

**Entwicklung eines standardorientierten  
Situational Judgement Tests zur Erfassung  
professioneller pädagogischer Kompetenz**

**Dissertation  
zur Erlangung des Doktorgrades  
der Naturwissenschaften**

**vorgelegt beim Fachbereich 05 – Psychologie  
der Johann Wolfgang Goethe-Universität  
in Frankfurt am Main**

**von**

**Ines Weresch-Deperrois  
aus Karlsruhe**

Frankfurt, 2013

(D 30)

vom Fachbereich 05 - Psychologie der

Johann Wolfgang Goethe - Universität als Dissertation angenommen.

Dekan: Prof. Dr. Rolf van Dick

Gutachter: Prof. Dr. Holger Horz  
Prof. Dr. Regina Vollmeyer

Datum der Disputation: 18.06.2014

## Inhalt

Inhalt.....	3
Zusammenfassung .....	5
Einleitung.....	8
1. Einordnung der Arbeit in den theoretischen und empirischen Rahmen.....	12
1.1 Definition: Kompetenz.....	12
1.2 Modellierung der professionellen pädagogischen Kompetenz .....	16
1.3 Die Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004) als „Modell“ der professionellen pädagogischen Kompetenz .....	20
1.4 Die empirische und theoretische Basis der Standards für die Lehrerbildung: Bildungswissenschaften .....	24
1.5 Die Funktion von Bildungsstandards .....	57
1.6 Die Evaluation der Lehrerbildung.....	60
2. Fragestellung .....	72
3. Die Operationalisierung professioneller pädagogischer Kompetenz im Rahmen des diagnostischen Prozesses .....	74
3.1 Situational Judgement Tests.....	75
3.1.1 Definition: Situational Judgement Test.....	75
3.1.2 Die Validität von Situational Judgement Tests.....	77
3.1.3 Situational Judgement Tests in Abgrenzung zu anderen situationsorientierten Verfahren.....	84
3.1.4 Situational Judgement Tests zur Erfassung professioneller pädagogischer Kompetenz.....	91
3.2 Generierung der Items .....	99
3.2.1 1. Schritt: Auswahl geeigneter Items aus dem Instrument KOSTA.....	100
3.2.2 2. Schritt: Empirische Basis der Items mit negativer Polung.....	102
3.2.3 3. Schritt: Entwicklung der Items .....	108
3.2.4 4. Schritt: Eignungseinschätzung der generierten Items.....	125
4. Ergebnisse .....	130
4.1 Eignungseinschätzung der Items durch ExpertInnen .....	130
4.1.1 Die ExpertInnengruppe .....	130
4.1.2 Die Bewertung der generierten Items .....	131
4.2 Explorative Datenanalyse und Gütekriterien.....	131
4.2.1 Methode .....	132
4.2.2 Durchführung der Untersuchung.....	132

4.2.3	Stichprobe .....	133
4.2.4	Explorative Datenanalyse und Gütekriterien .....	135
5.	Diskussion der Ergebnisse .....	162
5.1	Itemanalyse.....	162
5.2	Haupt- und Nebengütekriterien.....	164
6.	Zusammenfassung und Ausblick .....	181
7.	Literatur.....	192
	Anhang D: Als Basis fungierende Items des Instrumentes KOSTA sowie deren Trennschärfen.....	213
	Anhang A: Verzeichnis der Abbildungen .....	214
	Anhang B: Verzeichnis der Tabellen .....	216
	Anhang C: Die Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004) ....	217
	Anhang D: Als Basis fungierende Items des Instrumentes KOSTA sowie deren Trennschärfen.....	225
	Anhang E: Das Instrument der qualitativen Vorstudie „Qualitative Erfassung professioneller pädagogischer Kompetenzen im Lehrberuf“ .....	229
	Anhang F: Interrater-Reliabilität und deskriptive Ergebnisse der qualitativen Datenerhebung zur Gewinnung der Basis negativ gepolter Items .....	243
	Anhang G: Eignungseinschätzung der 102 Items des Situational Judgement Tests durch 6 ExpertInnen .....	250
	Anhang H: Der Situational Judgement Test .....	253
	Anhang I: Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests .....	270
	Anhang K: Zuordnung der Items zu Kompetenzen durch 8 ExpertInnen .....	303

## Zusammenfassung

Kompetenzen bezeichnen „erlernbare kontextspezifische Leistungsdispositionen, die sich funktional auf Situationen und Anforderungen in bestimmten Domänen beziehen“ (Klieme & Hartig, 2007, S. 17).

Aufgrund der Domänenspezifität von Kompetenzen ergeben sich entsprechend differentieller Aufgaben- und Anforderungsbereiche ähnliche – doch nicht vollständig deckungsgleiche – Konnotationen des Begriffs. Somit ist es notwendig, für verschiedene Domänen verschiedene Kompetenzdefinitionen zur Verfügung zu stellen.

Im Zuge eines Wandels der Input- zur Output-Orientierung (z. B. Klieme et al., 2007; Messner, 2004) (nicht nur) im Bereich der Lehrerbildung – mithin einer Abkehr der Steuerung durch Gesetze, Lehrpläne oder Ähnlichem hin zu einer Orientierung an Effekten und Wirkungen – wurden durch die Kultusministerkonferenz (2004) Standards für die Lehrerbildung: Bildungswissenschaften formuliert, die als Explikation professioneller pädagogischer Kompetenzen gelten können. Diese Standards definieren konkrete Kompetenzen im Sinne einer „Idealnorm“ für die Kompetenzbereiche (1) sach- und fachgerechte Planung und Durchführung von Unterricht, (2) Gestaltung von förderlichen Lernsituationen, Motivation, Bereitstellen von Transferoptionen, (3) Förderung des selbstbestimmten Lernens und Arbeitens der SchülerInnen, (4) Wissen um soziale und kulturelle Lebensbedingungen von SchülerInnen bzw. Einflussnahme auf deren individuelle Entwicklung, (5) Vermittlung von Werten und Normen, Unterstützung selbstbestimmten Urteilens und Handelns durch die SchülerInnen, (6) Lösung von Schwierigkeiten und Konflikten in Schule und Unterricht, (7) Diagnostik von Lernvoraussetzungen und –prozessen, gezielte Förderung sowie Beratung von SchülerInnen und Eltern, (8) Leistungserfassung aufgrund transparenter Beurteilungsmaßstäbe, (9) Bewusstsein über Anforderungen des Lehrberufs sowie dessen besondere Verantwortung und Verpflichtung, (10) Verständnis des Lehrberufs als ständige Lernaufgabe sowie (11) Beteiligung an Planung und Umsetzung schulischer Projekte und Vorhaben.

Durch die Formulierung dieser Standards (ebd.) sieht sich die Lehrerbildung mit der Aufgabe konfrontiert, Studierende des Lehramts so auszubilden, dass die resultierenden Kompetenzen der Studierenden den genannten Standards genügen.

Dies impliziert eine standardorientierte Evaluation der Lehrerbildung, welche nur mit adäquaten, validen Instrumenten der Kompetenzdiagnostik zu realisieren ist. Bislang wird eine Evaluation der Lehrerbildung (und somit der professionellen pädagogischen Kompetenz von Lehramtsstudierenden) meist durch Selbsteinschätzungsverfahren umgesetzt – eine Methodik, die im Bereich der Kompetenzdiagnostik kritisch diskutiert wird (z. B. Nerdinger, Blickle und Schaper, 2008). Der Bedarf an kontextnahen, standardorientierten Verfahren zur Diagnostik professioneller pädagogischer Kompetenz besteht folglich. Um eine anforderungsbezogene Messung solcher Kompetenzen realisieren zu können, eignen sich simulationsorientierte kompetenzdiagnostische Ansätze wie etwa Assessment Center, Videovignetten (z. B. Blomberg, 2011, oder Riese, 2010), situative Übungen oder Situational Judgement Tests (Schaper, Hilligus & Reinhold, 2009). Diese Verfahren sollen es ermöglichen, „kontextgebundene Wissensstrukturen zu aktivieren und so einer Erfassung zugänglich zu machen.“ (Seidel & Prenzel, 2007, S. 204)

Für den Einsatz mit großen Stichproben – wie sie etwa bei der Evaluation der Lehrerbildung anfallen würden – eignet sich jedoch der Großteil situationsorientierter Verfahren aus ökonomischen Gründen nicht (z. B. Seifert, Hilligus & Schaper, 2009). So ist beispielsweise eine Durchführung von Assessment Centers mit Studierendenkohorten aufgrund des Durchführungs- und Auswertungsaufwandes kaum möglich. Auch der Einsatz von Videovignetten als Items eines situationsorientierten Kompetenztests zeigt sich nicht nur in der Entwicklung (Videographie von Unterricht oder Nachstellen desselben, Auswahl und Schnitt anfallenden Materials, ...), sondern ebenso in der Darbietung in konkreten Testsituationen aufwendig. Situational Judgement Tests hingegen weisen die Vorteile situationsorientierter Verfahren auf und können zusätzlich aufgrund ihrer Darbietung im paper-pencil-Format problemlos auf große Stichproben angewendet werden.

In der vorliegenden Studie wurde ein solcher Situational Judgement Tests entwickelt, um die professionelle pädagogische Kompetenz Lehramtsstudierender *kontextnäher* erfassen zu können als dies durch Selbsteinschätzungsverfahren geleistet werden kann und um eine *ökonomische* Erfassung der professionellen pädagogischen Kompetenz Lehramtsstudierender zu ermöglichen. Ein Situational Judgement Test könnte im Rahmen des Lehramtsstudiums an Stelle von Self-Assessment-Instrumenten eingesetzt werden, um einerseits gegenüber den

Studierenden eine validere Rückmeldefunktion hinsichtlich der eigenen Kompetenzen bereitstellen zu können, andererseits, um den ausbildenden Hochschulen im Sinne einer Qualitätssicherung Informationen über die Wirkung und Wirksamkeit der Lehrerbildung liefern zu können.

Bei der Itemgenerierung wurden sowohl ExpertInnen als auch NovizInnen beteiligt, um durch die Items ein breites Spektrum an Kompetenzausprägungen reflektieren (z. B. Bruder, 2011). Zusätzlich wurden die schließlich entwickelten Items zur Absicherung der Inhalts- sowie Konstruktvalidität zwei ExpertInnengruppen zur Eignungseinschätzung vorgelegt.

Die vorliegende Studie hatte zum Ziel, den entwickelten standardorientierten Situational Judgement Test anhand verschiedener Außenkriterien (z. B. Prüfungsnoten, die Anzahl bislang abgeleiteter Praktika oder das Studienfach) konvergent und diskriminant zu validieren. Hierbei wurden konvergente Validitätsnachweise lediglich teilweise signifikant, wobei diskriminante Validitätsnachweise anhand des Studienfachs (Lehramtsstudium vs. Studium technisch-naturwissenschaftlicher Fächer) deutliche Signifikanzen und Effektstärken ergaben.

Erste Hinweise auf die Eignung des generierten Instruments zur standardorientierten Erfassung der professionellen pädagogischen Kompetenz bestehen folglich, wobei aufgrund teilweise unbefriedigender Ergebnisse der konvergenten Validitätsüberprüfung weitere Analysen dringend angezeigt sind.

## Einleitung

„Entscheidend und in gewisser Weise tatsächlich revolutionär für den Schulbereich ist es, sich bei der Steuerung nicht länger nur am Prinzip einer immer detaillierteren Vorgabe von Inputs (Gesetze, Lehrpläne, Erlasse, Stundentafeln, Ordnungen), sondern verstärkt an den Outputs bzw. Outcomes, also an tatsächlich erreichten Effekten und Wirkungen zu orientieren – und diese mit gesetzten Standards zu vergleichen.“  
(Terhart, 2002, S. 6)

Eine solche outputorientierte Sicherung der Bildungsqualität ist eine zentrale Aufgabe der Kultusministerkonferenz. Um diese Bildungsqualität überprüfen und sichern zu können, wurden Standards sowohl für den schulischen Unterricht als auch für Lehrkräfte formuliert, welche sich auf verschiedene Fächer in Schule (z. B. „Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss, Jahrgangsstufe 10“, KMK, 2003) und Hochschule (z. B. „Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung“, KMK, 2008) beziehen. Durch die Formulierung solcherlei verbindlicher Standards lässt sich die Zielerreichung hinsichtlich der Qualitätsanforderungen systematisch überprüfen, wobei

- a) einerseits durch „bundesweit geltende Bildungsstandards zur Entwicklung und Vergleichbarkeit der Qualität schulischer Bildung im föderalen Wettbewerb der Länder“ (KMK, 2003, S.5) die Qualität schulischer Bildung, die Vergleichbarkeit schulischer Abschlüsse sowie die Transparenz des Schulsystems gesichert wird.
- b) andererseits für den Bereich der Lehrerbildung durch die „*Standards für die Lehrerbildung: Bildungswissenschaften*“ (KMK, 2004) sowie die „*Ländergemeinsamen inhaltlichen Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung*“ (KMK, 2008) Kompetenzbereiche definiert wurden, über welche eine Lehrkraft verfügen soll. Die Standards für die Lehrerbildung: Bildungswissenschaften sind seit dem Ausbildungsjahr 2005/2006 in der Lehrerbildung der Bundesländer implementiert, wobei insbesondere die Studienordnungen der Lehramtsstudiengänge, der Vorbereitungsdienst (Referendariat) sowie die Fort- und Weiterbildung von



Lehrkräften betroffen sind. Diese Standards beziehen sich auf die Kompetenzbereiche (1) sach- und fachgerechte Planung und Durchführung von Unterricht, (2) Gestaltung von förderlichen Lernsituationen, Motivation, Bereitstellen von Transferoptionen, (3) Förderung des selbstbestimmten Lernens und Arbeitens der SchülerInnen, (4) Wissen um soziale und kulturelle Lebensbedingungen von SchülerInnen bzw. Einflussnahme auf deren individuelle Entwicklung, (5) Vermittlung von Werten und Normen, Unterstützung selbstbestimmten Urteilens und Handelns durch die SchülerInnen, (6) Lösung von Schwierigkeiten und Konflikten in Schule und Unterricht, (7) Diagnostik von Lernvoraussetzungen und –prozessen, gezielte Förderung sowie Beratung von SchülerInnen und Eltern, (8) Leistungserfassung aufgrund transparenter Beurteilungsmaßstäbe, (9) Bewusstsein über Anforderungen des Lehrberufs sowie dessen besondere Verantwortung und Verpflichtung, (10) Verständnis des Lehrberufs als ständige Lernaufgabe sowie (11) Beteiligung an Planung und Umsetzung schulischer Projekte und Vorhaben.

Eine Output-Orientierung im Rahmen der Lehrerbildung impliziert die Erfassung von Kompetenzen und Kompetenzentwicklungen der Studierenden vom Beginn des Studiums bis zum Eintritt in den Lehrberuf. Doch welche Kompetenzen muss eine (angehende) Lehrperson aufweisen? Eine pragmatische Antwort liefern die von der KMK (ebd.) verfassten elf Standards für die Lehrerbildung: Bildungswissenschaften. Diese Standards spannen einen Bezugsrahmen, an dem sich pädagogisches Handeln – und folglich die Ausbildung von Lehrpersonen – ausrichten soll. Freilich müssen diese Standards durch spezifische Kompetenzen und Handlungsskripte konkretisiert werden.

Um den individuellen Erfüllungsgrad dieser Standards bei Lehramtsstudierenden überprüfen zu können, müssen valide Instrumente der Kompetenzdiagnostik zur Verfügung stehen, welchen die durch die KMK (2004) formulierten Standards zugrundeliegen. Bislang wird eine Überprüfung der Wirksamkeit von Lehrerbildung hinsichtlich des erwünschten Kompetenzzuwachses der Studierenden insbesondere durch Selbsteinschätzungsverfahren (z. B. Abs, 2005) oder retrospektive Befragungen hinsichtlich der Relevanz verschiedener Aspekte der Ausbildung realisiert.

Die Validität von self-assessment-Methoden im Bereich der Kompetenzeinschätzung ist jedoch begrenzt, da beispielsweise einerseits soziale Erwünschtheit die Antworten verzerren könnte und andererseits eine von der konkreten Handlung abgekoppelte Einschätzung eigener Kompetenzen nur eingeschränkt valide gelingt. Solcherlei Instrumente messen nach Hartig und Jude (2007) sowie McClelland (1987) nicht die tatsächliche *Kompetenz*, sondern das *Selbstkonzept eigener Fähigkeiten*. Frese, Fay, Hilburger und Leng (1997) argumentieren am Beispiel des Konstruktes der persönlichen Initiative ähnlich, wenn sie sagen, dass traditionelle self-assessment-Verfahren zwar die *Wichtigkeit* erfassen, die Personen dem Konstrukt beimessen, nicht jedoch das *Ausmaß* tatsächlicher Initiative im Arbeitskontext.

Kompetenzen können durch simulationsorientierte Verfahren wie etwa Situational Judgement Tests (Weekley & Ployhart, 2006) realitätsnah erfasst werden: Im Rahmen dieses Verfahrens sollen die ProbandInnen kurz dargestellte Situationen – meist in Form eines Fragebogens – analysieren oder adäquate Handlungen zur Problemlösung generieren. Kompetenzen werden hiermit handlungsnah abgefragt (zur näheren Darstellung der Methode siehe Kap 3.1).

Nach Schaper, Hilligus und Reinhold (2009) sollten Verfahren der Kompetenzmessung innerhalb einer Domäne anforderungsbezogen und ganzheitlich die betreffenden Kompetenzfacetten erfassen. Den Autoren zufolge ist es für eine anforderungsbezogene Messung nötig, Erhebungsverfahren gemäß ‚simulationsorientierter‘ eignungsdiagnostischer Ansätze (z. B. in Form von Arbeitsproben, situativen Übungen oder Situational Judgement Tests) zu entwickeln. Simulationsorientierte Verfahren sollen es hierbei ermöglichen, „kontextgebundene Wissensstrukturen zu aktivieren und so einer Erfassung zugänglich zu machen.“ (Seidel & Prenzel, 2007. S. 204)

Ziel dieser Arbeit ist es, ein solches simulationsorientiertes Instrument zu entwickeln, welches entlang der KMK-Standards die professionelle pädagogische Kompetenz Lehramtsstudierender fachübergreifend und situations- sowie kontextbezogen erheben kann.

Folgender Frage soll somit nachgegangen werden:

- Lässt sich entlang der Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004) ein inhaltsvalider Situational Judgement Test entwickeln?

Kapitel 1 ordnet die Fragestellung in den theoretischen und empirischen Rahmen ein. Hier wird zunächst der Begriff der Kompetenz geklärt und auf den Bereich des Lehrberufs übertragen. Die Standards für die Lehrerbildung: Bildungswissenschaften der KMK (2004) werden vorgestellt und in den Bezug zur empirischen Forschungsergebnissen der Pädagogischen Psychologie gesetzt.

Kapitel 2 expliziert die Fragestellung der vorliegenden Untersuchung, bevor in Kapitel 3 die Schritte der Operationalisierung des Konstrukts „standardorientierte professionelle pädagogische Kompetenz“ vorgestellt werden. Hierbei wird zunächst die Methode „Situational Judgement Test“ dargelegt. Anschließend werden die Auswahl der Itembasis, die qualitative Vorstudie zur Gewinnung realitätsnaher Items sowie die Experteneinschätzung bezüglich der Eignung schließlich gestalteter Items erörtert.

In Kapitel 4 werden die Ergebnisse zur Analyse der Gütekriterien des generierten Situational Judgement Tests vorgestellt und diskutiert.

Kapitel 5 diskutiert schließlich das Vorgehen sowie die Ergebnisse und liefert einen Ausblick auf mögliche weiterführende Modifikationen und Forschungsaufgaben.

# 1. Einordnung der Arbeit in den theoretischen und empirischen Rahmen

Der Kompetenzbegriff ist in aller Munde. Da ist beispielsweise die Rede von „Medienkompetenz“, „Selbstorganisationskompetenz“, „naturwissenschaftlicher Kompetenz“, „interkultureller“ oder eben „pädagogischer Kompetenz“. Doch was ist das überhaupt – Kompetenz?

## 1.1 Definition: Kompetenz

Der lateinische Ursprung des Wortes „Kompetenz“, „competentia“, bedeutet sowohl Symmetrie, Konstellation als auch wörtlich Kompetenz. Das Verb „competere“ beschreibt gemeinsam/zugleich erstreben, zusammentreffen, etwas mächtig sein, ausreichen, entsprechen, zustehen, zukommen oder kämpfen.

