2014) greift auf die Dreiteilung von van Es & Sherin zurück und entwickelte das Konzept der „professionellen Unterrichtswahrnehmung“.<sup>18</sup> Dabei unterscheiden sie drei kognitive Tätigkeiten voneinander, mit denen unterschiedliche Dimensionen professioneller Unterrichtswahrnehmung bezeichnet werden: Beschreibung, Erklärung und Vorhersage (s. z.B. Seidel & Stürmer, 2014). Auf der untersten Stufe (Beschreiben) müssen „lehr-lernrelevante beobachtete Unterrichtssituationen“ (Jahn, Stürmer, Seidel, & Prenzel, 2014, S. 172) *beschrieben* werden. Auf der nächsthöheren Stufe (Erklären) müssen diese Beschreibungen „auf Basis wissenschaftlicher Theorien und Befunde vor dem Hintergrund von Lernwirksamkeit“ (ebd.) *erklärt* werden. Diese Erklärungen sollten dann auf der höchsten Stufe (Vorhersage) genutzt

---

16 Dabei kann sich der Informationsgehalt von semantischen Netzen hinsichtlich Umfang und Abstraktionsgrad deutlich unterscheiden, denn semantische Netze sind ein komplexes Gebilde von Teil-Ganzes-Strukturen und Mikro- und Makropropositionen, die aus einzelnen Schemata bestehen, die selbst wieder in umfassende Netze integriert sind. In der Literatur wird dann von *tief verstandenem* Wissen gesprochen, wenn die unterschiedlich komplexen Strukturen einer Person deutliche Verbindungen aufweisen und diese idealiter in Form von übergeordneten *Prinzipien* (Peterson & Comeaux, 1987, S. 327) oder in Form von *Abstraktionshierarchien* gespeichert sind (Berliner, 1987, S. 298; Frey, 2004; Leinhardt & Smith, 1985). Durch solche Abstraktionshierarchien besitzen ExpertInnen eine *Beweglichkeit* im Denken, die sie befähigt, über den Prozess des *Chunking*, mit einer vergleichsweise kleinen Anzahl hoch abstrakter Schemata Realität mehrperspektivisch wahrzunehmen (Borko & Livingston, 1989, S. 490). Diese Beweglichkeit offenbart sich aber nicht nur in inhaltlicher Hinsicht (Perspektivenwechsel), sondern äußert sich auch in der Fähigkeit, zwischen verschiedenen Ebenen (abstrakt vs. konkret, strukturell vs. funktional oder oberflächen- vs. prinzipienorientiert) je nach situativer Anforderung strategisch wechseln zu können (s. Krems, 1996, S. 83).

17 Dieser Einteilung von van Es und Sherin wird auch in anderen Studien gefolgt. Beispielsweise greifen König et al. (2014) auf diese Einteilung zurück, differenzieren aber nur zwischen zwei unterschiedlichen kognitiven Tätigkeiten, mit denen Informationen bei der Analyse von Unterricht verarbeitet werden: *noticing* und *interpreting*. Sie versuchen dann, Korrelationen zwischen dem pädagogischen Wissen und den kognitiven Tätigkeiten nachzuweisen.

18 Mit dem Konzept der „professionellen Unterrichtswahrnehmung“ wird beschrieben, „wie Lehrpersonen komplexe Unterrichtssituationen vor dem Hintergrund ihres Wissens um Lehren und Lernen beobachten und interpretieren“ (Jahn et al., 2014, S. 172). Zur Messung der professionellen Unterrichtswahrnehmung wird das Tool Observer (Seidel & Stürmer, 2014; Seidel, Blomberg, & Stürmer, 2010) eingesetzt, das ein inhaltlich valides, videobasiertes Instrument zur Erfassung der professionellen Wahrnehmung von Unterricht darstellt (s. z.B. Blomberg, Stürmer, & Seidel, 2011; Jahn et al., 2014; Seidel et al., 2010; Seidel et al., 2011; Stürmer & Seidel, 2015).

werden, um „Auswirkung auf weitere Lernprozesse von Schülerinnen und Schülern“ (ebd.) *vorhersagen* zu können.

Kersting (2008) und Kersting et al. (2012) unterscheiden drei Grade (Level) der Komplexität der Informationsverarbeitung, in denen die Tiefe der Interpretation zum Ausdruck kommt: Level 1 bezieht sich auf die (reine) Beschreibung von beobachteten Unterrichtsergebnissen. Level 2 umfasst Aussagen, die zwar analytische Schlussfolgerungen enthalten, aber nicht miteinander in Beziehung gesetzt werden, sodass sie (noch) kein zusammenhängendes Urteil bilden können („but without connecting analytical points to form a coherent argument“, Kersting, 2008, S. 849). Personen, die Level 3 erreicht haben, geben Antworten, die eine sehr umfassende, kohärente und integrierte Interpretation (hoher Komplexitätsgrad der Informationsverarbeitung) darstellen und sich durch das Erkennen von Ursache-Wirkungsbeziehungen (kognitive Tätigkeit) auszeichnen (s. Kersting, 2008; Kersting et al., 2012).

### 1.4.2 Fünfstufiges Kompetenzmodell nach Plöger und Scholl

Die herangezogenen Studien belegen, dass bei der Analyse von Unterricht Informationen in unterschiedlichem Komplexitätsgrad zu verarbeiten sind und dabei noticing, interpreting und decision making wichtige Facetten von Analysekompetenz darstellen.

Diese beiden Aspekte (die bei der Analyse von Unterricht zu unterscheidenden kognitiven Tätigkeiten des Wahrnehmens, Evaluierens und des Suchens nach Alternativen einerseits und die Komplexität der Informationsverarbeitung andererseits) berücksichtigen auch Plöger und Scholl in ihrem Modell der Komplexität der Informationsverarbeitung. Sie sehen die Analyse von Unterricht als einen Prozess der Informationsverarbeitung und konkretisieren diesen Prozess durch fünf Stufen.<sup>19</sup>

---

19 Bei der Modellierung ihres Modells greifen Plöger und Scholl (2014a) auf das *Modell of hierarchical complexity* von Commons et al. zurück. Dieses Modell von Commons et al. (s. z.B. Commons et al., 1982; Commons & Richards, 1984a, 1984b; Commons & White, 2003; Commons, Goodheart, Dawson, & Draney, 2008; Commons, Trudeau, Stein, Richards, & Krause, 1998) ist *inhaltsunabhängig*, also *formal* gehalten, und erhebt somit den Anspruch, auf sämtliche Inhaltsbereiche angewendet werden zu können, in denen *Informationen in organisierter Form verarbeitet* werden, wie dies auch auf die Analyse von Unterricht zutrifft. Organisierte Informationsverarbeitung ist immer dann erforderlich, „wenn mit zunehmender Anzahl von Inhaltselementen die Schwierigkeit ihrer Verarbeitung steigt“ (Plöger et al., 2015, S. 169). Dies ist dann der Fall, wenn die Anzahl der zu verbindenden Elemente größer wird und dadurch der Komplexitätsgrad der zu lösenden Aufgabe steigt. Um die *Art* der dazu erforderlichen Komplexität der Informationsverarbeitung besser beschreiben zu können, unterscheiden Commons et al. (s. z.B. Commons et al., 2008) zwischen *horizontaler* und *vertikaler* Komplexität.

Die *horizontale* Komplexität ist dadurch bestimmt, dass es lediglich zu einer willkürlichen Verkettung (chains) von Elementen kommt, sodass die zu verarbeitenden Informationen mehr oder weniger unverbunden bleiben. Im Gegensatz dazu ist die *vertikale* Komplexität nicht nur durch die steigende Anzahl von Elementen gekennzeichnet, sondern dadurch, dass zwischen diesen Elementen spezifische Verbindungen hergestellt werden. Diese Verbindungen sind keine nichtwillkürlichen (nonarbitrary) Verkettungen, sondern stellen eine *Koordination* dar (Commons, 2007), sodass weniger komplexe Einheiten in höheren organisiert werden und dadurch unterschiedliche Komplexitätsgrade der Informationsverarbeitung durch eine geordnete Abfolge von Stufen modelliert werden können. Die einzelnen Stufen bauen aufeinander auf, sodass sich höhere Stufen aus der Koordination von Elementen der darunterliegenden ergeben und über die Stufen hinweg die Komplexität in *quantitativer* wie *qualitativer* Hinsicht zunimmt. Das bedeutet, dass komplexere Aufgaben (= höhere Stufen des Modells) Leistungen voraussetzen, die für eine Stufe geringerer Komplexität kennzeichnend sind, sodass sich komplexere Aufgaben nur durch *Rekursion* (Commons et al., 2008, S. 183) auf weniger komplexe Subaufgaben (= niedrigere Stufen des Modells) lösen lassen.

Das Modell ist in *drei* Hauptstufen (römische Bezifferung) und *fünf* Unterstufen (arabische Bezifferung) unterteilt:

### Hauptstufe I: Analytische Kompetenz

Auf Stufe 1 und 2 sind die Analyseleistungen begrenzt auf die Erfassung *einzelner* Unterrichtsereignisse, die als sichtbare Handlungen und/oder Situationen wahrgenommen werden, sodass mit diesen beiden Stufen die *Sichtstruktur* (Aebli, 1971; Klieme, 2006; Kunter & Trautwein, 2013; Oser & Baeriswyl, 2002; Oser & Patry, 1994; Reyer, 2004; Seidel, 2003a, 2006b) von Unterricht abgebildet wird. Die Bezeichnung „Analytische Kompetenz“ betont dabei, dass die ProbandInnen in der Lage sind, *einzelne* Ereignisse aus dem Unterrichtsgeschehen herausfiltern und beschreiben zu können. Solche Leistungen können auch schon NovizInnen erbringen, wie die Auswertung des Tests (s. Kap. 1.5.2, S. 40) gezeigt hat. NovizInnen waren in der Lage, Situationen und Handlungen wahrzunehmen (Stufe 1) und auch den damit verbundenen Grad der Ausprägung (z.B. hohe vs. niedrige Motivation der SchülerInnen; hoher vs. niedriger Gesprächsanteil der Lehrperson) zu registrieren (Stufe 2).