Im Englischen wird das Wort „competence“ sowohl im Kontext von Befugnissen (Amtsbefugnis, Zuständigkeit) als auch im Kontext des „Könnens“ (Befähigung, Fähigkeit, Sachkunde, Tüchtigkeit) verwandt. Das französische „compétence“ bedeutet zum Einen die Amts- oder Entscheidungsbefugnis oder die Zuständigkeit, zum Anderen die Sachkenntnis. Zusätzlich wird hier der Begriff auch für die „Federführung“ oder den „Geschäftsbereich“ verwendet. Auch im Spanischen und Italienischen ergibt sich ein ähnliches Bild.

Nach Erpenbeck und von Rosenstiel (2007) werden die Begriffe „competentia“ und „Competenz“ zum ersten Mal in Zedlers Untersallexikon von 1753 im Sinne der heutigen Wortbedeutung verwandt. Der Duden (2004) definiert heute „Kompetenz“ einerseits als Befähigung, Begabung, Fähigkeit, Fertigkeit, Können, Qualifikation, Sachverstand, Sachverständnis, oder Talent, andererseits als Zuständigkeit oder Befugnis (Deutsche Akademie der Wissenschaften, 2003).

Diese Alltagsdefinition wird auf verschiedene Wissenschaftsbereiche transferiert und nimmt aufgrund deren differentieller Aufgaben- und Anforderungsbereiche sowie Foci ähnliche – doch nicht vollständig deckungsgleiche – Konnotationen an. „Kompetenz“ ist folglich sowohl dispositional als auch kontextualisiert und zeigt sich im Bewältigen spezifischer Anforderungen der jeweiligen Domäne.

Lehmann und Nieke (1999) unterscheiden dabei drei Bedeutungen des Begriffes „Kompetenz“:

- *Formale, auch juristisch genau definierte Kompetenzen, welche Zugang zu bestimmten Befugnissen gewähren.* Bezogen auf den schulischen Kontext erwerben folglich SchülerInnen mit einem Schulabschluss Zugang zu weiterführenden Bildungseinrichtungen (z. B. Erwerb der Hochschulzugangsberechtigung durch das Abitur), LehrerInnen erwerben die Berechtigung, ein bestimmtes Lehramt ausüben zu können (z. B. Abschluss eines schulspezifischen Lehramtsstudiums mit einer bestimmten Fächerkombination).
- *Kompetenzen als Fertigkeiten und Fähigkeiten im Sinne einer Expertise.* So können SchülerInnen jene Kompetenzen erreichen, die das Qualifikationsniveau für eine bestimmte Jahrgangsstufe bestimmen, LehrerInnen können im Unterricht sowohl fachlich als auch pädagogisch adäquat agieren.
- *Kompetenzen als Wettbewerbsvorteile.* Hier erscheinen Kompetenzen in Abhängigkeit von einer (sozialen) Bezugsnorm: SchülerInnen mit einem höheren Bildungsabschluss weisen einen Wettbewerbsvorteil gegenüber SchülerInnen mit niedrigerem Bildungsabschluss auf, LehrerInnen mit Zusatzkompetenzen können MitbewerberInnen oder KonkurrentInnen übertreffen.

### Sozialwissenschaftlicher Kompetenzbegriff

Als einer der Begründer des sozialwissenschaftlichen Kompetenzbegriffs kann Chomsky (1968) gelten, der den Begriff im Kontext der Sprachwissenschaft prägte. Er wendet sich gegen die behavioristische Sprachauffassung mit ihrer ausschließlichen Anerkennung direkt beobachtbarer sprachlicher Daten, indem er ein sprachbezogenes kognitives Wissens- und Überzeugungssystem annimmt, welches sich in der frühen Kindheit entwickle und sich in der Interaktion mit der Umwelt ausgestalte. Als Terminus technicus für dieses System führt Chomsky den Begriff der „*linguistischen Kompetenz*“ ein.

In einer Erweiterung der „*sprachlichen*“ hin zur „*kommunikativen*“ Kompetenz definieren Schneewind und Pekrun (1994) letztere als „zentrales Sozialisationsziel“.

„Kommunikative Kompetenz“ bestimmt hier den adäquaten Einsatz sozial-kognitiver Regeln und Strukturen um in kommunikativen Situationen bestehen zu können.

Beide Kompetenzbegriffe unterscheiden zwischen Kompetenz und Performanz: „Kompetenz“ bezeichnet verhaltenserzeugende Fähigkeiten, „Performanz“ bezeichnet eben jenes durch diese Fähigkeiten erzeugte beobachtbare Verhalten. Shohamy (1996, S. 148) betont: „...competence equals ability equals trait, while performance refers to the actual execution of tasks.“

McClelland (1973) forderte in den 70er Jahren des letzten Jahrhunderts eine kompetenz- anstelle einer intelligenzorientierten Diagnostik, die im Gegensatz zur dekontextualisierten Intelligenzdiagnostik die Situations- und Kontextabhängigkeit menschlichen Handelns explizit berücksichtigen sollte. McClelland definiert „Kompetenzen“ als die für eine spezifische Tätigkeit notwendigen Voraussetzungen und bleibt damit bewusst unscharf.

Nach Gieseke (2006) gilt Heinrich Roth als maßgeblicher Begründer einer empirischen Erziehungswissenschaft. Roths (1966, 1971) Kompetenzkonzept versteht Kompetenzen als im Laufe des Bildungs- und Erziehungsprozesses erworbene Dispositionen, welche „die Bewältigung von unterschiedlichen Aufgaben bzw. Lebenssituationen ermöglichen. Sie umfassen Wissen und kognitive Fähigkeiten, Komponenten der Selbstregulation und sozial-kommunikative Fähigkeiten wie auch motivationale Orientierungen.“ (Klieme & Hartig, 2007, S. 21).

Ebenso wie Roth betont auch Bandura (1990) den Kontextbezug des Kompetenzbegriffs, indem er postuliert, dass Wissen und Fähigkeiten sich deutlich von der situationsspezifischen *Anwendung* des betreffenden Wissens und der Fähigkeiten unterscheiden, und auch Bledow und Frese (2009) fordern, dass Kompetenzen nicht vom jeweiligen Ausführungskontext abstrahiert werden sollten, sondern stets in einem konkreten Kontext zur Anwendung kommen.

Auch Weinert (2001, S. 27) stellt die Situationsabhängigkeit des Begriffs heraus: Er definiert „Kompetenzen“ als „die bei Individuen verfügbaren oder durch sie erlernbaren Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und

verantwortungsvoll nutzen zu können.“ Diese Definition des Kompetenzbegriffs liegt der Expertise zu den nationalen Bildungsstandards zugrunde (Klieme et al., 2003).

Zusätzlich unterscheidet Weinert (ebd.) zwischen spezifischen, kontextabhängigen Kompetenzen auf der einen Seite und Schlüssel- (etwa Sprache) bzw. Metakompetenzen (etwa Lern- oder Steuerungsstrategien) mit hoher Generalisierbarkeit auf der anderen Seite (vgl. die Lernstrategieklassifikation von Weinstein und Mayer, 1986).

Aufgrund ihrer Kontextdependenz lassen sich Kompetenzen durch *Lernen* erwerben und durch Erfahrung bis zur Expertise ausbauen. Hierdurch unterscheiden sich Kompetenzen von Fähigkeiten und Dispositionen. Berufliche Kompetenzen lassen sich folglich nur im jeweiligen Aufgaben- und Anforderungskontext definieren und spiegeln somit die „Handlungsanforderungen in einer Domäne“ (Schaper, 2009, S. 172) wider (zu erfolgreichem Handeln im Unterricht siehe Kap. 1.4).

Gemein ist sämtlichen Definitionen die Betonung der Beziehung „Kompetenz – Kontext“, und so seien zusammenfassend Klieme und Hartig (2007) zitiert:

*Kompetenzen werden in dieser psychologischen Tradition als erlernbare kontextspezifische Leistungsdispositionen verstanden, die sich funktional auf Situationen und Anforderungen in bestimmten Domänen beziehen. Die Breite dieser Domänen bzw. der relevanten Situationen kann zwischen spezifischen Kompetenzen und Schlüsselkompetenzen variieren, aber grundlegend sind ein Kontextbezug und die Erlernbarkeit. (S. 17/18)*

In der vorliegenden Arbeit wird implizit diesem Kompetenzbegriff gefolgt, indem „Kompetenzen als erlernbare kontextspezifische Leistungsdispositionen verstanden“ werden. Auf der Folie der Lehrerbildung bedeutet dies, dass professionelle pädagogische Kompetenzen erlernbar im Rahmen des Lehramtsstudiums und seiner Praxisphasen sind und durch die spezifischen Anforderungen des Lehrberufs definiert werden. Diese Anforderungen werden durch die Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004) expliziert.

## 1.2 Modellierung der professionellen pädagogischen Kompetenz

Zur Beschreibung und Messung von Kompetenzen sind *Kompetenzmodelle* vonnöten, welche sich hinsichtlich der betrachteten Foci unterscheiden: Lassen sich verschiedene Kompetenzstufen unterscheiden? Welche Struktur weist Kompetenz auf? Wie entwickelt sich Kompetenz? Basiert ein Kompetenzmodell auf empirischen Erkenntnissen oder normativen Vorgaben? Bezieht sich ein Modell auf allgemeine oder spezifische Kompetenzfacetten? Der folgende Abschnitt stellt methodische Ansätze zur Klärung dieser Fragen vor und bezieht sich hierbei auf die Modellierung *professioneller pädagogischer Kompetenz*.

### Niveau-, Struktur- und Entwicklungsmodelle

*Kompetenzniveaumodelle* stellen unterschiedliche Ausprägungen von Kompetenzen in den Mittelpunkt des Interesses (Klieme & Hartig, 2006) und bewerten sowie klassifizieren diese Ausprägungen auf der Folie von Bildungsprozessen. Für jedes Niveau wird hierbei beschrieben, welche Aufgaben eines Kompetenztests auf einem bestimmten Kompetenzniveau gelöst werden können. Die Skala des Tests wird hierbei in verschiedene Kompetenzstufen (Niveaus) unterteilt.

Im Rahmen der Lehrerbildung werden Kompetenzniveaumodelle selten formuliert. Eine Ausnahme realisierten Blömeke, Kaiser und Lehmann (2008), indem sie auf der Basis der „Mathematics Teaching in the 21<sup>st</sup> Century“ (MT21)-Kompetenztests vier Kompetenzstufen fachbezogenen Wissens angehender MathematiklehrerInnen identifizierten.

*Kompetenzstrukturmodelle* ermöglichen eine umfassende Beschreibung und Messung von Kompetenzen in einer bestimmten Domäne (Klieme & Hartig, 2006). Hierbei wird untersucht, welche Teilkompetenzen innerhalb einer Domäne bzw. innerhalb bestimmter Situationen und Kontexte erforderlich sind, wobei davon ausgegangen wird, dass die einzelnen Teilkompetenzen miteinander korreliert sind.

Für den Bereich der Lehrerbildung liegen zahlreiche Kompetenzstrukturmodelle vor. Als Beispiel sei der Ansatz von Kunter, Baumert, Blum, Klusmann, Krauss und Neubrand (COACTIV, 2011, S. 32/33) vorgestellt: Professionelle pädagogische Kompetenz besteht den Autoren zufolge aus spezifischem, erfahrungsbasiertem deklarativem und prozeduralem Wissen, professionellen Werten, Überzeugungen,



subjektiven Theorien, normativen Präferenzen und Zielen, motivationalen Orientierungen sowie Fähigkeiten der Selbstregulation. Die Wissensanteile subsumieren Kunter et al. unter dem Begriff des „Professionswissens“, welches sich in die Kompetenzbereiche „Fachwissen“, „Fachdidaktisches Wissen“, „Pädagogisch-psychologisches Wissen“, „Organisationswissen“ und „Beratungswissen“ aufgliedert. Diese Bereiche werden wiederum in Kompetenzfacetten unterteilt: Tiefes Verständnis der Schulmathematik (Fachwissen); Erklärungswissen, Wissen über das mathematische Denken von SchülerInnen sowie Wissen über mathematische Aufgaben (Fachdidaktisches Wissen) und zuletzt Wissen um Leistungsbeurteilung, Wissen über Lernprozesse und Wissen über effektive Klassenführung (Pädagogisch-psychologisches Wissen). Die Bereiche Organisationswissen und Beratungswissen werden im Modell nicht weiter spezifiziert, da aus methodischen Gründen die Validierung eines entsprechenden Messinstrumentes nicht realisierbar war. In einem multimethodalen (Wissenstests, Fragebögen, Vignettentests), multiperspektivischen (LehrerIn – Unterricht – SchülerInnen) sowie multikriterialen (Unterrichtsqualität, Freude an Mathematik, Leistungsängstlichkeit,...) Vorgehen konnte die Struktur des Modells weitgehend bestätigt werden (Baumert et al., 2011).

Neben solcherlei mehr oder weniger breiten Kompetenzstrukturmodellen werden auch Teilfacetten professioneller pädagogischer Kompetenz hinsichtlich ihrer Unterstrukturen untersucht. Schmelzing, Fuchs, Wüsten, Sandmann und Neuhaus (2009) analysierten etwa das Reflexionswissen von Biologielehrkräften – diese Facette der Handlungskompetenz von Lehrkräften wird von den Autoren in die Unterfacetten „Unterrichtswahrnehmung & Aufmerksamkeitsfokus“, „Fachdidaktisches Wissen: intuitiv-alltagssprachlich“, „Fachdidaktisches Wissen: theoretisch-fachsprachlich“ sowie „Maßstäbe zur Beurteilung von Unterricht“ untergliedert.

*Kompetenzentwicklungsmodelle* können die Frage beantworten, in welcher Reihenfolge (Teil-) Kompetenzen erworben werden und liefern somit Informationen darüber, welche Kompetenzen auf welcher Entwicklungsstufe lehr- und erwartbar sind. Sofern die Stufen von Kompetenzniveaumodellen als „Entwicklungsstufen“ interpretiert werden können, stellen Niveaumodelle Vorlagen für Entwicklungsmodelle dar.

Für den Bereich der Lehrerbildungsforschung liegen nach Schaper (2009) bislang keine Kompetenzentwicklungsmodelle vor.

### Empirische vs. normative Modellierung

Kompetenzmodelle lassen sich zum Einen aufgrund empirischer Analysen, zum Anderen auf normativer Basis entwickeln (Schaper, 2009). Erstgenannter Ansatz entspricht einer induktiven Vorgehensweise, bei der von empirischen Beobachtungen auf allgemeine Sachverhalte geschlossen wird. So werden hier beispielsweise Anforderungsanalysen in einer bestimmten Domäne aufgrund von Beobachtungsdaten erstellt, die wiederum Rückschlüsse auf benötigte Kompetenzen erlauben. Der zweitgenannte Ansatz lässt sich mit einer deduktiven Strategie vergleichen, wobei hier vom Allgemeinen auf Einzelereignisse geschlussfolgert wird. Solcherlei allgemeine Aussagen können fachspezifische Curricula, lerntheoretische Modelle oder normative Bildungsziele wie etwa die Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004) sein.

Schaper (ebd.) schlägt einen integrativen Zugang bei der Kompetenzmodellierung vor, da sowohl der induktive als auch der deduktive Ansatz Schwächen aufweisen, die durch eine Kombination der beiden Vorgehensweisen kompensiert werden können: Würden Kompetenzmodelle lediglich aufgrund empirischer Analysen formuliert, so geriete aus dem Blickfeld, dass Bildungsziele normativ formuliert werden. Deduktiv orientierte Ansätze gewährleisten die theoretische Fundierung einer Modellannahme und systematisieren empirisch angefallene Daten – eine rein deduktive Vorgehensweise läuft allerdings Gefahr, an realen Anforderungen vorbeizugehen. Eine Kombination beider Vorgehensweisen könnte folglich nach Schaper (ebd.) eine theoretisch fundierte Kompetenzmodellierung auf Basis normativer Bildungsziele realisieren.

Im Rahmen der Lehrerbildungsforschung wurden Kompetenzmodelle sowohl auf induktiver, deduktiver als auch auf integrativer Basis entwickelt. Ein Beispiel für einen integrativen Ansatz stellen die von Oser und Oelkers (2001) verfassten Standards für die Lehrerbildung dar. Diese Standards bilden den Ausgangspunkt für die Entwicklung und Diskussion von Standards der Lehrerbildung im deutschsprachigen Raum:

Die Autoren formulierten 88 Standards für Lehrpersonen, welche den vier Kriterien „Theoriebezogenheit“, „Empirie“, „Qualität“ sowie „Praxistradition“ genügen mussten. Die Autoren bemühen sich somit um die Vernetzung induktiver („Empirie“) sowie deduktiver („Theoriebezogenheit“, „Praxistradition“) Aspekte. Als Ergebnis resultierten 88 Standards in 12 thematischen Gruppen: (1) LehrerIn-SchülerIn-Beziehung; (2) SchülerInnenunterstützendes Handeln; (3) Disziplinprobleme und SchülerInnenrisiken; (4) Aufbau sozialen Verhaltens; (5) Lernprozessbegleitung/Lernstrategien; (6) Gestaltung und Methoden des Unterrichts; (7) Evaluation und Leistungsmessung; (8) Medieneinsatz; (9) Team-Arbeit der Lehrerschaft; (10) Öffentlichkeitsarbeit; (11) Kräftehaushalt der Lehrperson sowie (12) Fachdidaktik.

Nach Schaper (2009) besteht hinsichtlich solcher integrativer Ansätze jedoch weiterhin Bedarf. Zusätzlich plädiert der Autor für einen multimethodalen Ansatz, im Zuge dessen quantitative (z. B. Ratings) und qualitative (z. B. ExpertInnendiskussionen) Verfahren kombiniert werden sollen, um zu einer möglichst erschöpfenden Erfassung relevanter Kompetenzfacetten zu gelangen.

#### Allgemeine vs. spezifische Kompetenzmodelle

Der Begriff der Kompetenz bezieht sich stets auf eine Korrespondenz zwischen Kontext und Fähigkeit/Fertigkeit, wobei beide Faktoren als veränderlich angesehen werden. Meist werden hierbei die vier Kompetenzklassen der Fach-, Methoden-, Sozial- und Personalkompetenz unterschieden (z. B. Terhart, 2007).

Hinsichtlich der Modellierung professioneller Kompetenz im Lehrberuf können zwei Vorgehensweisen gewählt werden: Allgemeine Kompetenzmodelle können auf den Lehrberuf übertragen werden, oder aber es können spezifische Modelle für den Lehrberuf bzw. differenziert für bestimmte Fachbereiche entwickelt werden.

Ein *allgemeines* Kompetenzmodell stellte Heinrich Roth (1966, 1971) zunächst für die Berufspädagogik: Roth identifizierte die Facetten der *Selbst-, Sach- und Methodenkompetenz*. Übertragen auf den Lehrberuf bezeichnet „Selbstkompetenz“ die Selbstregulations- Reflektions- und Kritikfähigkeit, „Sachkompetenz“ bezieht sich auf Fachwissen und „Methodenkompetenz“ benennt die professionelle pädagogische Kompetenz im Sinne einer Kompetenz zur effektiven Unterrichtsgestaltung.

Als exemplarisch für ein *spezifisches* Kompetenzmodell kann erneut das COACTIV-Modell (Krauss et al., 2004) gelten, welches spezifisch für die Erfassung der Kompetenz von MathematiklehrerInnen entwickelt wurde.

Terhart (2007) favorisiert für den Zweck der Entwicklung eines Instrumentes der Erfassung und Bewertung professioneller Kompetenz im Lehrberuf den Ansatz allgemeiner Kompetenzmodelle, „denn am Ende soll eine Beurteilung des Entwicklungsstandes der beruflichen Kompetenzen insgesamt stehen, nicht eine Beurteilung spezieller Teilkompetenzen für spezifische Gruppen von Lehrern.“ (ebd., S. 48).

Nach Terhart (2007) sind zur Erfassung professioneller Kompetenzen im Lehrberuf Einschätzungen der Vorliegens von Kompetenzen von „nicht vorhanden“ bis „voll ausgeprägt“ zwar in der LehrerInnenbeurteilung im Rahmen des Referendariats noch vorhanden, werden jedoch nicht mehr als hinreichend angesehen. Vielmehr seien Stufenmodelle heute „lege artis“, wobei die Entwicklung dieser Kompetenzstufen in einer Kombination aus normativer Konstruktion und empirischer Prüfung vor sich gehen sollte.

### **1.3 Die Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004) als „Modell“ der professionellen pädagogischen Kompetenz**

Die Standards für die Lehrerbildung wurden nicht als Modell der professionellen pädagogischen Kompetenz formuliert. Dennoch können sie als einem integrativen Kompetenzstrukturmodell ähnlich interpretiert werden, da sie zum Einen erforderliche Kompetenzen beschreiben sowie Kompetenzfacetten definieren (Kompetenzstrukturmodell) und zum Anderen sowohl auf empirischer Basis wissenschaftlicher Theorien formuliert wurden als auch normativ Vorgaben für den „Output“ der Lehrerbildung setzen (integrativer Ansatz; Terhart, 2005).

Auch Helmke betont den engen Bezug von „Kompetenz“ und „Standard“, wenn er konstatiert (2009, S. 144): „Standards sind nichts anderes als erwartete Kompetenzen.“ Standards stellen dabei den Maßstab bereit, anhand dessen der Ausprägungsgrad von Kompetenzen gemessen werden kann (Terhart, ebd.).

Standards für die Lehrerbildung: Bildungswissenschaften beziehen sich auf diejenigen wissenschaftlichen Disziplinen, welche sich mit Bildungs- und Erziehungsprozessen, Bildungssystemen sowie den entsprechenden Rahmenbedingungen für das LehrerInnenhandeln befassen. Das Sekretariat der ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland definiert:

Standards in der Lehrerbildung beschreiben Anforderungen an das Handeln von Lehrkräften. Sie beziehen sich auf Kompetenzen und somit auf Fähigkeiten, Fertigkeiten und Einstellungen, über die eine Lehrkraft zur Bewältigung der beruflichen Anforderungen verfügt. Aus den angestrebten Kompetenzen ergeben sich Anforderungen für die gesamte Ausbildung und die Berufspraxis. (KMK, 2004, S. 4)

Auch Oser (2004, S. 193) definiert ähnlich:

Ein Standard ist ein professionelles Kompetenzprofil, mit dem pädagogische Situationen zweckmäßig und sinnvoll bearbeitet und bewältigt werden können. Mit dem Begriff ‚Kompetenzprofil‘ ist nicht eine einzelne Kompetenz gemeint, sondern quasi eine Kompetenzgruppe, die durch die Situation bestimmt wird, deren einzelne Teile aber ohne das Ganze sinnlos oder unzweckmäßig sind.

Beide Definitionen nehmen Bezug auf die Situations- und Kontextabhängigkeit des Kompetenzbegriffs (siehe Kapitel 1.1) und bestimmen die Standards in der Lehrerbildung als Menge von Kompetenzen (sowie Einstellungen) zur Bewältigung beruflicher Anforderungen.

Die Grundlage der Standards für die Lehrerbildung: Bildungswissenschaften  
Die Standards für die Lehrerbildung: Bildungswissenschaften orientieren sich an den jeweiligen Schulgesetzen der Länder sowie dem Berufsbild der Lehrperson. Der Präsident der Kultusministerkonferenz sowie die Bildungs- und Lehrgewerkschaften (2000) formulierten ein solches Berufsbild mit dem Ziel, die qualitative Weiterentwicklung des Bildungssystems sowohl durch das Kultusministerium als auch durch die Lehrpersonen zu gestalten. Folgende Aspekte der Berufsstandsdefinition wurden festgehalten:

1. LehrerInnen müssen sich zukünftigen Anforderungen stellen, die sich beispielsweise durch den gesellschaftlichen Wandel zur Wissensgesellschaft, durch die Globalisierung oder „die Weiterentwicklung der demokratischen Kultur“ (ebd., S. 2) ergeben.
2. Lehrerinnen und Lehrer sind ExpertInnen des Lernens und gestalten Lehr- und Lernprozesse nach wissenschaftlichen Erkenntnissen, vermitteln Lernmethoden und evaluieren ihren Unterricht. Neben fachwissenschaftlichen Kenntnissen weisen Lehrkräfte außerdem pädagogische, psychologische, kommunikative und soziale Kompetenzen auf.
3. LehrerInnen handeln glaubwürdig und arbeiten mit ihrem Kollegium zusammen.
4. LehrerInnen sind sich ihrer Vorbildfunktion bewusst und handeln dementsprechend, um die Persönlichkeitsentwicklung ihrer SchülerInnen positiv zu beeinflussen. Sie kooperieren mit Eltern und Fachkräften auch außerhalb der Schule.
5. LehrerInnen beurteilen kompetent und gerecht. Sie fördern Leistungsfähigkeit und Anstrengungsbereitschaft der SchülerInnen durch adäquate Leistungsrückmeldungen.
6. LehrerInnen aktualisieren ihr Wissen ständig und nehmen Fortbildungsangebote wahr. Zudem pflegen sie Kontakte zur Berufswelt.
7. LehrerInnen beteiligen sich aktiv an der Schulentwicklung, arbeiten hierbei mit ExpertInnen zusammen, beraten Eltern und SchülerInnen und gestalten ein motivierendes Schulklima.
8. LehrerInnen „unterstützen die interne und externe Evaluation der Lehr- und Lernprozesse, der Gestaltung des Schulprogramms und des Schullebens.“ (ebd., S. 4) Sie können Evaluationsinstrumente fachgerecht einsetzen und die Ergebnisse der Evaluationen korrekt interpretieren. Wissenschaftlich fundierte Leistungsvergleiche dienen hier der Qualitätsanalyse.
9. LehrerInnen sollen durch die Öffentlichkeit bei der Ausübung ihrer verantwortungsvollen Tätigkeit unterstützt werden, wobei sich Bildungspolitik und –verwaltung verpflichten, adäquate Rahmenbedingungen zu schaffen.

## Die Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004)

Auf Basis dieses Berufsbildes wurden vier Kompetenzbereiche vorgegeben, welche als normative Zielvorgaben des Lehramtsstudiums benennen, über welches pädagogische Wissen und Können ein Absolvent des Lehramtsstudiums verfügen muss. Diese vier übergeordneten Kompetenzbereiche werden durch Standards sowohl für die theoretischen als auch für die praktischen Ausbildungsabschnitte konkretisiert (KMK, 2004, S. 7-13; die vollständigen Standards siehe Anhang C).

Den vier Kompetenzbereichen (1) „Unterrichten. Lehrerinnen und Lehrer sind Fachleute für das Lehren und Lernen.“, (2) „Erziehen. Lehrerinnen und Lehrer üben ihre Erziehungsaufgabe aus.“ (3) „Beurteilen. Lehrerinnen und Lehrer üben ihre Beurteilungsaufgabe gerecht und verantwortungsbewusst aus.“ sowie (4) „Innovieren. Lehrerinnen und Lehrer entwickeln ihre Kompetenzen ständig weiter.“ werden 11 „Unterkompetenzen“ zugeordnet:

### Kompetenzbereich (1) „Unterrichten“:

- sach- und fachgerechte Planung und Durchführung von Unterricht
- Gestaltung von förderlichen Lernsituationen, Motivation, Bereitstellen von Transferoptionen,
- Förderung des selbstbestimmten Lernens und Arbeitens der SchülerInnen,

### Kompetenzbereich (2) „Erziehen“:

- Wissen um soziale und kulturelle Lebensbedingungen von SchülerInnen bzw. Einflussnahme auf deren individuelle Entwicklung
- Vermittlung von Werten und Normen
- Unterstützung selbstbestimmten Urteilens und Handelns durch die SchülerInnen
- Lösung von Schwierigkeiten und Konflikten in Schule und Unterricht

### Kompetenzbereich (3) „Beurteilen“:

- Diagnostik von Lernvoraussetzungen und –prozessen
- gezielte Förderung sowie Beratung von SchülerInnen und Eltern,
- Leistungserfassung aufgrund transparenter Beurteilungsmaßstäbe

Kompetenzbereich (4) „Innovieren“:

- Bewusstsein über Anforderungen des Lehrberufs sowie dessen besondere Verantwortung und Verpflichtung
- Verständnis des Lehrberufs als ständige Lernaufgabe
- Beteiligung an Planung und Umsetzung schulischer Projekte und Vorhaben.

Zwar sind die Standards für die Lehrerbildung: Bildungswissenschaften als *Kriterienkatalog* bezüglich des „Outputs“ der Lehrerbildung formuliert. Dennoch können sie als einem integrativen Kompetenzstrukturmodell ähnlich interpretiert werden, da sie zum Einen erforderliche Kompetenzen beschreiben sowie Kompetenzfacetten definieren (Kompetenzstrukturmodell) und zum Anderen sowohl auf empirischer Basis wissenschaftlicher Theorien formuliert wurden als auch normativ Vorgaben für den „Output“ der Lehrerbildung setzen (integrativer Ansatz).

#### **1.4 Die empirische und theoretische Basis der Standards für die Lehrerbildung: Bildungswissenschaften**

Die Standards für die Lehrerbildung: Bildungswissenschaften wurden auf Basis von Modellen der professionellen pädagogischen Kompetenz entwickelt. Die empirischen Befundlagen zu den vier Kompetenzbereichen „Unterrichten“, „Erziehen“, „Beurteilen“ sowie „Innovieren“ sollen im Folgenden dargestellt werden. Die Auswahl der vorgestellten Modelle ist dabei beispielhaft zu verstehen und erhebt keinen Anspruch auf Vollständigkeit.

Unterrichten. Lehrerinnen und Lehrer sind Fachleute für das Lehren und Lernen.

Darüber, dass das Unterrichten die Kernaufgabe von Lehrpersonen sei, besteht in der scientific community Konsens (z. B. Baumert, Kunter, Blum, Klusmann, Krauss & Neubrand, 2011; Helmke, 2009; Terhart, 2000). Die Ansicht darüber, was „guten“ Unterricht auszeichnen solle, hat sich allerdings in den letzten Jahrzehnten stark gewandelt:

Das in den Fünfziger- und Sechzigerjahren des vorherigen Jahrhunderts vorherrschende Paradigma der *Lehrerpersönlichkeit* versuchte, Unterschiede in der erzieherischen Wirkung auf Personenmerkmale des/der Lehrers/in zurückzuführen



(Bromme, 1997). Die Ergebnisse dieser Forschung waren jedoch wenig ergiebig, da nach Mayr und Neuweg (2006) sowie Bromme (ebd.) ungenügende psychometrische Verfahren verwandt wurden, empirisch begründete Theorien zu Effekten von Persönlichkeitsmerkmalen fehlten sowie lediglich triviale Befunde (etwa, dass sich Grausamkeit des/der Lehrers/in ungünstig auf die SchülerInnen auswirke; Gudjons, 2000) resultierten.

Unter dem Einfluss des Behaviorismus in den Siebzigerjahren des letzten Jahrhunderts wurde diese personenzentrierte Sicht vom *Prozess-Produkt-Paradigma* abgelöst: Hier wurde die Wirkung eng umschriebener einzelner Lehrerfertigkeiten auf ebenso eng umschriebene Schülerleistungen untersucht (z.B. Brophy & Good, 1986). Diese Lehrerfertigkeiten wurden zunächst als unabhängig von der spezifischen Unterrichtssituation, folglich als losgelöst vom Kontext, erachtet – eine Auffassung, die allerdings erweitert werden musste, da effektives Lehren die Organisation einer großen Anzahl von Einflussfaktoren beinhalte, wobei das Lehrerverhalten sich stets ändernden Umständen anpassen müsse (Bennett, 1978, zit. nach Bromme, 1997, S. 185).

Den Paradigmenwandel der Lehr-Lernforschung zum *Expertenparadigma* innerhalb der 80er und 90er Jahre des letzten Jahrhunderts formuliert Bromme (1997) als Verschiebung des Focus´ von der Persönlichkeit des/der Lehrers/in zur Focussierung auf effektive Verhaltensweisen und zuletzt zur Sichtweise der Lehrperson als Experte/in mit begrenztem Einfluss. Dieser Perspektivenwandel impliziert, dass der direkte Einfluss des/der Lehrers/in auf die Lernergebnisse seiner/ihrer SchülerInnen weit geringer sei als bis dahin angenommen wurde, und somit änderte sich folglich ebenso das Anforderungsprofil des Lehrberufs: 1989 postuliert Slavin, dass die Aufgabe eines/einer Lehrers/in nicht das „Lehren“, sondern das Zur-Verfügung-Stellen von *Lerngelegenheiten* sei – eine Sichtweise, die eine deutliche Parallele aufweist zum Konstruktivismus mit dessen Verständnis des Lernens als aktive Konstruktion des Lernenden. Ein/e Lehrer/in wird folglich als Experte/in für die Gestaltung von Lerngelegenheiten gesehen. Dieser Auffassung folgend, ist es nach Terhart (2000) „realistisch und geboten, die Organisation von Lehren und Lernen als den Kernbereich des professionellen Mandats der Lehrerschaft zu betrachten.“ (S. 78)

Somit überrascht es nicht, dass auch im Rahmen der Forschung zu professioneller pädagogischer Kompetenz das Unterrichten zumeist im Focus des Interesses steht.

Um analysieren zu können, welche Aspekte professioneller pädagogischer Kompetenz effektiven Unterricht im Sinne der Beeinflussung von Schülerleistungen tatsächlich auszeichnen, ist empirische Unterrichtsforschung notwendig (Helmke, Helmke & Schrader, 2007). Auch an dieser Stelle sei auf die Komplexität der Lehr-Lernsituation hingewiesen: „Unterricht ist lediglich ein Angebot mit dem Ziel, aktiv zu lernen. Ob, wie gut und wie nachhaltig dieses Angebot schülerseits genutzt wird, hängt von einer Vielzahl unterrichtlicher, kontextueller und individueller Merkmale ab.“ (ebd., S. 54) Zudem erfordern unterschiedliche Lernziele unterschiedliche Lehr-Lern-Methoden – *den* guten Unterricht oder *die* gute Lehrmethode kann es demzufolge nicht geben. Aufgrund der Vielfalt möglicher Unterrichtsziele kann sich die Unterrichtsforschung nicht auf ein einziges Kriterium (etwa die Schülerleistung) beschränken, sondern sollte verschiedene Kriterien (beispielsweise „Leistungssteigerung“ und „Chancenausgleich“) innerhalb von Klassen kombinieren (ebd.).

Wie können Determinanten effektiven Unterrichts aufgedeckt oder Modelle der professionellen pädagogischen Kompetenz empirisch überprüft werden?

Die beiden grundlegenden Strategien der Unterrichtsforschung analysieren entweder Zusammenhänge zwischen Input- und Output-Variablen oder focussieren bestimmte Personengruppen:

a) Einfache *Korrelationsstudien* basieren auf dem Prozess-Produkt-Modell, welches Korrelationen zwischen Unterrichtsmerkmalen und Unterrichtsprodukten interpretiert. Diese Studien gelangen jedoch aufgrund des stark vereinfachenden Charakters des Modells an ihre Grenzen. In der Folge der großen Leistungsstudien wie etwa PISA oder TIMSS haben sich inzwischen jedoch methodische Standards durchgesetzt, die allgemein als „lege artis“ gelten (ebd.):

1. Die Modellierung mithilfe von Strukturgleichungsmodellen löst die Auswertung von Daten auf Basis einfacher Korrelationen ab.
2. Die anfallenden Daten sollten Mehrebenenanalysen (etwa auf Schüler- und Klassenebene) unterzogen werden.
3. Ein multimethodaler (Fragebogen- und Beobachtungsverfahren) sowie multiperspektivischer (z. B. SchülerInnen- vs. LehrerInnensicht) Zugang ist wünschenswert.

4. Theoretisch fundierte Konstrukte müssen in statistisch testbare Modelle überführt werden. Hierbei ergibt sich allerdings größtenteils das Problem der Multikollinearität der untersuchten Variablen.

b) *Expertenstudien* analysieren bestimmte Personengruppen und identifizieren im Kontext der Unterrichtsforschung beispielsweise diejenigen Lehrkräfte, deren SchülerInnen im Vergleich zu Lehrkräften mit gleichen Eingangsvoraussetzungen einen besonders hohen Leistungszuwachs aufweisen. Anschließend werden die Unterrichtsmerkmale des/der betreffenden Lehrers/in analysiert. Die Profile erfolgreicher LehrerInnen können somit erstellt und interpretiert werden, sofern eine repräsentative Vergleichsstichprobe von Lehrkräften vorliegt.

Durchgängig steht die pädagogisch-psychologische, empirische Unterrichtsforschung den eher wenig empirisch fundierten Allgemeinen Didaktiken oder Fachdidaktiken gegenüber – eine Kluft, die sich beispielsweise im Beibehalten schulspezifischer Maßnahmen (etwa Sitzenbleiben; Arnold, Koch-Priewe & Lin-Klitzing, 2007) ohne ausreichende empirische Grundlage ebenso spiegelt wie in der - zwar angegangenen, doch noch immer beklagten - Diskrepanz zwischen Theorie und Praxis des Lehramtsstudiums (zum direkten Transfer theoretischer Studieninhalte in die Lehrpraxis siehe z. B. Schocker-von Dittfurth, 2001).

Zu welchen Ergebnissen gelangt die aktuelle Unterrichtsforschung?

Hinsichtlich zentraler Bestimmungsstücke lernförderlichen Unterrichts besteht weitgehend Konsens: Relevant für die Förderung verständnisvollen Lernens sind insbesondere die *Bereitstellung von kognitiv anspruchsvollen Lerngelegenheiten, regelmäßige Lernkontrolle, individuelles Feedback, an individuelle Bedürfnisse und Rahmenbedingungen angepasster Unterricht* sowie *effizientes Zeit- und Klassenmanagement* (Kunter et al., 2011; z. B. Helmke, 2009; Brophy, 2000; Gudjons, 1998).

Das einflussreiche Modell professioneller Kompetenz der COACTIV-Forschungsgruppe (Baumert & Kunter, 2011) bezieht sich auf die von der KMK (ebd.) geforderte „sach- und fachgerechte Planung und Durchführung von Unterricht“, indem unter dem Aspekt „Professionswissen“ die an Shulman (1986) angelehnten Wissensarten „Fachwissen“, „Fachdidaktisches Wissen“ sowie „pädagogisch-psychologisches Wissen“ (neben Organisations- und

Beratungswissen) subsumiert werden. Die Autoren formulieren ihr Modell lediglich für den Bereich der Mathematik, da sie auf Basis empirischer Evidenz (z. B. TIMSS, Klieme, Schümer & Knoll, 2001) davon ausgehen, dass das Bereitstellen von Lerngelegenheiten bei ähnlicher Strukturierung des Unterrichts bzw. Methoden- und Inhaltswahl sich insbesondere durch die gestellten *fachspezifischen Aufgaben* zeigt. Baumert und Kunter gelangen somit zu dem Schluß, dass das Bereitstellen von Lerngelegenheiten nicht unabhängig vom Inhaltsbereich zu analysieren sei.

*Fachwissen* wird demnach von den Autoren als Kompetenzfacette „tiefes Verständnis der Schulmathematik“ konkretisiert und bildet somit die notwendige fachliche Basis des Unterrichtens.

Baumert und Kunter (ebd.) untergliedern das *fachdidaktische Wissen* in die Facetten „Erklärungswissen“ – hier geht es um das Wissen über effektives Erklären und Darbieten von Inhalten –, „Wissen über das mathematische Denken von SchülerInnen“ – hier ist das Wissen um typische Fehlvorstellungen, Fehler oder Strategien ebenso angesprochen wie das Diagnostizieren von Wissen und Verständnisprozessen – sowie „Wissen über mathematische Aufgaben“ – mithin das Wissen darüber, wie mathematische Aufgaben und Inhaltsbereiche strukturiert werden sollten bzw. welche Wissensbasis sie erfordern.

Zuletzt wird das fachübergreifende *pädagogisch-psychologische Wissen* in die Facetten „Wissen um Leistungsbeurteilung“, „Wissen über Lernprozesse“ sowie „Wissen über effektive Klassenführung“ unterteilt und schließlich in der Studie COACTIV-R (Voss, Kunter & Baumert, 2009) konkretisiert zu Klassenführung (classroom management), Unterrichtsmethoden, Diagnostik, Lernprozesse und Heterogenität.