| Stufe III: Prozesskompetenz      |   |
|----------------------------------|---|
|                                  | Experte   |
| Stufe 5                          | – kann den Stellenwert von didaktisch-methodischen Sinneinheiten vor dem Hintergrund des <i>gesamten</i> Unterrichtsprozesses einschätzen und für solche Einheiten begründbare Alternativen angeben |
| Stufe II: Synthetische Kompetenz |   |
| Stufe 4                          | – kann <i>einzelne</i> Handlungen oder Situationen zu <i>größeren</i> didaktisch-methodischen Einheiten verdichten und für solche Einheiten Alternativen angeben                                    |
| Stufe 3                          | – kann die Effekte bzw. Folgen, die <i>einzelne</i> Handlungen oder Situationen bewirken (können), identifizieren oder mit hoher Wahrscheinlichkeit kalkulieren bzw. voraussehen                    |
| Stufe I: Analytische Kompetenz   |   |
| Stufe 2                          | – kann den Grad der Ausprägung dieser Kategorien bestimmen  |
| Stufe 1                          | – kann einzelne Unterrichtsereignisse (Handlungen oder Situationen) in entsprechende Kategorien einordnen   |

beziehen sich auf die Tiefenstrukturen von Unterricht  
 beziehen sich auf die Sichtstrukturen von Unterricht

Abb. 4: fünfstufiges Modell der Analysekompetenz (aus: Plöger et al., 2015, S. 170)

## Hauptstufe II: Synthesekompetenz

Ab Stufe 3 aufwärts werden Analyseleistungen beschrieben, die die *Tiefenstrukturen* von Unterricht betreffen. Damit solche *Tiefenstrukturen* (Kunter & Trautwein, 2013; Oser & Patry, 1994; Reyer, 2004; Seidel, 2003a, 2003b) erfasst bzw. erschlossen werden können, müssen die ProbandInnen die sichtbaren Handlungen und/oder Situationen in einem *interpretativen* Akt zu didaktisch-methodischen Sinnzusammenhängen verdichten können (s. z.B. König & Lebens, 2012). Mit den Stufen 3 und 4 werden Grade der Analysekompetenz beschrieben, bei denen einzelne Handlungen mit weiteren einzelnen Handlungen oder mit dadurch hervorgerufenen Effekten zu größeren didaktisch-methodischen Einheiten *verbunden* werden und zu diesen Alternativen formuliert werden können. Bei ProbandInnen, die Leistungen erbringen, die die Stufen 3 und 4 betreffen, ist die Informationsverarbeitung deutlich komplexer organisiert als auf den beiden unteren Stufen. Durch das Zusammenfassen (synthetisieren) einzelner Aspekte werden diese beiden Stufen von Plöger und Scholl als „Synthesekompetenz“ bezeichnet.

## Hauptstufe III: Prozesskompetenz

Eine Person, die Stufe 5 erreicht hat, bezieht sich auf den *gesamten* zu analysierenden Unterrichtsprozess. Sie kann didaktisch-methodische Sinneinheiten vor dem Hintergrund des *gesamten* Unterrichtsverlaufs beurteilen und sowohl für solche Einheiten als auch für den gesamten Unterrichtsprozess begründbare Alternativen angeben. Daher wird diese Stufe auch als „Prozesskompetenz“ benannt, da hier der komplette Unterrichtsverlauf erfasst werden muss.

Das hier in aller Kürze vorgestellte fünfstufige Kompetenzmodell von Plöger und Scholl wird auch bei meiner Qualitativen Inhaltsanalyse zur Erfassung der Komplexität der Informationsverarbeitung genutzt. Die zum Einsatz kommenden Kategorien operationalisiere ich in Kapitel 3 und nenne an dieser Stelle die Kategorien, die zur Auswertung des Interviewmaterials zum Einsatz kommen werden:

- Verstehenstiefe – Level 1
- Verstehenstiefe – Level 2a
- Verstehenstiefe – Level 2b
- Verstehenstiefe – Level 3a
- Verstehenstiefe – Level 3b