Voss, Kunter und Baumert (ebd.) fanden empirische Nachweise dafür, dass die Mathematikleistung signifikant durch kognitive Aktivierung und classroom management prognostiziert wird. Die von der KMK geforderte Motivierung (operationalisiert als Leistungsängstlichkeit und Freude an Mathematik) der SchülerInnen kann den Autoren zufolge insbesondere durch die individuelle Unterstützung durch die Lehrperson und durch die Klassenführung beeinflusst werden.

König und Blömeke definierten in der Studie „Teacher Education and Development Study- Learning fo Teach Mathematics“ (TEDS-M, 2009) ein fünfdimensionales Modell fachunabhängigen pädagogischen Wissens, welches dem COACTIV-Modell deutlich ähnelt: Die Autoren postulieren die Faktoren „Strukturierung von Unterricht“, „Motivierung“, „Umgang mit Heterogenität“, „Klassenführung“ sowie „Leistungsbeurteilung“, wobei letzterer in den Standards der KMK nicht unter „Unterrichten“ gefasst, sondern als eigener Kompetenzbereich definiert wird.

Blömeke, Felbrich und Müller (2008) identifizierten innerhalb der internationalen Studie „Mathematics Teaching in the 21<sup>st</sup> Century“ (MT21) fünf an Mathematiklehrkräfte gestellte Anforderungsfelder, von welchen sich der Aspekt „A: Auswahl von Unterrichtsinhalten & Methoden/Sequenzierung des Lernprozesses“ direkt auf das Unterrichten bezieht (zu den weiteren Anforderungsbereichen s.u.). Die Autoren untergliedern diesen Anforderungsbereich in die Facetten „Auswahl und Begründung von unterrichtlichen Inhalten“ sowie „Planung und Evaluierung von Unterrichtsstunden“.

Brophy (2000) gelangt in den „Educational Practices Series“ der International Academy of Education auf Basis der bedeutsamsten Erkenntnisse der Unterrichtsforschung zu folgenden Prinzipien des effektiven Lehrens: Unterstützendes Klassenklima, das Bereitstellen von Lerngelegenheiten, eine Orientierung am Lehrplan, Kohärenz der Inhalte, ein durchdachter Unterrichtsplan, Übung und Anwendung des Gelernten, Unterstützung der Lerntätigkeit, Lehren von Strategien, Nutzung von Formen des kooperativen Lernens, zielorientierte Beurteilung sowie klar kommunizierte Leistungserwartungen. Helmke (2004) definiert – ebenfalls unter Rückgriff auf die empirische Unterrichtsforschung einen fast deckungsgleichen Katalog, fügt allerdings in einer Überarbeitung (Helmke, 2009) die Facetten „*Kompetenzorientierung*“ sowie „*Umgang mit Heterogenität*“ hinzu. Auch Baumert (2003) nennt klar strukturierte Aufgabenstellungen und Erklärungen, ein angemessenes Tempo, gelungene Anpassung an individuelle Lernvoraussetzungen sowie die Gestaltung einer positiven sozialen Atmosphäre als relevant für „guten Unterricht“ und bezieht beim Aspekt der Unterrichtsführung explizit die Störungsprävention mit ein.

Zur theoriegeleiteten Analyse von Unterricht verwandten Prenzel et al. (2002) die Methode der videogestützten Unterrichtsbeobachtung, um komplexe Unterrichtsskripts – bestimmte kulturabhängige Unterrichtsmuster und

wechselseitige Lehr- und Lernprozesse – erfassen zu können. Die Autoren beziehen sich auf den Physikunterricht und nennen vier relevante Aspekte der Unterrichtsgestaltung, bezeichnen sie allerdings als „Problembereiche des deutschen Physikunterrichts“ (S. 143): (1) Zielorientierung, (2) Begleitung des Lernens und die Unterstützung zielgerichteter mentaler Aktivitäten, (3) Berücksichtigung von Schülervorstellungen und das Umgehen mit Fehlern oder Fehlvorstellungen, (4) Einsatz und Einbettung von Lehrer- oder Schülerexperimenten und Thematisierung naturwissenschaftlicher Prozesse. Die Zielorientierung wird dabei von Prenzel et al. als „Explikation und Transparenz von Lehrzielen als Basis für die Konstruktion von Lernzielen“ spezifiziert. Die Begleitung des Lernens beinhaltet die Diagnose, die Überwachung des Lernprozesses sowie Rückmeldung durch die Lehrkraft oder die Lernenden selbst im Sinne selbstgesteuerten Lernens. Der Faktor des Experimentierens bezieht sich auf Denk- und Arbeitsweisen bzw. Vorstellungen von den Besonderheiten der Naturwissenschaften und lässt sich somit nur auf naturwissenschaftliche Fächer übertragen. Die beiden erstgenannten Faktoren „Zielorientierung“ sowie „Begleitung des Lernens“ beziehen sich dabei direkt auf den Kompetenzbereich „Unterrichten“ der KMK (ebd.). Beide eignen sich empirischen Analysen zufolge als Prädiktoren des schülerseitigen Leistungszuwachses.

Schuler, Hell und Trapmann (2007) identifizierten im Rahmen einer Anforderungsanalyse für die Universität Hohenheim aufgrund von Literaturrecherchen sowie ExpertInnenurteilen 15 „Anforderungsdimensionen“ (vgl. MT21, Blömeke et al., 2008) – Anforderungen, welche an LehrerInnen gestellt werden – von denen sich fünf auf den Aspekt des Unterrichtens beziehen, nämlich das didaktische Wissen und Handeln, Planung und Organisation, Fachwissen, Fachinteresse und die praktische Anwendung fachlicher Inhalte.

Auch Bromme (1997) bedient sich des Konzepts der Identifikation von Anforderungen und unterscheidet drei Anforderungsebenen, die sich dem/der Lehrer/in stellen:

(1) Die Organisation und Aufrechterhaltung einer Struktur von LehrerInnen- und SchülerInnenaktivitäten, (2) die Entwicklung des Stoffes im Unterricht, (3) die Organisation der Unterrichtszeit sowie (4) professionelles Wissen.

Unter (1) werden beispielsweise sowohl die Störungsfreiheit, ein situationsangemessener Wechsel zwischen Arbeitsformen als auch ein angemessenes Anspruchsniveau subsummiert. (2) basiert auf der Überwindung der „Sender-Metapher“ mit dem/der Lehrer/in als Sender/in einer an die EmpfängerInnen (SchülerInnen) adaptierten Botschaft hin zur gemeinsamen, durch den/die Lehrer/in strukturierten Entwicklung des Stoffes. Bromme (ebd.) erwähnt hier insbesondere die thematische Gliederung des Unterrichts sowie das Lernen in Schülergruppen. Aufgrund der Vorstrukturierung des Unterrichts, etwa durch die Länge der Unterrichtsstunden oder das Curriculum, besteht (3) das „time-management“ (Bromme, 1997, S. 193) in der Gestaltung der zeitlichen Abfolge und Dauer einerseits der Stoffbehandlung und andererseits der Schülerbeiträge. (4) rekurriert auf Forschungsergebnisse beispielsweise von Carlsen (1987, zit. nach Bromme, 1997, S. 1995) oder Dobey und Schafer (1984), die eine größere Schülerbeteiligung bei größerem Wissen des/der Lehrers/in und eine weniger direkte Steuerung des Unterrichts durch LehrerInnen mit besserem Fachwissen beobachteten. Zudem werde die diagnostische Kompetenz von LehrerInnen durch das Wissen um Schülerfehlvorstellungen, Lernstrategien, Verständnisschwierigkeiten sowie die kognitive Entwicklung von SchülerInnen beeinflusst.

Nicht unerwähnt bleiben soll zuletzt das einflussreiche QuAIT-Modell von Slavin (1997): Der Autor identifiziert die Merkmale „Qualität der Instruktion – *Quality*“, „Angemessenheit der Instruktion – *Appropriateness*“ (beispielsweise im Hinblick auf den Entwicklungsstand der SchülerInnen), „Motivation – *Incentives*“ und „Zurverfügungstellen von Lernzeit – *Time*“ als relevant für die Unterrichtsqualität.

Die oben genannten Modelle und Merkmalskataloge professioneller pädagogischer Kompetenz basieren auf empirischen Erkenntnissen und theoretischen Annahmen. Bildungsziele werden jedoch *normativ* formuliert, und außerdem bestehen nach Terhart (2007) weiterhin normative, emotional geprägte und konkurrierende Idealvorstellungen darüber, was „guter Unterricht“ sei. Professionelle pädagogische Kompetenz bewegt sich also im Spannungsfeld zwischen empirischer und normativer Definition. Die normative Perspektive sollte folglich nicht außer Acht gelassen werden, da sie zum Einen Ziele für Schüler- und Lehrerbildung setzt, zum Anderen implizit oder explizit Lehrerhandeln beeinflusst. Eine didaktisch-normative Perspektive des professionellen pädagogischen Handelns nehmen beispielsweise Meyer (2004) oder das historische Modell von Leuchtenberger (1909) ein – ebenso

wie aus internationaler Sicht das einflussreiche National Board for Professional Teaching Standards (2009), da die dort genannten Standards und Qualitätsbereiche als normative Vorgaben gelten können:

Meyer (2004) postuliert aus didaktischer Sicht zehn Merkmale guten Unterrichts: Klare Strukturierung des Unterrichts, hoher Anteil echter Lernzeit, lernförderliches Klima, inhaltliche Klarheit, sinnstiftendes Kommunizieren, Methodenvielfalt, individuelle Förderung, intelligentes Üben, transparente Leistungserwartungen sowie eine vorbereitete Umgebung.

Aus historischer Perspektive seien einige Kriterien des „Vademecums für junge Lehrer“ (Leuchtenberger, 1909) zitiert. Ähnlich den Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004) wurde auch hier eine Idealnorm für KandidatInnen und jüngere LehrerInnen (in Form von pädagogischen und didaktischen Grundregeln) realisiert, wobei sich deutliche Entsprechungen zu aktuellen normativen und empirisch gewonnenen Modellen zeigen:

- Verknüpfung von neuem Stoff mit bereits Bekanntem
- Scaffolding
- Transfer
- Adäquate Nutzung von Medien
- Stellen von offenen Fragen
- Einbezug sämtlicher SchülerInnen
- Vorbildfunktion des/der Lehrers/in
- Vorurteilsfreie Förderung aller SchülerInnen ohne Blick auf sozioökonomischen Stand, Elternhaus oder persönliche Sympathie
- Positiver Umgang mit eigenen Fehlern
- Wecken von Interesse
- Selbstreflexion
- Unterlassen von Spott, Ironie, Moralpredigten
- Gründliche Vorbereitung von Klassenarbeiten zusammen mit den SchülerInnen

Zuletzt zeigt ein Blick in die USA zu den Standards des einflussreichen National Board for Professional Teaching Standards (2009) sehr detaillierte Standards für nahezu allen Fächer und Klassenstufen. Fachübergreifend werden fünf Kernstandards postuliert: (1) LehrerInnen fühlen sich gegenüber SchülerInnen und



deren Lernen verpflichtet; (2) LehrerInnen kennen die Inhalte ihrer Fächer und wissen, wie diese gelehrt werden; (3) LehrerInnen sind verantwortlich für die Anleitung und Überwachung des Lernens von SchülerInnen; (4) LehrerInnen reflektieren Ihre Arbeit systematisch und lernen aus ihrer Erfahrung sowie (5) LehrerInnen sind Mitglieder von Lerngemeinschaften.

Bei aller Kritik an der mangelnden empirischen und theoretischen Basis normativer Modelle des Unterrichtens sind doch die Parallelen zwischen empirischer Forschung und normativen Modellen augenfällig. Interdisziplinär, international sowie historisch lässt sich somit ein Konsens hinsichtlich Merkmalen effektiven Unterrichts finden. Hinsichtlich zentraler Bestimmungsstücke lernförderlichen Unterrichts besteht weitgehend Konsens: Die Planung und Durchführung von Unterricht auf der Basis fachlichen und fachdidaktischen Wissens bzw. das classroom management und der Umgang mit heterogenen Schülervoraussetzungen, Lernunterstützung und Motivierung werden von nahezu allen Modellen als relevant postuliert. Diese Merkmale stimmen mit den durch die KMK genannten Kompetenzfacetten der fach- und sachgerechten Planung und Durchführung von Unterricht, der adäquaten Gestaltung von Lernsituationen zur Unterstützung des Lernprozesses, der Motivation von SchülerInnen sowie der Förderung selbstbestimmten Lernens und Arbeitens überein, wenngleich auch die Abgrenzung der in den verschiedenen Modellen genannten Unterrichtsmerkmale von anderen KMK-Kompetenzbereichen nicht trennscharf gelingt.

Doch welche „Produkte“ im Rahmen des Produkt-Prozess-Paradigmas (etwa der Lernzuwachs auf Schülerseite) können durch welche Facetten des Unterrichtens überhaupt erwartet werden?

Nach Küstings, Billich und Lipowsky (2009) erklären Merkmale des Unterrichts, des/der Lehrers/in und der Klassenzusammensetzung zwischen 10 und 30 Prozent der Leistungsunterschiede von SchülerInnen auf, wobei SchülerInnen mit ungünstigen Lernvoraussetzungen besonders von einer Verbesserung der Unterrichtsqualität zu profitieren scheinen.

Die Befundlage hinsichtlich eines Effektes von Fachwissen auf Schülerleistungen zeigt sich Küstings, Billich und Lipowsky (ebd.) zufolge uneinheitlich. Das fachdidaktische Wissen leistet der COACTIV-Studie zufolge jedoch einen

substantiellen Beitrag zur Schülerleistung (Blum & Krauss, 2008). Zu ähnlichen Ergebnissen gelangen beispielsweise auch Hill, Rowan und Ball (2005).

Eine effektive, störungsminimierende Klassenführung hängt der Metaanalyse von Wang, Haertel und Walberg (1993) zufolge von allen untersuchten Unterrichtsmerkmalen am stärksten mit der Klassenleistung zusammen. Bezüglich der kognitiven Aktivierung der Lernenden, also einer elaborierten Auseinandersetzung der Lernenden mit dem Stoff, lassen sich aufgrund der dünnen Forschungslage bislang noch keine eindeutigen Aussagen treffen. Klieme, Schümer und Knoll (2001) zeigten jedoch beispielsweise positive Effekte zwischen kognitiv aktivierendem Unterricht und dem Lernerfolg der SchülerInnen in Mathematik.

Der Begriff des „*Unterrichtsklimas*“ wird nach Künsting, Billich und Lipowsky (2009) vielfältig operationalisiert. Wird eine positive Lehrer-Schüler-Interaktion (Ansprechbarkeit des/der Lehrers/in bei Problemen, Orientierung an kriterialen und individuellen Bezugsnormen,...) als Indikator für das Unterrichtsklima untersucht, so zeigen sich nach Grewe (2007, zit. nach Künsting, Billich & Lipowsky, 2009) die größten Zusammenhänge zwischen affektiven Verhaltensweisen der SchülerInnen und dem Unterrichtsklima. Bezüglich der Schülerleistungen lassen sich jedoch häufig keine Effekte des Unterrichtsklimas nachweisen, sofern für weitere Unterrichtsmerkmale kontrolliert wird.

Insgesamt zeigt sich die Forschungslage zwischen professioneller pädagogischer Kompetenz von Lehrpersonen und Schülerleistungen uneinheitlich (Künsting, Billich & Lipowsky, 2009), was jedoch beispielsweise aufgrund methodischer Mängel zahlreicher Untersuchungen (z. B. mangelnde Gütekriterien von Messinstrumenten) bzw. der komplexen Variablengefüge auf Lehrer-, Schüler-, Klassen- und Schulebene nicht verwundern muss (Wuttke, 2009).

Lernfortschritte von SchülerInnen eindeutig auf Grade der Lehrerkompetenz rückführen zu wollen, gestaltet sich insgesamt äußerst schwierig. Nach Terhart (2007) werden – je nach Studie – zwischen 3 und 23 Prozent der Varianz von Schülerleistungen durch den Faktor „LehrerIn“ aufgeklärt. Einer der aussagekräftigsten Faktoren für Schülerleistungen ist das Vorwissen der SchülerInnen. Da die Lehrperson nur eine von einer Vielzahl an Determinanten der Schülerleistung ist und sich zudem schwer von anderen Einflussgrößen isolieren lässt, schlägt Terhart (ebd.) vor, nicht die Schülerleistung, sondern das

Zurverfügungstellen von Lerngelegenheiten als Kriterium professioneller Kompetenz von Lehrpersonen zu verwenden.

Erziehen. Lehrerinnen und Lehrer üben ihre Erziehungsaufgabe aus.

Unter diesem Kompetenzbereich wird die Vermittlung von Werten und Normen, die Unterstützung selbstbestimmten Urteilens und Handelns von SchülerInnen ebenso zusammengefasst wie die Kenntnis über die sozialen und kulturellen Lebensbedingungen von SchülerInnen sowie die Einflussnahme auf deren individuelle Entwicklung im Rahmen der Schule. Zudem sollen LehrerInnen Lösungsansätze für Schwierigkeiten und Konflikte in Schule und Unterricht finden.

Der Wandel des Lehrerbilds vollzog sich nicht nur innerhalb der Beziehung „Lehrerhandeln – Schülerleistung“, sondern erweiterte den Lehrerauftrag zusätzlich um *sozialpädagogische Aufgabenstellungen* (Gudjons, 2000): So besteht etwa die Forderung, dass „Schule“ aufgrund einer abgeschwächten Erziehungsfunktion der Familie zunehmend Anteile der schulischen Unterrichts- und Wissensvermittlungsfunktion an eine schulische Erziehungs- und Beratungsfunktion abtreten solle (Gudjons, 2000; Terhart, 2000).

Nach Kunter et al. (2011) lassen sich „Erziehen“ und „Unterrichten“ als Kernaufgabe von Lehrpersonen nicht voneinander trennen. Die Autoren interpretieren lehrerseitige Erziehungsaufgaben als unmittelbar aus dem Unterrichten folgend: Zum Einen erziehe „Schule“ durch die kognitive Aktivierung während des Unterrichts (Lernen, Problemlösen, Leistungserbringung) bzw. durch die Setzung von Bewertungsmaßstäben. Gleichzeitig stelle schulischer Unterricht eine Erziehung zu reflexivem Denken sowie zum Begründen und Erklären bzw. zum ausdauernden Arbeiten und Üben dar. Zum Anderen erziehe Schule zur Einhaltung sozialer Regeln, indem beispielsweise bestimmte Umgangsformen, Pünktlichkeit, geordnete Unterrichtsführung oder das Verhalten innerhalb verschiedener Sozialformen eingefordert würden. Kunter et al. (ebd.) nennen als weitere Erziehungsinhalte Aufmerksamkeit, Anstrengung, Geduld, Leistungsmotivation, Zielorientierung, Belohnungsaufschub, Selbstregulation, Emotionskontrolle, soziale Rücksichtnahme, Hilfsbereitschaft, Verantwortungsübernahme, Kooperation sowie Konfliktbewältigung.

Die *Moralerziehung* stellt einen eher normativen Aspekt der Standards für die Lehrerbildung: Bildungswissenschaften dar, lässt sich allerdings auch aus

empirischer Werte als Dimension der von Blömeke, Felbrich und Müller (2008) im Rahmen der Studie „MT21“ identifizierten Anforderungen an Lehrkräfte finden („foster the development of morals and values“). Wird die schulische Erziehung im Sinne Baumerts und Kunters (ebd.) verstanden, so besteht die schulische Werteerziehung vor allem in der Erziehung zur Einhaltung sozialer Regeln bzw. zu arbeits- und leistungsbezogenen Fähigkeiten. Die Erziehung zur Findung und Begründung eigener Werte und Werturteile, wie von Rekus (2009) oder Weinert (2000) gefordert, tangieren die Autoren lediglich in den Aspekten des reflexiven Denkens sowie des Begründens und Erklärens.