## 1.5 Methode und Ergebnisse der Untersuchung von Plöger und Scholl (2014)

### 1.5.1 Anlage der Studie

Das im vorangegangenen Teilkapitel vorgestellte fünfstufige Kompetenzmodell wurde von Plöger und Scholl zur Konstruktion eines Tests genutzt, mit dem die Analysekompetenz von Lehrpersonen gemessen wurde.<sup>20</sup> Dieser Test (paper and pencil-Verfahren) bestand aus 23 Items (Plöger & Scholl, 2014b, S. 273), zu deren Beantwortung die ProbandInnen didaktisch-methodisches und fachliches (incl. fachdidaktisches) Wissen aktivieren mussten, um so die *Lernwirksamkeit* des unterrichtlichen Handelns angemessen einschätzen zu können. Exemplarisch soll hier ein Item vorgestellt werden, für dessen Lösung didaktisch-methodisches Wissen (hier Exemplarizität [Einheit von Konkretem und Abstraktem]) notwendig ist (s. Plöger & Scholl, 2014b, S. 280): „Zu Beginn der Unterrichtsstunde lässt der Lehrer Beispiele für Lichtbrechung nennen und führt dann selbst weitere an (mittels bildlicher Darstellung). Wie beurteilen Sie diesen Einstieg?“

Den ProbandInnen wurde als Stimulus eine videographierte Unterrichtsstunde vorgeführt (11. Klasse, Fachoberschule, Fach Physik; Thema: Snelliussches Brechungsgesetz). In dieser Physikstunde sollten die SchülerInnen das Brechungsgesetz nachentdecken: Zu Beginn der Unterrichtsstunde präsentiert die Lehrperson verschiedene Phänomene, in denen Brechung auftritt, in Form von Bildern. Im Anschluss daran werden die SchülerInnen gebeten, ein Experiment zu planen, mit dem das Brechungsgesetz nachentdeckt werden kann. Die SchülerInnen führen dann das Experiment durch, werten ihre Ergebnisse aus und formulieren mit Hilfe des Lehrers das Snelliussche Brechungsgesetz, das mittels Beispielrechnung (Berechnung des Brechungsindex aus gegebenen Winkeln) überprüft wird. Zum Abschluss der Unterrichtsstunde demonstriert die Lehrperson den SchülerInnen die Funktionsweise einer Fresnel-Linse und deklariert diese Linse als „Anwendung“ des Brechungsgesetzes.

Dieses Video wurde vornehmlich aus zwei wichtigen Gründen eingesetzt:

- 1) Das Video ist insofern repräsentativ, weil es typische Schwächen von Physikunterricht im Sinne der IPN-Videostudie (*Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik*, Seidel & Prenzel, 2007; Seidel, Prenzel, Rimmel, Schwindt et al., 2006; Seidel, Prenzel, Rimmel, Dalehefte et al., 2006) zeigt. Negativ fallen vor allem „die fehlende Einbettung des Experiments in einen hypothesengeleiteten Prozess, die starke Lehrerzentrierung, die daran gekoppelte reproduktive Gesprächsführung, die mangelhafte Phasierung der Stunde unter dem Gesichtspunkt der Lernaktivitäten und die geringe explizite und implizite Zielorientierung auf“ (Plöger & Scholl, 2014a, S. 93).
- 2) Da das Video eine komplette Unterrichtsstunde zeigt, ist die Komplexität der wahrnehmbaren bzw. erschließbaren Situationen und Handlungen entsprechend groß, sodass genügend Äußerungs- und Anknüpfungsmöglichkeiten für die ProbandInnen entstehen. Dieses Maß an Komplexität stellt zusammen mit den Schwächen des gezeigten

---

<sup>20</sup> Die Studie war als Querschnittsstudie angelegt. Die Daten wurden im Zeitraum von Juli 2010 bis März 2011 erhoben.

Unterrichts insbesondere für ExpertInnen einen hohen Anreiz dar, mögliche *Handlungsalternativen* zu entwickeln (s. Gruber, 1994).

Die Gesamtstichprobe (N = 800) umfasste vier Teilgruppen mit je 200 Personen, die an die berufsbiographischen Entwicklungsmöglichkeiten im Bereich der LehrerInnenbildung angelehnt waren: Lehramtsstudierende (an Universitäten aus sechs Bundesländern), ReferendarInnen (an Studienseminaren aus vier Bundesländern), LehrerInnen (an Gymnasien eines regionalen Bezirks) und Fach- und SeminarleiterInnen (an Studienseminaren aus sechs Bundesländern).<sup>21</sup>

Um überprüfen zu können, inwieweit das *fachliche* (incl. *fachdidaktische*) Wissen Einfluss auf die Analyseleistungen von Unterricht hat, wurde die Stichprobe zusätzlich hinsichtlich der Fachzugehörigkeit der ProbandInnen in drei Gruppen ausdifferenziert:

- Personen, die das Fach Physik unterrichten bzw. studieren,
- Personen, die mindestens ein naturwissenschaftliches Fach und/oder das Fach Mathematik unterrichten bzw. studieren,
- Personen, die weder ein naturwissenschaftliches Fach und auch nicht das Fach Mathematik unterrichten bzw. studieren.