Nach Rekus (ebd.) ist die Erziehung zur Herstellung selbstverantwortlichen Urteilens sowie Handelns zentral, wobei die Grundsätze a) der Diskussion konkreter, lebensnaher Beispiele (statt der Ausrichtung an abstrakten Themen, wie etwa Gesetzen), b) der Aufforderung zum selbstständigen Urteilen (statt der Vorgabe des „richtigen“ Urteils durch die Lehrkraft), c) die Bereitstellung von Möglichkeiten der Perspektivenübernahme (statt Verharren auf der eigenen Sichtweise) und d) der Unterstützung bei der Umsetzung von eigenen Werten in Handlungen Selbstverantwortlichkeit fördern sollen. Diese Grundsätze lassen sich Schirp (2003) zufolge im Rahmen der Schule durch die Etablierung (1) sozialer Strukturen und Modelle, von (2) wertereфлектиierenden Lerngemeinschaften, sowie durch die (3) Herstellung von Handlungsfreiräumen und Kooperationen erreichen. (1) bezieht sich auf die gemeinsame Schaffung sozialer Vereinbarungen – Gudjons (1998) bezeichnet dies als Besprechungen und Aufforderungen anstelle von Vorschriften – bzw. durch Gelegenheit zur selbstgesteuerten, auch klassenübergreifenden Arbeit. Die Installierung von (2) Lernergruppen, welche frei und ohne Bereitstellung der „Lösung“ durch die Lehrperson diskutieren können, fördert einen reflektiven Umgang mit Werten und Normen. Hier ist etwa die Arbeit mit moralischen Dilemmata (Kohlberg, 1995) angezeigt, die sich zwar auf die unmittelbaren Lebenswelten der SchülerInnen beziehen, jedoch nicht alleine im Schulalltag stehen bleiben sollten. (3) spricht insbesondere die Schaffung von „Helfer- und Unterstützungssystemen“ (Schirp, S. 14) an, beispielsweise im Rahmen von „peer mediations“ zur selbstständigen Lösung von Konflikten.

All diese Methoden finden natürlich im Rahmen des Unterrichts statt – insofern schließt sich der Kreis zu Baumert und Kunter (ebd.). Zusätzlich werden durch die genannten Aspekte jedoch Werte und Normen bzw. deren selbstgesteuerte Entwicklung explizit thematisiert.

Beim Umgang mit *Konflikten* ist es Imhof (1985) zufolge nötig, Konflikte zunächst zu erkennen und sie klar darstellen zu können. Anschließend müssen Lösungswege gesucht oder aber Unveränderlichkeiten akzeptiert werden. Der offene Umgang mit Konflikten ist also ausschlaggebend (Busch & Todt, 2006), und zusätzlich ist nach Oser (2006) bei der Lösung von Konflikten die Übernahme verschiedener Perspektiven besonders relevant.

Eine typische Form von Konflikten und Gewalt an Schulen ist das „Bullying“ – aggressives Verhalten eines oder mehrerer Schüler/s gegen ein hilfloses Opfer über einen längeren Zeitraum hinweg (Scheithauer & Bull, 2009). Das Modell des Bullying von Olweus (2002) beschreibt zunächst auf Seiten der gewaltbereiten SchülerInnen („Bullys“) Persönlichkeitsmerkmale und den familiären Kontext als relevant. Verstärkendes Verhalten auf Seiten des Opfers, der MitschülerInnen sowie von LehrerInnen und Eltern führe dann zu einem „Teufelskreis des Bullying“: Opfer reagieren meist passiv auf die Übergriffe und zeigen mangelnde Problemlösefertigkeiten, indem sie beispielsweise gegenüber den Lehrpersonen sowie Eltern schweigen. MitschülerInnen reagieren entweder nicht oder aber unterstützen die Bullys aktiv, wobei Lehrkräfte sich inkonsequent verhalten oder die Vorfälle ignorieren, und Eltern können schließlich aufgrund mangelnder Information nicht eingreifen.

Auf Schülerseite wird Aggressivität nach Scheithauer und Bull (ebd.) diesem Modell folgend etwa durch Defizite in der Fähigkeit zur Perspektivenübernahme oder Empathie sowie durch unzureichende soziale Normen begünstigt. Auf Seiten der Schule lassen sich mangelnde schulweite und explizite Verhaltensregeln bzw. der inkonsequente Umgang mit Regeln als Ursachen identifizieren. Klassenklima und Lehrerverhalten sind Einflussgrößen auf der Ebene der Klasse (ebd.).

Folgerichtig sollte Konfliktprävention durch ein respektvolles Klassenklima bzw. ebenso handelnde „Lehrer- und Mitschülermodelle“, das Formulieren und konsequente Einhalten von Verhaltensregeln auf Schul- und Klassenebene sowie einen Einbezug der Eltern umgesetzt werden. Auftretende Konflikte sollten transparent gemacht und gemeinsam gelöst werden.

*Disziplinprobleme im Unterricht*, definiert als Störungen, welche die effektive Lernzeit einschränken, sind negativ mit der Strukturiertheit des Unterrichts korreliert (z. B. Klieme & Rakoczy, 2008). Der erfolgreiche Umgang mit Störungen ist

weniger abhängig von spezifischen Reaktionen auf bereits erfolgtes Fehlverhalten, sondern von einer Prävention von Störungen, welche sich durch „Allgegenwärtigkeit“ (Kounin, 1976) beschreiben lässt.

Die *„Kenntnis über die sozialen und kulturellen Lebensbedingungen von SchülerInnen sowie die Einflussnahme auf deren individuelle Entwicklung im Rahmen der Schule“* spricht letztendlich die Forderung nach einem angemessenen Umgang mit Heterogenität an, die insbesondere auf Erkenntnissen der PISA-Studie (z. B. PISA-Konsortium, 2007) beruht. Die Ergebnisse von PISA legten offen, dass der soziale und sprachliche Hintergrund von SchülerInnen die Schülerleistung stark – zu stark – beeinflusst.

Ein Modell des Umgangs mit Heterogenität stellt Leutner (1992) vor: Der Autor unterscheidet zunächst zwischen dem *Zweck* und der *Umsetzung* der Anpassung des Unterrichts an heterogene Schülergruppen. Der Zweck der Adaption kann darin bestehen, Kompetenzdefizite auszugleichen (z. B. durch Stützunterricht), Defizite in Lern- und Leistungsvoraussetzungen zu beseitigen (z. B. mangelnde Lernmotivation, Anstrengungsbereitschaft oder Selbstwirksamkeitsüberzeugung) oder darin, Stärken und Interessen von SchülerInnen zu nutzen, indem entsprechende Methoden verstärkt genutzt werden. Diese Adaptionzwecke können dann auf verschiedene Weise realisiert werden: Lernziele können individuell je nach Lernvoraussetzung gesetzt, Lehr-Lern-Formen können an die Voraussetzungen der Lernenden angepasst oder der Umfang der Lernzeit kann individuell zur Verfügung gestellt werden.

Hinsichtlich welcher Lernermerkmale zeigt sich Heterogenität überhaupt? Hier werden insbesondere das Vorwissen, der Sprach- und Migrationshintergrund sowie der Entwicklungsstand diskutiert (Helmke, 2009). Lerner mit verschiedenen Ausprägungen auf diesen Variablen reagieren unterschiedlich auf verschiedene Unterrichtsformen: Je weniger Vorwissen und Sprachkompetenz SchülerInnen aufweisen, desto eher profitieren sie von klar strukturiertem Unterricht (Weinert & Helmke, 1984). Die Sprach- und Lesekompetenz stellt eine wichtige Determinante des Schulerfolgs dar. Somit kann insbesondere die Förderung der Lesefähigkeit von SchülerInnen mit Migrationshintergrund zu einer Angleichung von Schulleistungen führen (Billmann-Mahecha & Tiedemann, 2006).

Zusammenfassend zeigt die internationale Forschungslage zur Effektivität binnendifferenzierten Unterrichts nach Rossbach und Wellenreuther (2002), dass eine Individualisierung per se nicht zu besseren Schülerleistungen führt. Wichtiger sei, dass innerhalb der Binnendifferenzierung mit angemessenen Methoden (z. B. Auswahl geeigneter Lernmaterialien, Zusammenstellung „passender“ Lerngruppen, Adaption der Instruktionmethode auf das Vorwissen...) gearbeitet würde. Die Auswahl und Umsetzung adäquater Methoden bzw. deren Evaluation ist nach Helmke (2009) nur auf Basis diagnostischen Wissens zu leisten. Dem Diagnostizieren und Beurteilen widmet die KMK einen eigenen Kompetenzbereich.

Der Kompetenzbereich „Erziehen“ zeigt sich weniger einheitlich als der Kompetenzbereich „Unterrichten“. Die Facetten „Vermittlung von Werten und Normen“, „Unterstützung selbstbestimmten Urteilens und Handelns“, „Kenntnis über die sozialen und kulturellen Lebensbedingungen von SchülerInnen und Einflussnahme auf deren individuelle Entwicklung im Rahmen der Schule“ sowie „Finden von Lösungsansätzen für Schwierigkeiten und Konflikte in Schule und Unterricht“ lassen sich zwar empirisch untermauern, jedoch wurden diese Merkmale bislang nicht in einem „Modell des Erziehens“ miteinander in Beziehung gesetzt. „Erziehen“ gemäß der KMK (ebd.) lässt sich somit schwierig als homogener Kompetenzbereich interpretieren.

Beurteilen. Lehrerinnen und Lehrer üben ihre Beurteilungsaufgabe gerecht und verantwortungsbewusst aus.

Die Standards für die Lehrerbildung: Bildungswissenschaften fassen unter dem Kompetenzbereich „Beurteilen“ die Diagnose von Lernvoraussetzungen und Lernprozessen, die Leistungserfassung auf der Grundlage transparenter Beurteilungsmaßstäbe sowie die individuelle Förderung von SchülerInnen und die Beratung von Lernenden und deren Eltern zusammen.

Kunter und Voss (2011) bezeichnen dieses Kompetenzfeld als „Konstruktive Unterstützung“. Diese sei notwendig, um die Nutzung aktivierender Lerngelegenheiten erst zu ermöglichen. Voraussetzung sei auf Lehrerseite die Fähigkeit zur Feststellung und Struktur von Verständnisproblemen. Die Rückmeldung der Diagnostik an die SchülerInnen solle sich wertschätzend und geduldig gestalten. Gudjons (1998) benennt dies als „Unterstützung und Ermutigung aufgrund von Lernstandsdiagnosen“ und das „Anerkennen individueller Lernwege“.

Die *individuelle Förderung* von SchülerInnen – und somit der Umgang mit Heterogenität, wie er beispielsweise von Helmke (2009, s.o.) als Determinante effektiven Unterrichts definiert wurde – hängt Kunter und Voss (ebd.) zufolge zum Einen von der Beziehungsqualität zwischen SchülerIn und LehrerIn ab. Zum Anderen bestehe eine Abhängigkeit von der Strukturierung des Unterrichts dergestalt, dass er für die SchülerInnen mit verschiedenen Voraussetzungen (z. B. kognitiver, emotionaler, motivationaler oder sozialer Art) bewältigbar ist. Hier sprechen die Autoren beispielsweise die Methodik des Scaffolding oder den konstruktiven Umgang mit Fehlern (z. B. Schröder, 2007) an. Innerhalb eines Modells der Unterrichtsqualität aus den drei Faktoren „Potenzial zur kognitiven Aktivierung“, „Klassenführung“ und „Konstruktive Unterstützung“ untergliederten Kunter und Voss den letztgenannten Faktor in die Facetten „nicht-evaluativer Umgang mit Fehlern“, „respektvoller Umgang“, „Sozialorientierung“, „Geduld“ sowie „Anpassung an Schülerschwierigkeiten“.

Weinert (2000) bezeichnet *diagnostische Kompetenz* als die Fähigkeit, den Kenntnisstand, Lernfortschritte und –probleme sowie die Angemessenheit bestimmter an die SchülerInnen gestellter Aufgaben beurteilen zu können, um so das Handeln im Unterricht an diagnostischen Erkenntnissen ausrichten zu können. Allerdings, so der Autor, sei diese Kompetenz bei verschiedenen Lehrkräften äußerst verschieden ausgeprägt – die Ursache hierfür sei darin zu suchen, dass diagnostische Kompetenz in der Lehrerbildung kaum vermittelt würde.

In der Diagnostik gelten die Gütekriterien der Objektivität, Reliabilität und Validität. Dies jedoch gelte für die Diagnostik von Lehrkräften nur eingeschränkt, so schreiben Weinert und Schrader (1986): Diagnostische Urteile von Lehrpersonen können und dürfen ungenau sein, sofern sich die LehrerInnen dieser Ungenauigkeit und Subjektivität (Tent, 2006; Wild & Krapp, 2001) bewusst seien. Weniger gehe es um eine akkurate Zustandsdiagnostik, eher darum, *Änderungen* in Verhalten, Motivation und Wissen der SchülerInnen zu beachten.

Die Beachtung dessen, dass diagnostische Rückmeldungen an die SchülerInnen Lernvorgänge beeinflussen, ist nach Rheinberg, Bromme, Minsel, Winteler und Weidenmann (2001) relevant. Auch eine optimistische Erwartung der Lehrkraft an die SchülerInnen ist hierfür ebenso positiv bedeutsam wie eine Attribution von Misserfolgen auf mangelnde Anstrengung – nicht auf mangelnde Fähigkeit (Weinert & Schrader, 1986).



Die Standards für die Lehrerbildung: Bildungswissenschaften sprechen eine Facette der diagnostischen Kompetenz an, die nach Baumert und Kunter (2006) wohl bedeutsam für eine unterstützende kognitiv aktivierende Gestaltung des Unterrichts ist: Die Diagnostik nicht alleine von Lernergebnissen, sondern bereits des *Lernprozesses*. Im Rahmen des COACTIV-Modells (Voss & Kunter, 2011) lokalisieren die Autoren diagnostische Fähigkeiten als Kombination aus den fachdidaktischen und pädagogisch-psychologischen Wissensfacetten „Wissen über mathematisches Schülerdenken“, „Wissen über mathematische Aufgaben“ sowie „Wissen um Leistungsbeurteilung“.

Die *diagnostische Kompetenz* von Lehrkräften zeigt sich Helmke zufolge (2009) auf sechs Dimensionen: Nicht nur die Merkmale von SchülerInnen, sondern ebenso die Merkmale von Aufgaben müssen diagnostiziert werden (Dimension der Person- und Aufgabenmerkmale), um eine Passung zwischen Fähigkeit und Anforderung erreichen zu können. Sowohl fachliche Leistung als auch fachübergreifende lern- und leistungsbezogene Merkmale (z. B. Lernstrategien, Motivation,...) bedürfen einer Diagnose (Dimension des fachlichen/überfachlichen Bezugs). Nicht alleine Individuen, sondern ebenso ganze Klassen werden bezüglich ihres Leistungsniveaus diagnostiziert, um Schülerleistungen in Bezug zur entsprechenden Kohorte setzen zu können (Dimension der Ebene). Die prognostische Komponente diagnostischer Urteile spricht die Dimension „Status vs. Potenzial“ an: Neben dem Status Quo muss Diagnostik auch eine Aussage darüber leisten, was ein Lernender maximal zu erreichen im Stande wäre (auch: Gudjons, 2000). Die diagnostische Leistung bezieht sich neben einer Diagnostik anderer Personen auch auf eigenes Verhalten und Erleben (z. B. die Beurteilung des eigenen Unterrichts – wobei LehrerInnen hier wohl Defizite hinsichtlich einer validen Einschätzung aufweisen (z. B. Clausen, 2002) – dies bezeichnet Helmke (ebd.) als die Dimension der „Selbst- und Fremddiagnose“. Zuletzt bezieht sich Diagnostik stets auf soziale, individuelle oder kriteriale Bezugsnormen, deren adäquate Auswahl ausschlaggebend ist für die Güte des diagnostischen Urteils (Dimension der Bezugsnorm). Nach Baumert und Kunter (2006) sowie Spinath (2004, 2005) sind diese Dimensionen allerdings sogar innerhalb eines Fachbereichs unabhängig, woraus die Autoren schließen, dass es „eine diagnostische Kompetenz“ nicht gebe.

Brünken (2009) fasst den Forschungsstand zur diagnostischen Kompetenz Lehrender folgendermaßen zusammen: Hinsichtlich der Leistungsdiagnostik schneiden Lehrende zufriedenstellend ab, sofern sie soziale Bezugsnormen

anlegen. Intelligenz diagnostizieren Lehrende für höhere Leistungsniveaus treffend – insgesamt ist die Beurteilungsgüte hinsichtlich kognitiver Merkmale von SchülerInnen umso besser, je deutlicher die Merkmale mit Leistung korreliert sind. Gräsel, Krolak-Schwerdt, Nölle und Hörstermann (2010) fanden außerdem Hinweise darauf, dass Lehrkräfte soziale Merkmale (z. B. die Unterstützung durch die Eltern) in ihre Urteilsbildung mit einbeziehen.

Möglichkeiten zur Verbesserung der diagnostischen Kompetenz beschreibt Bromme (1997): Je mehr Lehrende über Schülerfehlvorstellungen, Lernstrategien, Verständnisschwierigkeiten sowie die kognitive Entwicklung von Schülern wissen, desto valider diagnostizieren sie Schülerleistungen. Eine ständige Bewusstmachung von typischen Urteilsfehlern (z. B. inadäquate Bezugsnormen, Haloefekte oder subjektive Theorien über Zusammenhänge zwischen Merkmalen) sollte Helmke (2009) zufolge für LehrerInnen selbstverständlich sein.

Diagnostische Kompetenz ist auf verschiedene Weise bedeutsam für gelingende Lernprozesse: 1. bilden diagnostische Ergebnisse den Rahmen für die Gestaltung von Unterricht, 2. besitzen diagnostische Fähigkeiten Auswirkungen auf die Lernleistung von SchülerInnen (z. B. Brunner, Anders, Hachfeld und Krauss, 2011) und 3. beeinflussen diagnostische Urteile im Form von Notenvergaben grundsätzlich die Bildungsbiographie von Lernenden. Aufgrund der hohen Relevanz dieses Kompetenzbereichs ist eine stärkere Vermittlung diagnostischen Wissens und Könnens durch die Lehrerbildung dringend angezeigt (z. B. Brünken).

Dass die *Zusammenarbeit mit Eltern* im Rahmen der Standards für die Lehrerbildung: Bildungswissenschaften Platz findet, ist Ergebnis empirischer Evidenz (z.B. Wild, Hofer & Pekrun, 2001; Krumm, 2006; Blömeke, Felbrich & Müller, 2008) zum Einfluss von Merkmalen der Familie (z.B. die Qualität der Eltern-Kind-Beziehung oder lern- und leistungsbezogene Einstellungen der Eltern) auf die Leistungsentwicklung und Lernmotivation von SchülerInnen. Ebenso beeinflusst eine enge Zusammenarbeit zwischen Eltern und LehrerInnen die Leistung und Motivation von SchülerInnen positiv (z.B. Balli, Demo & Wedman 1998).

Allerdings besteht die Elternberatung von LehrerInnen meist in Gesprächen an Elternabenden bzw. Elternsprechtagen und hat vorwiegend den Leistungsstand der SchülerInnen oder Erziehungsprobleme zum Inhalt (Wild, 2003). Arnold und Pätzold (2002) definieren Beratung im Schulkontext als Geben von Hilfestellungen bei Lern-

und Entwicklungsproblemen, und Griese (2009) bezeichnet die Elternarbeit als den Hauptbestandteil lehrerseitiger Beratung. Der Wert dieser Gespräche wird durch Eltern allerdings als eher gering eingeschätzt (Krumm, 1996), obwohl sie sich Hilfe und Unterstützung erhoffen. Weiter – und eher im Sinne der Standards für die Lehrerbildung: Bildungswissenschaften – fasst Keller (2003) die Beratung, indem Elternberatung, Lernberatung oder die Beratung von KollegInnen benannt werden.

Auf Basis der Diagnostik von Lernergebnissen und –prozessen oder Schülerverhalten wird von Lehrenden mehr und mehr eine Beratungsfunktion wahrgenommen (Schnebel, 2007). Das „Beratungswissen“ wird im COACTIV-Modell (Baumert & Kunter, 2011) als ein Aspekt des Professionswissens definiert, allerdings nicht weiter spezifiziert oder empirisch erfasst. Bruder, Klug, Hertel und Schmitz (2011) entwickelten dagegen explizit ein Modell der Beratungskompetenz von Lehrenden im Bereich der Lernberatung. Auch diese Autoren betonen die Fähigkeit zur Diagnostik als Voraussetzung für eine Unterstützung – in diesem Falle durch Beratung.

Bruder et al. (ebd.) nehmen fünf Dimensionen der beraterischen Kompetenz an: *Berater-Skills* beziehen sich auf das Strukturieren des Beratungsgesprächs sowie den Einsatz von Gesprächsstrategien (z. B. aktives Zuhören oder Paraphrasieren). *Ressourcen- und Lösungsorientierung* bezeichnen das Wissen und Können im Bereich der Ziel-, Lösungs- und Ressourcenorientierung in einer Beratung, *Diagnostizieren und Pädagogisches Wissen* das Erkennen und Definieren eines Problems sowie die Ursachensuche und anschließend das Wissen um Lösungsstrategien. Die Dimension der *Kooperation* bezeichnet das kooperative Handeln im Umgang mit den Ratsuchenden bzw. eine systemische Orientierung, sowie *Coping* die Fähigkeit, in Gesprächen mit Kritik und schwierigen Beratungssituationen umgehen zu können.

Die Standards für die Lehrerbildung: Bildungswissenschaften fassen unter dem Kompetenzbereich „Beurteilen“ die Diagnose von Lernvoraussetzungen und Lernprozessen, die Leistungserfassung auf der Grundlage transparenter Beurteilungsmaßstäbe sowie die individuelle Förderung von SchülerInnen und die Beratung von Lernenden und deren Eltern zusammen.

Die Kompetenzfacetten „diagnostische Kompetenz“ sowie „Beratung“ zeigen sich empirisch gut fundiert. Allerdings scheint es erneut schwierig, beide Facetten unter einem KMK-Kompetenzbereich des „Beurteilens“ subsumieren zu können, da die

Forschung zu Diagnostik und Beratung bislang erst in Ansätzen integriert wurde (z. B. Kunter & Voss, 2011).

Innovieren. Lehrerinnen und Lehrer entwickeln ihre Kompetenzen ständig weiter.

Der letzte Kompetenzbereich der Standards für die Lehrerbildung: Bildungswissenschaften beinhaltet das Verständnis des Lehrberufs als ständige Lernaufgabe bzw. als Profession mit besonderen Anforderungen, mit besonderer Verantwortung und Verpflichtung. Weiterhin ist in diesem Kompetenzbereich die Planung und Umsetzung schulischer Projekte angesprochen.

Dem Lernen durch *Weiterbildung* kommt angesichts sich immer schneller fortentwickelnder Wissensbestände große Bedeutung zu. Die Forschungslage zur Weiterbildung von LehrerInnen ist allerdings äußerst dünn, was durchaus an den Methoden zur Evaluation von Weiterbildungsveranstaltungen liegen könnte: Wie häufig im Rahmen solcher Evaluationen der Fall, wird am Ende des Weiterbildungsangebots meist nur die Zufriedenheit der TeilnehmerInnen abgefragt. Dass zwischen diesem subjektiven Kriterium und dem doch eigentlich interessierenden, nämlich der Schülerleistung, kein Zusammenhang besteht, zeigten beispielsweise Jacob und Lefgren (2004).

Nicodemus, Jäger und Bodensohn (2010) legen eine der wenigen Untersuchungen zur Wirksamkeit von Lehrerfortbildungen vor: Die Autoren focussieren den Transfer der in der Fortbildung „Mathematik Anders Machen“ (Deutsche Telekom Stiftung) vermittelten Inhalte und kommen zu dem Schluss, dass dieser Transfer eher nicht stattfindet. Die teilnehmenden LehrerInnen gaben hierbei an, dass die Gründe hierfür weder im Widerstand von Kollegen noch am mangelnden Praxisbezug der vermittelten Inhalten lägen. Die Studie bleibt leider eine weitere Erklärungssuche schuldig.

Über die Ursachen der wenig ermutigenden Ergebnisse bzw. der dünnen Forschungslage bezüglich der Weiterbildung von LehrerInnen kann spekuliert werden, dass a) die Evaluationspraxis auf Ebene 1 nach Kirkpatrick (1998) in Form einer Abfrage der „customer satisfaction“ stehen bleiben und b) der Weiterbildung von Lehrkräften sowohl im Rahmen der Ausbildung als auch durch die Lehrenden selbst kein ausreichender Stellenwert zukommt, wie beispielsweise von Helmke (2009) angenommen. Leutwyler und Ettlín (2008) fordern somit zur

Systematisierung der Weiterbildung von Lehrenden und als Grundlage für systematische Evaluationen „Standards für die Lehrerweiterbildung“.

Eine weitere Lernform stellt das Lernen aufgrund von Reflektion dar. Die *Reflektionsfähigkeit* bezüglich des eigenen Tuns wird von verschiedenen Seiten als relevant für den Lehrberuf angesehen. Die wichtige Rolle der Selbstreflektion im Bereich diagnostischer Urteilsbildung wurde oben bereits angesprochen. Nach Helmke (2007) ist Reflektionsfähigkeit jedoch für die Entwicklung des gesamten Unterrichts relevant und ein „unabdingbares Merkmal der Lehrperson“ (Helmke, 2009): Der Autor betont dabei das datengeleitete, kooperative Formulieren von Hypothesen über Funktionsmuster bzw. das Hinterfragen eigener subjektiver Theorien. Auch nach Seidel und Prenzel (2007) stellt die Fähigkeit zur Reflektion des Unterrichtsgeschehens eine wichtige Komponente pädagogischer Professionalität dar. Eine vielversprechende Methode ist hierbei die Videoanalyse des eigenen – oder fremden – Unterrichts (z. B. Prenzel, Seidel et al., 2002). Auch zur Diagnose der Reflektionsfähigkeit von Lehrkräften eignen sich Videosequenzen: Schmelzing et al. (2009) fanden beispielsweise, dass das Reflexionswissen von Biologielehrkräften durch einen videogestützten Test erfasst werden kann.

Ein Blick in die USA zu den Standards des einflussreichen National Board for Professional Teaching Standards (2009) zeigt, dass ebenso aus internationaler Perspektive die Reflektion des eigenen Tuns als Kernkompetenz von Lehrenden betrachtet wird: „Proposition 4: LehrerInnen reflektieren Ihre Arbeit systematisch und lernen aus ihrer Erfahrung.“

Lehmann und Nieke (1999) erweitern den Reflektionsbegriff um die – den Autoren zufolge – nötige Selbstreflektion von Lehrkräften: Hier werden Aspekte des beruflichen Selbstkonzept bzw. der Berufsethik angesprochen bzw. die Reflektion von eigenen Überzeugungen (etwa über Mathematik als reines Anwenden von Lösungsschemata; Köller, Baumert & Neubrand, 2000) und subjektiven Theorien beispielsweise über das Zustandekommen von Lernergebnissen. Dabei sind subjektive Theorien äußerst änderungsresistent und leiten zudem das Handeln stärker als wissenschaftliche Theorien (Rheinberg, Bromme, Minsel, Winteler & Weidenmann, 2006).

Baumert und Kunter (2006) stellen in ihrem Modell der professionellen Handlungskompetenz die Kompetenzfacetten „Wissen und Können“ den Facetten

„Werthaltungen und Überzeugungen“ sowie „Motivation und Selbstregulation“ gegenüber. Hinsichtlich der Kompetenzfacetten „Werthaltungen und Überzeugungen“ treffen die Autoren Unterscheidungen zwischen *Wertbindungen, epistemologischen Überzeugungen, subjektiven Lehr- und Lerntheorien* sowie *Zielsystemen für Curriculum und Unterricht*. Diese Überzeugungen und Werthaltungen bedürfen aufgrund ihres rein subjektiven Charakters keinerlei Rechtfertigung und Validierung, sondern basieren alleine auf individuellen „Richtigkeitsannahmen“ und zeigen sich über die Zeit hinweg als recht stabil.

Bislang ist weithin unklar, inwiefern die Werthaltungen von Lehrkräften deren professionelles Handeln beeinflussen, und auch hinsichtlich epistemologischer Überzeugungen sowie Zielsystemen besteht dahingehend kein Konsens (Baumert & Kunter, 2006).

In Bezug auf das Lehrerhandeln werden *motivationale Aspekte* insbesondere mit Blick auf *Kontrollüberzeugungen und Selbstwirksamkeitserwartungen, intrinsische Motivation* sowie *Selbstregulation* untersucht. Schmitz und Schwarzer (2000) gelangen zu dem Schluss, dass *Selbstwirksamkeitsüberzeugungen* die psychische Regulationsfähigkeit von Lehrkräften maßgeblich beeinflussen. Insgesamt zeigt sich die Forschungslage jedoch uneindeutig.

Die *Selbstregulationsfähigkeit* stellt einen wichtigen Aspekt der professionellen Handlungskompetenz von Lehrerinnen und Lehrern dar. Forschungen zum Belastungserleben und Coping zeigen dies deutlich auf (z. B. Schaarschmidt, 2005). Klusmann, Kunter, Trautwein und Baumert (2006) zeigten beispielsweise, dass hohes Engagement sowie hohe Distanzierungsfähigkeit mit hoher Berufszufriedenheit, geringer emotionaler Erschöpfung, persönlichem Wohlbefinden sowie unterstützender Unterrichtsführung einhergehen. Klusmann bzw. die Forschungsgruppe um COACTIV (2011) sehen das erfolgreiche Haushalten mit den eigenen Ressourcen aufgrund der verschiedensten, sich überschneidenden Anforderungen an die Lehrperson als Teil der professionellen pädagogischen Kompetenz. Als Kriterien nutzen die Autoren einerseits das „berufliche Wohlbefinden“, andererseits die Gestaltung des Unterrichts. Lehrpersonen des „Gesundheitstyps“ (hohes Engagement bei gleichzeitig hoher Widerstandsfähigkeit) schnitten dabei bei empirischen Analysen auf beiden Kriterien am Besten ab.

Auch aufgrund ihrer *Vorbildfunktion* (z. B. Hurrelmann, 2006) kommt Lehrpersonen besondere Verantwortung zu, deren sie sich bewusst sein sollten (z. B. Blömeke, Felbrich & Müller, 2008). Nach Helmke (2009) wirken Lehrkräfte als Modell und beeinflussen somit direkt – neben der indirekten Beeinflussung über die Gestaltung des Unterrichts – die Persönlichkeitsentwicklung von SchülerInnen. Ebenso betonen Schuler, Hell und Trapmann (2007) auf Basis ihrer Anforderungsanalyse des Lehrberufs die Vorbildfunktion der Lehrkraft (hierzu auch Hurrelmann, 2006) bzw. die Relevanz authentischen Verhaltens. Bezüglich der Sprachnutzung seien Lehrpersonen neben den Eltern sogar die wichtigsten Modelle, so Helmke (2009). Vor allem aber wirken LehrerInnen über ihre Einstellung zum Fach: Zeigen Lehrkräfte Begeisterung und Wertschätzung für ihr Fach und vermitteln sie, dass es wichtig und interessant sei, so erhöht sich die Wahrscheinlichkeit, dass SchülerInnen diese Einstellungen übernehmen (Gettinger & Kohler, 2006)

Der *Kooperation zwischen den Lehrkräften* wurde im Zuge eines veränderten Lehrerbilds größere Bedeutung beigemessen (Gudjons, 2000), und inzwischen ist eine hohe Relevanz der Lehrerkooperation für die Schulqualität kaum mehr strittig (Steinert et al., 2006).

Allerdings zeigen sich die empirischen Ergebnisse zum Zusammenhang „Lehrerkooperation – Schülerleistung“ nicht konsistent. Nach Scheerens und Bosker (1997) könnte dies an der bislang uneinheitlichen Modellierung des Konstrukts liegen. Als Teil der Schulentwicklung wird Lehrerkooperation jenseits der Orientierung an Schülerleistungen als Merkmal der Organisationsqualität gesehen, welches über eine Verbesserung des Unterricht und der Lehr- und Lernkultur schließlich auch die Lernergebnisse von SchülerInnen verbessern kann. Die Struktur von „Schule“ an sich scheint aber diese Zusammenarbeit aufgrund der Einteilung des Arbeitsinhaltes in Fächer, Jahrgangsstufen und Klassen sowie des Charakters der Lehrtätigkeit als „Einzelarbeit“ zu erschweren (Steinert et al., ebd.).

Seashore, Louis und Leithwood (1998) identifizieren als Determinanten für eine Erhöhung der schulischen Effektivität eine gemeinsame Zielorientierung, eine Schwerpunktsetzung auf das schülerseitige Lernen, die offene Diskussion individueller Unterrichtspraktiken und die regelmäßige Diskussion und Kooperation in Fragen des Curriculums, des Unterrichts, der Lernfortschritte sowie der Leistungsbeurteilung. Die Autoren benennen zudem das reziproke Feedback als Quelle sozio-emotionaler Unterstützung und der Kompetenzerweiterung. Rolf

(2011) gelangt zu den für die *Schulentwicklung* relevanten drei Faktoren Zielorientierung, wechselseitiges Feedback sowie Teamarbeit. Einen besonderen Stellenwert bemisst der Autor ebenso den Führungskräften der Schule – nicht überraschend, da Organisationsentwicklung nur gelingen kann, wenn die Leitung bei der Zielerreichung als Vorbild fungiert. Hier schließt Schönig (2002) an, welcher die Organisationskultur als ein Schlüsselkonzept für die Qualitätserhöhung durch Schulentwicklung bezeichnet.

Mit Bezug auf die theoretischen Konzepte der Schuleffektivitäts- und Schulentwicklungsforschung formulieren Steinert et al. (ebd., S. 191) ein theoretisch begründetes Modell der Lehrerkoooperation mit drei Facetten:

- *Schulorganisation*: Pädagogische Orientierung und Zielkonzept der Schule, Koordination der schulischen Förderangebote, Information und Kommunikation, Aufgabenverteilung und Entscheidungsbeteiligung, Berichterstattung und Evaluation
- *Personalmanagement und Professionalisierung*: Rekrutierung, Fortbildung und Training der Lehrkräfte, Beratung und Diagnose des Lehrerhandelns.
- *Unterrichtsorganisation*: Curriculum, Fachinhalte, Fach-Didaktik und Methodik, Beratung und Diagnose der Lernentwicklungen und individuelle Förderung der SchülerInnen

Zudem identifizierten die Autoren empirisch fünf Niveaustufen der Lehrerkoooperation, welche bei (1) vereinzeltm Austausch mit unklarer Zielkonzeption („Fragmentierung“) beginnen und dann über (2) fach- und jahrgangsspezifisch abgestimmtes Lehrerhandeln sowie ein globales Zielkonzept („Differenzierung“); (3) die partielle Kooperation zur Planung und Durchführung des Unterrichts, Austausch über Fachinhalte und –didaktik und Notenmaßstäbe sowie Selbstevaluation („Koordination“); (4) ein detailliertes Zielkonzept, umfassend abgestimmtes Lehrerhandeln innerhalb und zwischen Jahrgangsstufen und Fächern, umfassende Kooperation bei Unterrichtsplanung und -durchführung, wechselseitige Beratung über fachliche und überfachliche Inhalte, Didaktik und Diagnostik („Interaktion“) zu einem (5) systematischen Zielkonzept und abgestimmten Lehrerhandeln, Transparenz und wechselseitiger Adaptivität im Unterrichtshandeln, systematischer Beobachtung von Lehrerhandeln und Lernentwicklungen, Selbst- und Fremdevaluation, systematischer Fortbildung („Integration“) reichen.



Kelchtermans (2006) unterscheidet zwischen Kooperation und Kollegialität und kritisiert, dass beide Begriffe häufig denkungsgleich verstanden würden. Kooperation, so der Autor, beschreibe kooperatives Handeln, während Kollegialität die Beziehungen zwischen den MitarbeiterInnen der Schule bezeichne. Beide Aspekte bedingen sich allerdings gegenseitig und seien unmittelbar vom Kontext der Schule abhängig.

Die *Lehrerkooperation* wird als wichtiges Element der Schulentwicklung (z. B. auch Blömeke, Felbrich & Müller, 2008) interpretiert, welches (indirekt) auf Schülerleistungen positiv einwirken könne.

Das Wissen über die Rahmenbedingungen des Lehrens und Lernens bezeichnen Baumert und Kunter (2011) als *Organisationswissen* und unterteilen dieses auf Basis schultheoretischer und organisationssoziologischer Literatur in fünf Facetten: Das Wissen über (1) das Bildungssystem und dessen Rahmenbedingungen, (2) die Steuerung und Governance, (3) Schulorganisation, Schulökologie, Schulverfassung, Rechtsstellung von SchülerInnen, Eltern, LehrerInnen, Aufgaben der Schulleitung, (4) Schulqualität und –effektivität (5) Schultheorien.

Der Kompetenzbereich „Innovieren“ lässt sich untergliedern in die Facetten „Weiterbildung“, „Selbstreflektion“, „Selbstregulation“ sowie „Schulentwicklung“. Eine lose Verknüpfung dieser Unterkompetenzen könnte sich über die Facette „Selbstreflektion“ als Basis für das Erkennen von Optimierungspotentialen auf verschiedenen Ebenen herstellen. Dies bleibt jedoch spekulativ, und so muss auch für diesen Kompetenzbereich das Merkmal der Heterogenität gelten.

### Wissen vs. Können

In den Standards für die Lehrerbildung: Bildungswissenschaften wird jede der elf „Unterkompetenzen“ sowohl auf die theoretischen als auch auf die praktischen Ausbildungsabschnitte bezogen, indem die betreffende Kompetenz in Wissens- und Könnensinhalte übersetzt wird (Abbildung 1 zeigt ein Beispiel).

**Kompetenzbereich: Unterrichten****Lehrerinnen und Lehrer sind Fachleute für das Lehren und Lernen.****Kompetenz 1:**

Lehrerinnen und Lehrer planen Unterricht fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.

Standards für die theoretischen Ausbildungsabschnitte	Standards für die praktischen Ausbildungsabschnitte
<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- kennen die einschlägigen Bildungstheorien, verstehen bildungs- und erziehungstheoretische Ziele sowie die daraus abzuleitenden Standards und reflektieren diese kritisch.</li> <li>- kennen allgemeine und fachbezogene Didaktiken und wissen, was bei der Planung von Unterrichtseinheiten beachtet werden muss.</li> <li>- kennen unterschiedliche Unterrichtsmethoden und Aufgabenformen und wissen, wie man sie anforderungs- und situationsgerecht einsetzt.</li> <li>- kennen Konzepte der Medienpädagogik und – psychologie und Möglichkeiten und Grenzen eines anforderungs- und situationsgerechten Einsatzes von Medien im Unterricht.</li> <li>- kennen Verfahren für die Beurteilung von Lehrleistung und Unterrichtsqualität.</li> </ul>	<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- verknüpfen fachwissenschaftliche und fachdidaktische Argumente und planen und gestalten Unterricht.</li> <li>- wählen Inhalte und Methoden, Arbeits- und Kommunikationsformen aus.</li> <li>- integrieren moderne Informations- und Kommunikationstechnologien didaktisch sinnvoll und reflektieren den eigenen Medieneinsatz.</li> <li>- überprüfen die Qualität des eigenen Lernens.</li> </ul>

Abbildung 1. Wissen und Können innerhalb der KMK-Kompetenz "Unterrichtsplanung und -durchführung" (Beispiel)

Wie stehen Wissen und Können miteinander in Beziehung?

Dass beide Komponenten grundlegend die professionelle pädagogische Kompetenz determinieren, darüber besteht weitgehende Übereinstimmung (Baumert & Kunter, 2011). Das COACTIV-Modell der Autoren basiert beispielsweise auf der klassischen Unterscheidung zwischen deklarativem („Wissen, was“) und prozeduralem Wissen („Wissen, wie“), und auch Mayer (2003) setzt Kompetenz mit deklarativem und prozeduralem Wissen gleich. Simonton (2003) unterscheidet Wissen und skills (Fertigkeiten), die gemeinsam erfolgreiches Handeln ermöglichen, wohingegen Ceci, Barnett und Kanaya (2003) Kompetenzen als skills definieren, die in Laufe des Lebens erworben wurden. Klieme und Hartig (2007) nehmen eine Verknüpfung deklarativen und prozeduralen Wissens mit Fertigkeiten, Einstellungen (beliefs) sowie Regulationskomponenten im kompetenten Handeln an: „In diesem Sinne

kann Kompetenz verstanden werden als die Verbindung von Wissen und Können in der Bewältigung von Handlungsanforderungen.“ (S.19)

Erpenbeck und von Rosenstiel (2007) definieren Kompetenzen als subjektzentrierte „Selbstorganisationsdispositionen“ (S. XIX) in Abgrenzung zu sachverhaltszentrierten „Qualifikationen“: Letztere würden in normierbaren Prüfungssituationen sichtbar, welche das aktuelle Wissen widerspiegeln, während erstere sich im realen Handeln zeigen. Die Autoren unterscheiden also klar zwischen der Wissens- und der Kompetenzdimension.