**Tab. 1:** Zusammensetzung der Stichprobe nach Fachzugehörigkeit (in Anlehnung an Plöger & Scholl, 2014b, S. 280)<sup>22</sup>

|   | Studierende | Referendare | Lehrpersonen | Fach- & SeminarleiterInnen |
|---|-------------|-------------|--------------|----------------------------|
| Physik als Unterrichtsfach                                      | 8           | 7           | 36           | 20                         |
| Mathematik und/oder mindestens ein naturwissenschaftliches Fach | 90          | 40          | 56           | 58                         |
| Kein naturwissenschaftliches Fach                               | 102         | 153         | 108          | 122                        |

Der Untersuchung von Plöger und Scholl lagen zwei *Hypothesen* zugrunde (s. Plöger & Scholl, 2014a, 2014b):

- Hypothese 1: „*Experten (Fach- und Seminarleiter) erreichen die höchsten Stufen der Analysekompetenz, während die Kompetenz von Novizen (Studierenden) über die ersten zwei Stufen nicht weit hinausgeht.*“<sup>23</sup> (Plöger & Scholl, 2014b, S. 279, kursiv im Original)
- Hypothese 2: „*Personen, die das Fach Physik oder Mathematik bzw. ein weiteres naturwissenschaftliches Fach unterrichten, erzielen aufgrund ihrer fachlichen bzw. fachdidaktischen Kompetenzen höhere Analyseleistungen als Personen, die keines dieser Fächer unterrichten.*“ (Plöger & Scholl, 2014b, S. 279, kursiv im Original)

21 Die Stichprobe für die Studierenden, ReferendarInnen und Fach- und SeminarleiterInnen wurde als Gelegenheitsstichprobe gezogen, die als Klumpen erhoben wurde. Die Gruppe der Lehrpersonen wurde als Klumpenstichprobe durch eine zweistufige Zufallsauswahl erhoben.

22 Zur Harmonisierung hinsichtlich der Stichprobenzusammensetzung wurde die ursprüngliche Tabelle von Plöger und Scholl (2014b) angepasst.

23 Als Rangfolge hinsichtlich der Stichprobe wird dabei erwartet, dass Fach- und SeminarleiterInnen die höchsten Stufen der Analysekompetenz erreichen. Ihnen sollten LehrerInnen und ReferendarInnen folgen. Die niedrigste Ausprägung der Analysekompetenz sollten Studierende aufweisen.

### 1.5.2 Zentrale Ergebnisse der Studie von Plöger und Scholl (2014)

Die 23 Items bildeten eine raschhomogene Skala mit einer Reliabilität von .79 (Plöger & Scholl, 2014a).<sup>24</sup> Die nachfolgenden Ergebnisse stellen die Ergebnisse im Hinblick auf die angeführten Hypothesen dar (s. Plöger & Scholl, 2014a, 2014b, 2014c, 2016):

- Hinsichtlich der Höhe der Analysekompetenz gibt es deutliche Unterschiede zwischen den ProbandInnengruppen: Studierende erzielen einen relativ niedrigen Fähigkeitswert, Fach- und SeminarleiterInnen dagegen den höchsten. Zwischen diesen beiden Gruppen liegen die ReferendarInnen und die LehrerInnen.<sup>25</sup>
- Die Studierenden können (von wenigen Ausnahmen abgesehen) nur die Items der ersten beiden Unterstufen des fünfstufigen Modells lösen. Die Fach- und SeminarleiterInnen erreichen durchschnittlich die höchsten Leistungen und sind in der Lage, auch Items zu lösen, die der fünften Stufe des Modells zuzuordnen sind.
- Darüber hinaus konnte nachgewiesen werden, dass das fachliche Wissen eine zentrale Voraussetzung für die Analyse der videografierten Unterrichtsstunde ist, denn Personen, die das Fach Physik unterrichteten, erreichten signifikant höhere Analyseleistungen als Personen, die dieses Fach nicht unterrichteten ( $[F(2, 788) = 5,973, p < 0,01]$ ). Allerdings gab es unter den besten 30 Personen auch „Ausreißer“, also „auch einige wenige, die exzellente Analyseleistungen erzielen, obwohl sie weder Physik noch Mathematik und/oder ein anderes naturwissenschaftliches Fach unterrichten.“ (Plöger & Scholl, 2016, S. 183)

---

24 Da sich die Itemskala als *raschhomogen* (Reliabilität von .79, s. Plöger & Scholl, 2014a) erwies, können die vorliegenden Summenwerte im Raschmodell als eine *suffiziente* Statistik gewertet werden. Der Grundgedanke dabei ist, dass die gelösten Aufgaben als Indikator für die Personenfähigkeit herangezogen werden: „Eine Person mit einer niedrigen Ausprägung dieser Fähigkeit wird z.B. nur die einfachen Aufgaben eines Tests lösen können. Eine Person mit einer höheren Fähigkeit hingegen wird sowohl die einfachen als auch die schwierigen Aufgaben lösen können – und damit insgesamt mehr Aufgaben lösen.“ (Strobl, 2010, S. 15)

25 Die über die Varianzanalyse ermittelten Leistungen der vier ProbandInnengruppen unterschieden sich signifikant voneinander ( $F[3, 796] = 84,554, p < .001$ ). Eine anschließende Post-Hoc-Analyse (Tukey-HSD-Test) zeigte, dass sich Studierende signifikant von allen anderen Gruppen unterschieden ( $p < .01$ ). „Dagegen unterscheiden sich die Lehrpersonen nur von den Studierenden signifikant. Referendarinnen und Referendare und Seminar-/Fachleiterinnen und -leiter unterscheiden sich signifikant sowohl von den Studierenden als auch voneinander, allerdings nicht von den Lehrpersonen.“ (Plöger & Scholl, 2014a, S. 104)

## 1.6 Quantitativer Validierungsversuch von Plöger, Scholl und Seifert (2015)

Im Sinne der oben referierten Erkenntnisse aus der Expertiseforschung waren diese Ergebnisse zu erwarten (s. Kap. 1.4.1). Insbesondere der große Unterschied zwischen den Leistungen der Fach- und SeminarleiterInnen (ExpertInnen) und der Studierenden (NovizInnen) war ein Beleg dafür, dass der Test eine klare Unterscheidung von „Extremgruppen“ zulässt, von denen vorab angenommen werden konnte, dass die eine Gruppe das in Frage stehende Merkmal (hier Analyseleistung) in hohem Maße besitzt, während dies bei der anderen Gruppe gerade nicht der Fall ist.

Im Sinne eines solchen *well-known-group-Vergleichs* (Cronbach & Meehl, 1955; DeVellis, 1991; Hattie & Cooksey, 1984; Rubin & Babbie, 2015) stellte dieses Testergebnis somit einen wichtigen Beleg für die Validität des zugrunde liegenden Konstrukts Analysekompetenz dar.<sup>26</sup> Um dieser Frage der Validität des Konstrukts weiter nachzugehen, führten Plöger et al. (2015) zusätzlich Konfirmatorische Faktorenanalysen (CFA) durch.

Im Folgenden erläutere ich diesen quantitativen Validierungsversuch eingehend, weil er den *zentralen* Bezug zu meiner Studie darstellt, in der ich auf qualitativem Wege prüfen möchte, ob die von mir erhobenen empirischen Daten ebenfalls mit den Annahmen des Konstrukts interpretiert werden können.

### 1.6.1 Modellprüfung mittels Konfirmatorischer Faktorenanalysen

Zunächst greife ich auf die Grafik aus der Einleitung (s. Abb. 1, S. 15) zurück und nenne noch einmal die zentralen Annahmen, die das Konstrukt Analysekompetenz bestimmen: Das Konstrukt Analysekompetenz wird über zwei Dimensionen, eine inhaltliche und eine formale, mit entsprechenden Faktoren definiert. Die inhaltliche Dimension ist dabei durch zwei Faktoren näher bestimmt: das *didaktisch-methodische* (DM) und das *fachliche* (incl. *fachdidaktische*) (F) Wissen. Die formale Dimension wird über den Grad der für die Analyseleistungen notwendigen Komplexität der Informationsverarbeitung abgebildet. In Anlehnung an das fünfstufige Kompetenzmodell (s. Abb. 4, S. 36) ist die formale Dimension dabei durch drei Faktoren gekennzeichnet (in aufsteigender Reihenfolge), die den oben beschriebenen Hauptstufen des Modells entsprechen: Die Stufen 1 und 2 des Modells repräsentieren den Faktor der *Analytischen Kompetenz* (AK), die Stufen 3 und 4 stellen die *Synthetische Kompetenz* (SK) dar und die Stufe 5 steht für den Faktor der *Prozesskompetenz* (PK).

Je nach Wahl bzw. Kombination dieser Dimensionen und der zugehörigen Faktoren ergeben sich mindestens fünf mögliche Strukturmodelle, die das Konstrukt repräsentieren und deren Validität im Rahmen der CFA geprüft wurde (s. Plöger et al., 2015):

---

<sup>26</sup> Eine zentrale Rolle hatte die Frage der *inhaltlichen* Validität bereits bei der Testkonstruktion und -auswertung gespielt. Die inhaltliche Validität der Testitems wurde über ein Rating von ca. 40 Seminar- und FachleiterInnen gesichert. Diese Personen waren an der Diskussion und Prüfung der akzeptierbaren Antworten auf die einzelnen Items beteiligt. Nach Abschluss der Diskussion konnte ein *Auswertungsmanual* erstellt werden, das die Grundlage für die Kodierung der Antworten der ProbandInnen bildete.