Shulman (1986) trifft bezüglich pädagogischen *Wissens* die einflussreiche Unterscheidung von *fachspezifischem Inhaltswissen*, *pädagogischem Inhaltswissen* sowie *fachdidaktischem Wissen*. Diese Gliederung spiegelt sich überwiegend in der Struktur der Lehramtsstudiengänge mit deren fachwissenschaftlichen, fachdidaktischen und bildungswissenschaftlichen Modulen.

Bromme (1997) differenziert in seiner „Topologie des professionellen Wissens“ fünf Wissensbereiche des/der Lehrers/in: (1) fachliches Wissen, (2) curriculares Wissen, (3) die „Philosophie des Schulfaches“ (Auffassungen über die Nützlichkeit eines Faches), (4) pädagogisches Wissen und (5) fachspezifisch-pädagogisches Wissen. Die Verschmelzung dieser verschiedenen Wissensbereiche im Handeln von LehrerInnen bezeichnet Bromme (ebd.) als „das Besondere des professionellen Wissens von Lehrern gegenüber dem kodifizierten Wissen der Fachdisziplinen.“ (S. 198) Die Wirkung dieses Wissens zeige sich dabei nicht direkt im Sinne einer kausalen Beziehung „Wissen → Handeln“, sondern vielmehr als „eine Veränderung der kategorialen Wahrnehmung von Unterrichtssituationen.“ (Bromme, ebd., S. 199).

Hinsichtlich des „Lehrerwissens“ muss zu weiten Teilen auf die professionsbezogene Expertiseforschung zurückgegriffen werden (Baumert & Kunter, 2006). Baumert und Kunter (ebd.) fassen deren Ergebnisse zusammen und gelangen zu dem Resultat, dass (1) professionelles Wissen domänenspezifisch und ausbildungs- bzw. trainingsabhängig sei, (2) Expertenwissen gut vernetzt und hierarchisch strukturiert sei, (3) in professionellen Domänen Expertenwissen in Konzepten, Schemata und Skripts sowie episodisches Wissen strukturiert sei, (4) professionelles Expertenwissen kontextabhängig sei und adaptives Verhalten ermögliche, (5) Prozeduren sowohl automatisiert als auch flexibel seien und (6)

professionelle Expertise auf der Basis theoretischen Wissens fuße. Weite Bereiche des Lehrerwissens seien dabei dem praktischen Wissen und Können zuzuordnen, wobei ein großer Teil der Lehrerexpertise sich als theoretisches Wissen darstelle. Praktische Expertise im Sinne des „Könnens“ sei erst dann erreicht, wenn theoretisches Wissen und Erfahrungswissen vollständig integriert seien.

Blomberg (2011) bezweifelt, dass das von den Universitäten vermittelte theoretische Wissenschaftswissen adäquat in praktisches Handlungswissen überführt werden könne, und auch Helmke, Helmke und Schrader (2007) erwähnen die Diskrepanz zwischen dem Wissenschaftswissen, welches Lehramtsstudierenden von den ausbildenden Universitäten vermittelt wird, und dem praktischen Handlungswissen, welches von den traditionsgeprägten (Fach-)Didaktiken übernommen werde – eine Perspektive, die beispielsweise auch Schocker-von Ditfurth (2001) einnimmt.

Kolbe (2004) konstatiert, dass wissenschaftliches Wissen nicht in die Anwendungssituation übertragen werde, sondern dass im konkreten Handlungskontext ein „eigenes professionelles Wissen“ über die Interpretation und Beurteilung der unterrichtlichen Situation entstehe. Nach Kolbe (ebd.) sei das professionelle Wissen von Lehrkräften ein unbewusstes: Unter Zeit- und Handlungsdruck müssen Wissensstände abrufbar sein, die zur situativen Deutung und Handlung befähigen, ohne dafür auf bewusste Handlungsregeln oder Handlungsentscheidungen zurückgreifen zu müssen. Für den Autor (ebd.) entstehen Interpretations- und Handlungsmuster aus der Erfahrung. Diese Muster seien Basis für berufliches Handeln.

Allerdings betont Kolbe (ebd.), dass sich dieses *Erfahrungswissen* (folglich die Muster des Interpretierens und Handelns) nur durch eine Reflexion in Bezug auf *wissenschaftliches Wissen* weiterentwickeln könne. Hierdurch würde das Erfahrungswissen hinterfragt und vertieft.

Kolbe (2004) definiert das professionelle Erfahrungswissen von Lehrkräften also nicht als „Wissen um kausal formulierbare Regelmäßigkeiten über den Zusammenhang von Lehrerhandeln und Wirkungen auf Lernerseite“, nicht als „zweckrational angewendetes, quasi-technisches Regelwissen“ (S. 218), sondern als situativ angewandte Interpretations- und Handlungsmuster. Der Autor postuliert sogar: „Wissen ist nicht ‚handlungssteuernd‘“ (ebd.). Zusammenfassend lässt sich

nach Kolbe (ebd.) professionelles Wissen verstehen als Wissen, welches über die „Logik des Berufshandelns rekonstruierbar ist“ (S. 221).

Diese Sicht des professionellen pädagogischen Wissens und Handelns weist deutliche Parallelen zu Weinerts (2001) Definition des Kompetenzbegriffs auf. Dort kann die professionelle Kompetenz von (angehenden) Lehrpersonen verstanden werden als das erfolgreiche Problemlösen in Lehr- und Lernumgebungen. Bei Kolbe (2004) rufen Lehrkräfte „unter Handlungsdruck“ („*Problem*“) situations- und kontextspezifische Interpretations- und Handlungsmuster („*Lösung*“) ab.

Das Verhältnis von Wissen und Handeln bzw. Theorie und Praxis ist dieser Sicht zufolge kein lineares (so auch Bastian & Helsper, 2000). Der mangelnde Transfer des an den Universitäten und Hochschulen vermittelten Wissens auf die pädagogische Praxis verwundert somit nicht länger. Selbstverständlich darf hieraus nicht der Schluss gezogen werden, dass theoretisches Wissenschaftswissen für das Lehrerhandeln nicht relevant sei – vielmehr stellt dieses Wissen einerseits den Rahmen für die notwendige Reflexion des eigenen Handelns, andererseits die Begründung für das Handeln und Interpretieren (beispielsweise von Entwicklungsverläufen) (Bastian & Helsper, 2000).

Die vorliegende Arbeit orientiert sich an Kolbes (2004) Konzept des professionellen Wissens als situationsspezifische Interpretations- und Handlungsmuster. Der entwickelte Instrument „Simulationsorientierte Erfassung pädagogischer Kompetenzen im Lehrberuf“ verzichtet folglich auf die Erfassung *reiner* Wissensaspekte, sondern nähert sich der professionellen Kompetenz von (angehenden) Lehrkräften situativ und handlungsbasiert durch die Vorgabe von simulationsorientierten Items, deren Beantwortung freilich aufgrund einer Wissensbasis erfolgen muss.

Resümee: Empirische Evidenz für die vier Kompetenzbereiche „Unterrichten“, „Erziehen“, „Beurteilen“, „Innovieren“, bzw. die Unterscheidung zwischen Wissen und Können

Standards als outputorientierte Zielvorgaben in der Lehrerbildung werden häufig kritisiert: Sie seien nicht ausreichend empirisch abgesichert und theoretisch kaum fundiert (Tulodziecki & Grafe, 2006). Diese Kritik scheint angesichts der breiten Forschungslage zu den von der KMK formulierten Standards für die Lehrerbildung:

Bildungswissenschaften nicht angemessen. Ebenso argumentiert Oser (2003; 2005) bezüglich der Kritik an dem von ihm formulierten ersten deutschsprachigen Standardkonzept, dass diese Standards zwar nicht aus einem Kompetenzmodell abgeleitet, jedoch einzelnen fundierten Theorien zuordenbar seien.

Zutreffend ist allerdings der Einwand Baumerts und Kunters (2006), dass die Standards hinsichtlich ihrer Begrifflichkeit und der Einteilung in die vier Kompetenzbereiche „Unterrichten“, „Erziehen“, „Beurteilen“ und „Innovieren“ nicht mit empirischen Ergebnissen pädagogisch-psychologischer Forschung übereinstimmen: Einerseits überschneiden sich Kompetenzbereiche stark, andererseits werden unter einem Kompetenzbereich Konstrukte subsumiert, welche aus pädagogisch-psychologischer Forschungsperspektive unabhängig sind:

Im Kompetenzbereich „Erziehen“ wird die Vermittlung von Werten und Normen sowie die Unterstützung selbstbestimmten Urteilens und Handelns von SchülerInnen ebenso subsumiert wie der Umgang mit heterogenen Klassengruppen und die individuelle Förderung. Zuletzt ist eine Kompetenzfacette angesprochen, die sich auf den lösungsorientierten Umgang mit Konflikten bezieht. Der Kompetenzbereich fasst somit drei Konzepte zusammen, welche aus Sicht der pädagogisch-psychologischen Forschung zum Einen eher unabhängig voneinander sind (Werteerziehung, Umgang mit Heterogenität/individuelle Förderung, Konfliktmanagement) bzw. zum Anderen eng mit Konstrukten anderer Kompetenzbereiche zusammenhängen: Der Umgang mit Heterogenität bzw. die Individualförderung überschneiden sich mit der Gestaltung von Unterricht, indem Unterrichtsmaterialien bzw. Methoden auf die Heterogenität von Klassen abgestimmt werden müssen. Zuvor steht allerdings eine weitere überschneidende Kompetenz: Die Fähigkeit, Heterogenität zu diagnostizieren und entsprechende Fördermaßnahmen und Unterrichtsformen adäquat implementieren zu können.

Die Struktur des Kompetenzbereichs „Beurteilen“ ist praktisch orientiert: Aus einer erfolgten Diagnose über Schülermerkmale leitet sich im Schulalltag der Umgang mit heterogenen Schülervoraussetzungen im Unterricht sowie eine Beratungsleistung an SchülerInnen oder Eltern ab. Aus Sicht der pädagogisch-psychologischen Forschung stellen „Diagnostik“, „Umgang mit Heterogenität“ und „Beratung“ jedoch getrennte Konstrukte (allerdings mit engem Bezug zur Basis der Diagnostik) dar. Demnach ist auch dieser Kompetenzbereich aus empirischer Sicht heterogen

formuliert. Trotzdem zeigt sich eine empirische und theoretische Fundierung hinsichtlich der einzelnen Konstrukte.

Der letzte Kompetenzbereich „Innovieren“ erweist sich als ebenso wenig eindimensional. Weiterbildung, Reflektionsfähigkeit und Selbstregulation sowie die Beteiligung an schulischer Entwicklung werden unter dem Begriff des „Innovierens“ subsumiert. Durch den zugrunde liegenden Gedanken des „Weiterentwickelns“ oder der Reflektionsfähigkeit als Basis für Weiterentwicklung können die Facetten immerhin lose geklammert werden. Wiederum sind trotzdem sämtliche Unterkompetenzen zumindest aus theoretischer Warte Aspekte professioneller pädagogischer Kompetenz.

Die Ansprüche an ein elaboriertes Modell der professionellen pädagogischen Kompetenz erfüllen die Standards für die Lehrerbildung: Bildungswissenschaften somit nicht, sondern sie präsentieren einen Anforderungskatalog an (zukünftige) Lehrkräfte, der allerdings empirisch und theoretisch gut fundiert ist, wie zusammenfassend die durch Blömeke, Felbrich und Müller (2008) identifizierten Anforderungen an Mathematiklehrkräfte zeigen: Die Autoren gelangen im Rahmen der internationalen MT21-Studie zu Anforderungsdimensionen, die nahezu deckungsgleich zu den Standards für die Lehrerbildung: Bildungswissenschaften sind: Anforderungsbereich A wird bezeichnet mit „Themen- und Methodenwahl, Sequenzierung des Lernprozesses“, Bereich B lautet „Leistungserfassung, Schüler- und Elternberatung“, Bereich C ist mit „Unterstützung der sozialen, moralischen und emotionalen Entwicklung“, Bereich D mit „Schulentwicklung“ und Bereich E mit „Professionsethos“ überschrieben. Mit Ausnahme der Bereiche D und E, welche in den Standards für die Lehrerbildung: Bildungswissenschaften gemeinsam einen Kompetenzbereich bilden, übereinstimmen die Anforderungs- bzw. Kompetenzbereiche nahezu vollständig – insbesondere auch beim Blick auf die Unterkompetenzen, welche die einzelnen Bereiche beinhalten. Tabelle 1 stellt beide Konzepte einander gegenüber.

Tabelle 1. Vergleich der Anforderungen gemäß MT21 (Blömeke et al., 2008) mit den Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004)

<b>Berufliche Anforderungen gemäß MT21 (Blömeke et al., 2008)</b>	<b>Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004)</b>
<p><b>A: „Themen- und Methodenwahl, Sequenzierung des Lernprozesses“</b></p> <ul style="list-style-type: none"> <li>- Auswahl und Begründung des Unterrichtsinhaltes</li> <li>- Planung und Begleitung von Unterricht</li> </ul>	<p><b>Unterrichten</b></p> <ul style="list-style-type: none"> <li>- fach- und sachgerechte Unterrichtsplanung und -durchführung</li> <li>- Bereitstellen von Lerngelegenheiten</li> <li>- Schülermotivation</li> <li>- Förderung selbstbestimmten Lernens und Arbeitens</li> </ul>
<p><b>B: „Leistungserfassung, Schüler- und Elternberatung“</b></p> <ul style="list-style-type: none"> <li>- Diagnostik von Lernvoraussetzungen, Schülerleistungen, Lernprozessen, Fehlvorstellungen</li> <li>- Leistungserfassung</li> <li>- Schüler- und Elternberatung</li> <li>- Umgang mit Fehlern, Feedback</li> </ul>	<p><b>Beurteilen</b></p> <ul style="list-style-type: none"> <li>- Diagnostik von Lernvoraussetzungen und Lernprozessen</li> <li>- Schüler- und Elternberatung</li> <li>- Leistungserfassung aufgrund transparenter Beurteilungsmaßstäbe</li> </ul>
<p><b>C: „Unterstützung der sozialen, moralischen und emotionalen Entwicklung“</b></p> <ul style="list-style-type: none"> <li>- Aufbau Lehrer-Schüler-Beziehung</li> <li>- Förderung der Werte- und Moralentwicklung</li> <li>- Umgang mit Schülerrisiken</li> <li>- Prävention von und Umgang mit Disziplinproblemen</li> </ul>	<p><b>Erziehen</b></p> <ul style="list-style-type: none"> <li>- Umgang mit Heterogenität</li> <li>- Individuelle Förderung</li> <li>- Vermittlung von Werten und Normen, Erziehung zu selbstbestimmtem Urteilen und Handeln</li> <li>- Finden von Lösungsansätzen für Schwierigkeiten und Konflikte</li> </ul>
<p><b>D: „Schulentwicklung“</b></p> <ul style="list-style-type: none"> <li>- Initiieren/ Erleichtern von Kooperation</li> <li>- Verständnis von Schulevaluation</li> </ul>	<p><b>Innovieren</b></p> <ul style="list-style-type: none"> <li>- Planung und Umsetzung schulischer Projekte und Vorhaben</li> </ul>
<p><b>E: „Professionsethos“</b></p> <ul style="list-style-type: none"> <li>- Akzeptanz der besonderen Verantwortung von Lehrkräften</li> </ul>	<p><b>Innovieren</b></p> <ul style="list-style-type: none"> <li>- Bewusstsein der besonderen Anforderungen des Lehrerberufs</li> <li>- Wahrnehmung des Berufs als ein öffentliches Amt mit besonderer Verantwortung und Verpflichtung</li> <li>- Wahrnehmung des Berufs als ständige Lernaufgabe</li> </ul>



Die Standards für die Lehrerbildung: Bildungswissenschaften unterscheiden für jede Unterkompetenz zwischen universitären sowie praktischen Ausbildungsabschnitten und vollziehen somit eine Trennung zwischen universitär vermitteltem Wissen und praktischem Können. Dass Wissen eine zentrale Komponente von professioneller pädagogischer Kompetenz ist, darüber herrscht breiter Konsens. Bezüglich des Transfers universitären Wissens in praktisches Handlungswissen bestehen jedoch erhebliche Zweifel (z. B. Blomberg, 2011; Helmke, Helmke & Schrader; 2007). Die von der KMK (2004) vorgenommene Unterscheidung vertieft diese Kluft weiter und übersieht zudem, dass „Können“ eine Kombination aus deklarativem und prozeduralem Wissen darstellt, welche im Lehramtsstudium auch gemeinsam vermittelt werden müsste (Kunter, 2011). Die Formulierung von Standards hätte hier eine Vorbildfunktion einnehmen können, indem Kompetenzen eben gerade nicht in universitäre und praktische Ausbildungsabschnitte hätten getrennt werden können.

Trotz der Kritik an heterogenen bzw. sich überschneidenden Kompetenzbereichen und der Unterscheidung zwischen universitär und praktisch vermittelten Kompetenzfacetten definieren die Standards für die Lehrerbildung: Bildungswissenschaften fundierte Zielkriterien für die Lehrerbildung und stellen aufgrund ihres hohen, international bestätigten Anforderungsbezugs einen verlässlichen Rahmen für professionelles pädagogisches Handeln.

## **1.5 Die Funktion von Bildungsstandards**

Die Formulierung von Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004) impliziert eine Professionalisierung des Lehrerberufs mit dem Ziel, Lehrkräfte auszubilden, denen definierte, auf einem Lehrleitbild begründete Kompetenzen zu eigen sein sollen (Terhart, 2002). Klieme und Rakoczy (2008) definieren Bildungsstandards als durch Bildungsziele legitimierte Instrumente der Professionalisierung.

Heidenreich (1999) beschreibt „Professionalisierung“ als die Systematisierung und Institutionalisierung von Berufsausbildungen und der Weiterentwicklung des betreffenden Wissensfelds. Zudem werden die betreffenden Tätigkeitsbereiche für die Angehörigen der entsprechenden Profession „reserviert“.

Oevermann (1996, zit. nach Bauer, 2000, S. 60) leitet die Professionalisierungsbedürftigkeit der Pädagogik daraus ab, dass pädagogisches Handeln ein Potential hinsichtlich „der Weichenstellung der Biographie von Schülern“ besitze. Nach Oevermann (1997) sehen sich Pädagogen mit dem Grundproblem konfrontiert, dass pädagogische Handlungen zwar den Lernenden anleiten, gleichzeitig aber die Autonomie des Lernenden nicht gefährden sollen. Ziele pädagogischer Einflussnahme seien hierbei primär die Wissensvermittlung, sekundär die Normvermittlung. Zwischen Lehrenden und Lernenden solle ein „Arbeitsbündnis“ geschlossen werden, das die Beziehung zwischen beiden strukturiert. Oevermann (ebd.) definiert die Beziehung zwischen Lehrendem und Lernenden hierbei sowohl als „diffus“ als auch als „spezifisch“: Zwar seien pädagogische Beziehungen auf bestimmte Ziele hin organisiert, zeitlich, räumlich und sächlich begrenzt, trotzdem weisen diese Beziehungen diffuse (nicht zielgerichtete, emotionale, personale) Anteile auf, die nicht als Fehlervariable, sondern vielmehr als notwendig für den pädagogischen Erfolg erachtet werden sollen.

Zusätzlich zu ihrer Funktion innerhalb der Professionalisierung stellen Standards ein Instrument der staatlichen Steuerung dar: Über die Erreichung der durch Bildungsstandards formulierten Zielkriterien müssen Bildungsinstitutionen Rechenschaft ablegen (Helmke & Schrader, 2006). Entscheidend für diese Neustrukturierung ist der Wandel von einer Input-Steuerung durch Gesetze, Lehrpläne, Erlasse, Stundentafeln, Ordnungen hin zur Output-Orientierung mit an Standards zu messenden Effekten und Wirkungen (Terhart, 2002).