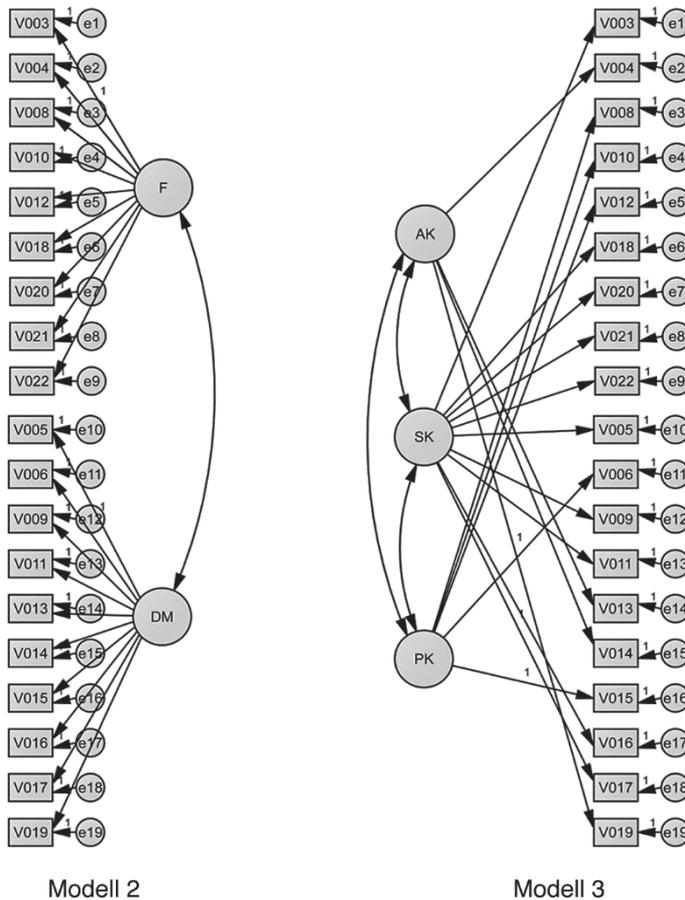
*Modell 1: eindimensionales Konstrukt (Analysekompetenz als g-Faktor)*

Mit diesem Modell wird die Varianz aller manifesten Variablen durch einen Generalfaktor erklärt. Aufgrund der einfachen Struktur des Modells hatten Plöger et al. (2015) auf eine Abbildung verzichtet.

*Modell 2 (s. Abb. 5, S. 42): eindimensional-zweifaktorielles Konstrukt (didaktisch-methodisches [DM] und fachliches Wissen [F] als zwei Faktoren der inhaltlichen Dimension)*  
 Dabei laden die Items jeweils auf einem der Faktoren der inhaltlichen Dimension (entweder auf didaktisch-methodischem oder auf fachlichem Wissen).

*Modell 3 (s. Abb. 5, S. 42): eindimensional-dreifaktorielles Konstrukt (analytische Kompetenz [AK], Synthetische Kompetenz [SK] und Prozesskompetenz [PK] als drei Faktoren („Stufen“) der formalen Dimension)*

Bei diesem Modell laden die Items jeweils auf einem der drei Faktoren der formalen Dimension (Komplexität der Informationsverarbeitung).

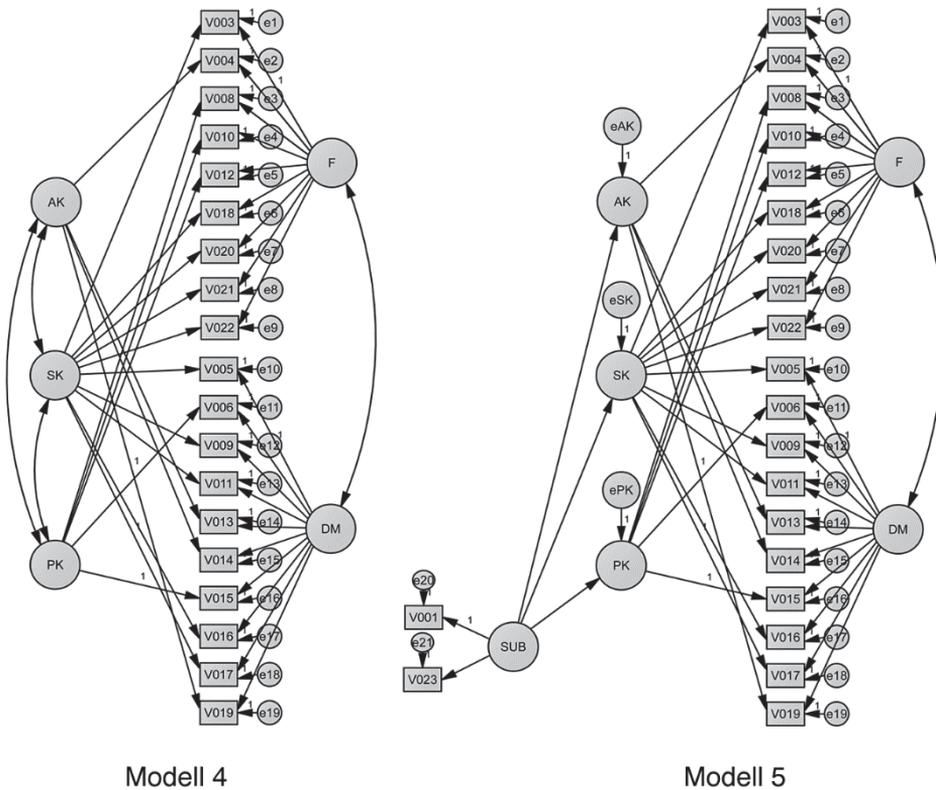


**Abb. 5:** Modell 2 und Modell 3 (aus: Plöger et al., 2015, S. 174)

Während in den Modellen 2 und 3 *keine* Doppelladungen auf der inhaltlichen und formalen Dimension zulässig waren, werden solche Doppelladungen in den Modellen 4 und 5 berücksichtigt.

*Modell 4* (s. Abb. 6, S. 43): *zweidimensionales Konstrukt* (Integration der inhaltlichen Dimension [mit zwei Faktoren] und der formalen Dimension [mit drei Faktoren])

Modell 4 kombiniert die Modelle 2 und 3 und stellt somit ein zweidimensionales Konstrukt dar, mit dem „die Varianz der manifesten Variablen durch die Faktoren der beiden unabhängigen Dimensionen erklärt werden“ (s. Plöger et al., 2015, S. 174) soll. Dabei erfasst jedes Item einerseits die *inhaltliche* Qualität des Wissens (didaktisch-methodisch *oder* fachliches Wissen) und andererseits dessen *formale* Qualität (Komplexitätsgrad der zu verarbeitenden Informationen).



**Abb. 6:** Modell 4 und Modell 5 (aus: Plöger et al., 2015, S. 175)

Es gilt aber die Restriktion, dass innerhalb der inhaltlichen Dimension (didaktisch-methodisches *und* fachliches Wissen) keine Doppelladungen erlaubt sind. Jedes Item soll *schwerpunktmäßig* entweder das didaktisch-methodische Wissen *oder* das fachliche Wissen erfassen. Dadurch können die Items dem entsprechenden Faktor der inhaltlichen Dimension zugeordnet werden. „Da aber davon auszugehen ist, dass die von jedem Item

erzeugte Varianz einen Rest des anderen Faktors mit abdeckt, wird diesem Sachverhalt durch die Korrelation von didaktisch-methodischem und fachlichem Wissen Rechnung getragen.“ (ebd., S. 175)

*Modell 5* (s. Abb. 6, S. 43): *zweidimensionales Konstrukt*

In Modell 5 wird, in Ergänzung zu Modell 4, das sogenannte *Subitizing*, also die Fähigkeit von ExpertInnen, Situationen schnell erfassen und sicher beurteilen zu können (s. Bromme, 1992), im Zusammenhang mit der formalen Dimension als *exogene Variable* herangezogen. Dazu wird zu Beginn und am Ende des Tests das identische Item eingesetzt, um mit dem globalen Urteil über die Unterrichtsstunde die „Varianz in den drei Faktoren der Komplexität der Informationsverarbeitung“ (Plöger et al., 2015, S. 175) aufzuklären.<sup>27</sup>

Hinsichtlich dieser fünf Modelle waren die durchzuführenden CFA an zwei zu prüfende Hypothesen geknüpft (s. Plöger et al., 2015, S. 175f.):

- *Hypothese 1*: „Die Modelle 1, 2 und 3 stellen keine adäquate Modellierung des Konstrukts Analysekompetenz dar.“ (ebd., S. 175)
- *Hypothese 2*: „Die Modelle 4 und 5 bilden das Konstrukt Analysekompetenz angemessen ab.“ (ebd., S. 176)

Um die Faktorenanalysen durchzuführen, wurde den einzelnen Faktoren der zu bildenden Modelle Testitems als Messindikatoren zugeordnet „unter der Voraussetzung, dass die zur Lösung der betreffenden Items notwendigen Anforderungen den theoretischen Annahmen entsprechen.“ (ebd., S. 173). Tabelle 2 stellt die Aufteilung der Items auf die jeweiligen Faktoren dar.

---

<sup>27</sup> Der Subitizing-Faktor wird über zwei Items zur *globalen* Bewertung der Unterrichtsstunde operationalisiert: „Das erste und letzte Item des Tests messen die Fähigkeit des Subitizing auf einer fünfstufigen Likertskala ...“ (Plöger et al., 2015, S. 177). Beide Items sind identisch formuliert („Wie bewerten Sie diese Unterrichtsstunde insgesamt?“) und zielen darauf ab, ob das rasche Urteil zu Beginn des Fragebogens über die Güte der zu analysierenden Unterrichtsstunde korrekt ist *und* mit der Bewertung der selbigen am Ende des Tests übereinstimmt (s. Plöger & Scholl, 2014a, S. 94).