Eine Output-Orientierung im Rahmen der Lehrerbildung impliziert die Erfassung von Kompetenzen und Kompetenzentwicklungen der Studierenden vom Beginn des Studiums bis zum Eintritt in den Lehrberuf. Hierdurch sind neben der Notwendigkeit neuer Steuerungsinstrumente aufgrund veränderter Rahmenbedingungen für Bildungssysteme (z. B. Einführung der Kosten-Leistungsrechnung und stärker ökonomisierte Strukturen) Prinzipien des Qualitätsmanagements angesprochen, welche sich mehr und mehr auch in Bildungsinstitutionen durchsetzen (müssen).

Um dem Anspruch eines Steuerungsinstrumentes genügen zu können, müssen Bildungsstandards bezüglich einer Ausprägung der durch sie genannten Kompetenzen überprüfbar und graduierbar sein (Klieme & Rakoczy, ebd.; Terhart, 2005) Hierfür ist es nötig, eine kriterienorientierte Skalierung zu implementieren, die

definiert, welche Kompetenzausprägung als ungenügend, minimal oder maximal gilt (Terhart, ebd.).

### Kritik an der Standardkonzeption

Bildungsstandards sehen sich einigen grundsätzlichen Kritikpunkten ausgesetzt. Diese beziehen sich zunächst auf die *fehlende theoretische Basis* und *theoriegeleitete Modellierung* oder *mangelnde Vollständigkeit* (Herzog, 2005). Die bereits oben erfolgte Diskussion der Standards für die Lehrerbildung: Bildungswissenschaften zeigte bereits, dass zumindest die Kritik an einer fehlenden theoretischen Basis hier nicht gelten kann, da zwar kein ausgewiesenes Modell, jedoch ein Bezug zu breiter empirischer und theoretischer Forschungsbasis gegeben ist (hierzu auch Oser, 2003; 2005). Bezüglich der mangelnden Vollständigkeit besteht weitgehend Einigkeit darüber, dass sich verschiedene relevante Kompetenzbereiche schlicht nicht in Standards überführen lassen (wie etwa Persönlichkeitsmerkmale; Frey & Jung, 2011; Oser, ebd.). Zudem erheben Standards keinen Anspruch auf Vollständigkeit, sondern seien offen konzipiert bzw. eine Diskussionsgrundlage für die Weiterentwicklung des Bildungssystems (Terhart, 2007).

Ein weiterer Kritikpunkt besteht in der Befürchtung einer *Normierung* und *Reduktion* von Bildungsprozessen und -inhalten. Die Standards für die Lehrerbildung: Bildungswissenschaften sind jedoch – wie bereits angemerkt – offen formuliert und lassen Raum für regional- und situationsangemessene konkrete Ausgestaltungen. Zudem stellen Standards lediglich Zielkriterien bezüglich übergreifender Kernbereiche dar, welche individuell erweitert und ergänzt werden können (Tulodziecki & Grafe, 2006).

Eine durch die Setzung von Standards implementierte *Output-Orientierung* stößt ebenso auf breite Kritik. Genügt es, alleine die Standarderreicherung der AbsolventInnen als einziges Kriterium zu betrachten? Nach Beck (2006) sollte diese Perspektive durch eine Input-, Throughput- sowie mehrstufige Output-Orientierung ergänzt werden: Der Autor schlägt vor, ebenso die Auswahl von Studierenden und den Eingang von Mitteln (Input), die Ausbildungsprozesse (Throughput) sowie die aufgrund der Ausbildung erreichten psychischen Dispositionen (Output Stufe I), Handlungen (Output Stufe II) und Produkte (Output Stufe III) zu focussieren. Auch Tulodziecki und Grafe (ebd.) halten eine zusätzliche Prozessperspektive für unabdingbar.

Um – ggf. neben anderen Kriterien – den an Standards ausgerichteten Output der Lehrerbildung überprüfbar zu machen, müssen die Standards für die Lehrerbildung: Bildungswissenschaften evaluierbar sein. Die Erfassung von Kompetenzen stellt jedoch die größte Herausforderung sowohl der Forschung zu professioneller pädagogischer Kompetenz als auch im Zuge des Standardkonzepts dar (z. B. Baumert & Kunter, 2011), da die Messung von Kompetenzen sich aufgrund des nötigen situativen Kontextes schwierig gestaltet. Adäquate Erfassungsinstrumente stehen bislang zwar für Teilkompetenzen, nicht jedoch für die Standards für die Lehrerbildung: Bildungswissenschaften zur Verfügung (Weiteres hierzu siehe Abschnitt 3.1.3).

Die Kritik der mangelnden theoretischen Basis sowie einer Normierung und Reduzierung von Ausbildungsinhalten und –prozessen kann für die Standards für die Lehrerbildung: Bildungswissenschaften abgewiesen werden. Zwar sind diese Standards ein Instrument im Rahmen einer berechtigterweise kritisierten eindimensionalen Output-Orientierung, dies tangiert jedoch nicht das Vorliegen der Standards per se, sondern die *Umsetzung der Standardevaluation*. Problematisch ist im Zuge dessen der Mangel an Instrumenten zur Überprüfung des Erreichens der durch die Standards für die Lehrerbildung: Bildungswissenschaften vorgegebenen Zielkompetenzen.

## 1.6 Die Evaluation der Lehrerbildung

Gleichwohl sollte die Lehrerbildung sich selbst jedoch Klarheit verschaffen, was sie eigentlich selbst erreichen *kann* und *will* – und wie sie sich selbst Rechenschaft darüber ablegt, ob sie das Gewollte auch *erreicht*. Dies für unmöglich, überflüssig und in jedem Fall: für gefährlich zu halten, tut der pädagogischen Sache keinen guten Dienst. Sich solchen Überlegungen zu entziehen, indem man sich die Rede vom „Technologiedefizit“ wie einen Schutzmantel umlegt, indem man Wirkungsdiffusität zum positiven Kennzeichen aller echten pädagogischen Maßnahmen erklärt und parallel dazu die Zurückweisung allein der Frage nach Wirkung und Wirkungsermittlung für die moralisch-kognitive Grundqualifikation von Erziehungswissenschaftlern hält, wird nicht weiterhelfen. (Terhart, 2005, S. 279)

Nach Terhart (2002, S. 5) wird durch die Formulierung von Standards für die Lehrerbildung das expliziert, was *implizit* „schon immer Bestandteil der Lehrerbildungsdiskussion war“: Traditionelle Vorstellungen über den/die „gute/n Lehrer/in“ führen über die aus diesem *Lehrerleitbild* abgeleiteten *Lehrerkompetenzen* zur Konzeptualisierung einer *Lehrerausbildung*, die zur Erreichung eben jener Kompetenzen führen soll. Im letzten – und relevantesten Schritt – soll ein/e „gute/r Lehrer/in“ mit den durch die entsprechende Lehrerbildung angebahnten Kompetenzen *höhere Lerngewinne bei seinen/ihren SchülerInnen* erzeugen als ein/e „weniger gute/r Lehrer/in“.

Letztendlich ergibt sich aus dieser Wirkungskette die Frage, inwiefern Lehrerbildung die intendierten Wirkungen – nämlich Lerngewinne auf Seiten der SchülerInnen – tatsächlich erzielt. Bei der Beantwortung dieser Frage können Cochran-Smith und Fries (2002) zufolge zwei verschiedenen Perspektiven eingenommen werden:

Der Ansatz der *Deregulierung* zum Einen focussiert die Evaluation des Lehrerhandelns („teacher evaluation“). Für den staatlichen Rückzug aus der Verantwortung für die Lehrerbildung wird hier deshalb argumentiert, weil in Anbetracht angeblich unklarer Wirkzusammenhänge zwischen Lehrerbildung, Lehrerhandeln und Schülerleistung ein staatlich definiertes und finanziertes Ausbildungsprogramm nicht zweckdienlich sei. Vielmehr sollen für Lehrpersonen lediglich Zulassungsstandards formuliert werden, deren Erfüllung unabhängig von der Ausbildung der Person zur Einstellung auf Probe berechtigen solle. Die Beurteilung der solcherart Eingestellten solle (größtenteils) aufgrund der erreichten Schülerleistung erfolgen, wobei unerheblich sein solle, *warum* ein/e Lehrer/in bessere Schülerleistungen erzeuge als ein/e andere/r – wichtig sei alleine der *Output*. Ballou und Podgursky (1999) fassen diese Perspektive prägnant zusammen: „Die Fähigkeit von Lehrern scheint eher eine Funktion angeborener Talente als eine Funktion der Lehrerbildung zu sein. Lehrer selbst sagen, dass dies so ist. ... Darüber hinaus scheinen Lehrer, die auf alternativem Weg zum Lehrerberuf kommen, ähnlich effektiv zu sein wie solche, die eine vollständige Ausbildung durchlaufen haben. Die legt es nahe, dass – verglichen mit anderen Faktoren – die Ausbildung nicht sehr viel zur Leistung von Lehrern beiträgt“ (ebd., S. 57)

Zum Anderen entspricht der Ansatz der *Professionalisierung* im Grunde der oben skizzierten impliziten Vorstellung einer Definition von Lehrerkompetenzen aufgrund

eines überkommenen Lehrerleitbildes und der daraus folgenden Gestaltung der Lehrerbildung. Diese Lehrerbildung (mit einer Kombination aus Wissenschaft und Praxis) solle qualifizierte Lehrkräfte hervorbringen, die wiederum durch ihr kompetentes Handeln die Lernleistungen ihrer SchülerInnen positiv beeinflussen sollen. Im Mittelpunkt des evaluativen Interesses steht hierbei die Lehrerbildung. Die aktuelle Forschungslage scheint hierbei für den Ansatz der Professionalisierung zu sprechen: „Quantitative Analysen zeigen, dass Maßzahlen für die Qualität der Lehrerausbildung und der Lehramtsprüfung bei weitem die stärksten Korrelationen mit Schülerleistungen in Lesen und Mathematik aufweisen. ... Diese Analyse legt nahe, dass solche Strategien, die in Lehrerbildung, –lizenzierung und Lehrerfortbildung investieren, einen wichtigen Beitrag zur Qualifikation und Kompetenz der Lehrerausbildung erbringen“ (Darling-Hammond, 2000, S. 1).

Trotz dieser ermutigenden Ergebnisse ist die kontrollierte Erfassung der Wirkungskette zwischen Lehrerbildung, Lehrerkompetenz und –handeln sowie Schülerleistungen sowohl theoretisch als auch methodisch äußerst diffizil und zeit- sowie ressourcenaufwendig.

Der Nachweis eines direkten kausalen Zusammenhangs zwischen den drei genannten Faktoren Lehrerbildung – Lehrerkompetenz – Schülerleistung ist aufgrund der Komplexität der Lehr-Lernsituation, nahezu unkontrollierbarer Umwelteffekte, der großen zeitlichen Spannweite und Anderem mehr kaum zu leisten. Sollten allerdings tatsächlich solche Zusammenhänge gefunden werden, so besteht die Gefahr, dass andere, weniger leicht messbare Effekte in der Lehrerbildung durch die Konzentration auf nachgewiesenermaßen erfolgreiche Größen vernachlässigt werden könnten (Terhart, 2002). Terhart (ebd.) empfiehlt somit, aufgrund formulierter Standards für die Lehrerbildung zunächst lediglich Zusammenhänge zwischen den ersten beiden Aspekten, Lehrerbildung und Lehrerkompetenz, zu ermitteln – ein Ansatz, der durch die von den Lehrerbildungszentren durchgeführte Evaluation studentischer Kompetenzentwicklung inzwischen in Ansätzen verfolgt wird, wobei Terhart allerdings konstatiert: „Die *Lehrerbildung* ist bislang in Deutschland noch nicht mit Blick auf ihre kurz- und langfristigen Wirkungen ernsthaft evaluiert worden.“ (ebd, S. 9)

Bis in die 60er-Jahre des letzten Jahrhunderts wurden Lehrpersonen insbesondere an normativen, idealistisch geprägten Leitbildern gemessen (z. B. Helmke, Helmke

& Schrader, 2007), die durch die formalen Standards staatlicher Lehramtsprüfungsordnungen ergänzt wurden. Durch die Wende der Pädagogik zur Erziehungs- und schließlich Bildungswissenschaft wandelte sich – zumindest offiziell – das Bild des/der Lehrers/in vom „ganzheitlich-erzieherisch motivierten“ (Terhart, 2002) Erzieher zu wissenschaftlich orientierten Lehr- und LernexpertInnen, was eine Abkehr vom personalistischen hin zu einem „technischen“ Ansatz implizierte, im Zuge dessen prinzipiell jeder die für die Ausübung des Lehrberufs notwendigen Kompetenzen erlernen kann. Hierdurch ergab sich selbstverständlich die Notwendigkeit der Formulierung neuer Richtmaße als Basis einer Evaluation, wie sie nun beispielsweise in Form der Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004) verbindlich vorliegen. Wenn diese Standards als Zielvorgaben gelten, so müssen zur Überprüfung der Zielerreichung Instrumente entwickelt werden, die eine Evaluation von pädagogischen Kompetenzen leisten können, und diese Instrumente müssen - um praktisch bedeutsam zu sein - auf einem Kompetenzmodell basieren, welches diese Standards beinhaltet.

Nach Jäger, Bodensohn und Frey (2008) müssen Hochschulen, welche zukünftige Lehrpersonen ausbilden, ein Qualitätsmanagement umsetzen, das im Sinne einer Qualitätssicherung quer- und längsschnittlich Rückmeldungen an Lehrende und Lernende beispielsweise über erreichte Kompetenzen, Leistungen bzw. die Lehrqualität liefert. Zudem soll ein Qualitätsmanagement die Entwicklung pädagogischer Kompetenzen der Studierenden sowohl an die Lernenden selbst als auch an steuernde Instanzen sowie die Wissenschaft rückmelden. Ein Instrument, welches pädagogische Kompetenzen sowie deren Entwicklung valide erfassen kann, entspricht folglich dem Ansatz des Qualitätsmanagements. König (2010) konstatiert jedoch, dass der Mangel an empirischen Untersuchungen zur Wirksamkeit der Lehrerbildung (und somit zur Kompetenzentwicklung der Lehramtsstudierenden) regelmäßig beklagt werde.

Im Bereich der *Personalbeurteilung und -entwicklung* liegt bereits eine Vielzahl von Konzepten und Instrumenten der Kompetenzdiagnostik vor (z. B. Erpenbeck & von Rosenstiel, 2007). Die deutsche empirische Forschung zur Kompetenzerfassung im pädagogischen Bereich schreitet voran, und so zeigen sich inzwischen erste Ergebnisse beispielsweise von Krauss, Kunter, Brunner, Baumert, Blum, Neubrand, Jordan & Löwen (COACTIV, 2004). Auch Seifert, Hilligus und Schaper (2009) liefern mit ihrem Messinstrument zur Erfassung pädagogischer Kompetenzen ein Beispiel für einen möglicherweise valideren Ansatz, welcher in Form einer Kombination von

Wissens- und Situational Judgement Test objektiver als self-assessment-Verfahren und reliabel die pädagogischen Kompetenzen von Lehramtsstudierenden erhebt. Hinsichtlich der Validierung stehen jedoch weitere Analysen noch aus. Zudem lässt sich der Bezug zu den KMK-Standards (2004) nur indirekt herstellen.

Zwar ist die internationale – und gerade die amerikanische – Forschung zur Erfassung professioneller pädagogischer Kompetenzen weitaus breiter aufgestellt (z. B. Wilson & Youngs, 2005), doch lassen sich deren Ergebnisse und Instrumente nicht ohne Weiteres auf den deutschen Sprachraum übertragen, da sich zum Einen in den USA der Weg zum Lehrberuf sehr heterogen gestaltet und zum Anderen das Niveau der Lehrerbildung in Deutschland weitaus höher liegt (Terhart, 2007).

Im deutschen Sprachraum wird professionelle pädagogische Kompetenz größtenteils durch Selbsteinschätzungsverfahren (z. B. Abs, 2005) oder retrospektive Befragungen hinsichtlich der Relevanz verschiedener Aspekte der Ausbildung geleistet ohne die Ebene des tatsächlichen Handelns zu tangieren (Seifert, Hilligus & Schaper, 2009), wobei die Validität von Selbst- und Fremdeinschätzungen im Rahmen der Kompetenzdiagnostik kritisch diskutiert wird (Nerdinger, Blickle und Schaper, 2008). Weinert und Schrader (1986) berichten beispielsweise, „dass erhebliche Zweifel an der Fähigkeit von Lehrern zur Selbstdiagnose ihres Unterrichts bestehen“. Auch Terhart (2009) bezeichnet die Ergebnisse solcher auf Selbstauskünften basierenden Studien als „unzuverlässig“ (S. 433) und fordert „ein Instrumentarium, mit dessen Hilfe man das Vorliegen bzw. den Ausprägungsgrad dieser erwünschten Kompetenzen identifizieren könnte.“ (ebd.)

Bezüglich der validen Erfassung allgemeiner pädagogischer Kompetenzen entlang der KMK-Standards besteht folglich weiterhin Bedarf, und auch Saterdag (2004) betont, dass Studierende im Laufe ihrer Ausbildung Hinweise bezüglich Ihrer Eignung und eines etwaigen Weiterbildungsbedarfs erhalten sollen. Diese Informationen können nur dann geliefert werden, wenn sie auf einer validen, reliablen und objektiven Erhebung der studentischen Kompetenzen beruhen. Nach Terhart (2007) gestaltet sich die Beurteilung professioneller pädagogischer Kompetenz von Lehrpersonen trotzdem schwierig,

- da im Kontext von „Schule“ eine Definition dessen, was „guter Unterricht“ sei, aufgrund normativer, emotional geprägter und konkurrierender



- Idealvorstellungen kaum zu leisten sei. Im Rahmen der empirischen Forschung besteht hier jedoch inzwischen ein breiter Konsens.
- da auf der Folie normativer Idealbilder die Möglichkeit der erfolgskritischen Erfassung pädagogischer Kompetenz allgemein angezweifelt werden könnte.
  - da die Aufgabenstruktur eines/r Lehrers/in sich äußerst komplex gestalte und somit sowohl schwierig zu erfassen als auch zu beurteilen sei.
  - da die Variablen, die das Kriterium „professionelle Kompetenz“ bedingen (könnten), vielfältig seien, teilweise außerhalb der Lehrperson selbst liegen (etwa Rahmenbedingungen) und hinsichtlich des Grades ihrer möglichen Veränderbarkeit noch Unsicherheit bestehe.
  - da bislang etwaige Kompetenzdifferenzen sich nicht in der beruflichen Laufbahn niederschlagen und somit zu fragen sei, warum eine Kompetenzbeurteilung überhaupt geleistet werden solle.
  - da sich der Lehrberuf hinsichtlich verschiedener Schulstufen und –formen sowie unterschiedlicher Fächerkombinationen und Fachdidaktiken sehr heterogen zeigt.

Während des Studiums und des Referendariats finden zwar sowohl an Hochschule und Universität als auch in den Praxisphasen durchaus Kompetenzbeurteilungen statt, bezüglich der Beurteilungspraxis des pädagogischen Handelns in den praktischen Phasen der Lehrerbildung durch FachleiterInnen existiert nach Terhart (2007) jedoch keine empirische Forschung.

Ein Beispiel für die Evaluation von Lehrerbildung findet sich bei Oser und Oelkers (2001): Absolventen des Lehramtsstudiums bzw. Lehrpersonen mit einem Jahr Berufserfahrung wurden von auf drei Dimensionen hinsichtlich ihrer subjektiven Einschätzung der 88 von den Autoren formulierten Standards (siehe Abschnitt 1.2) für die Lehrerbildung befragt:

1. Dimension: *Verarbeitungstiefe*. Hierbei wurde abgefragt, inwieweit die Standards im Wissen, Handeln und Reflektieren des Ausgebildeten verankert sind bzw. inwieweit die Standards angeeignet wurden.
2. Dimension: *Bedeutung*. Welche Bedeutung misst der Ausgebildete den Standards zu?
3. Dimension: *Beachtung*. Hierbei interessiert, im welchem Ausmaße der Ausgebildete die Standards im praktischen Handeln beachtet.

Die Auswertungsergebnisse von 1.286 Fragebögen zeigten sich wenig ermutigend: „Die Verarbeitungstiefe der meisten Standards ist gering, viele werden überhaupt nie oder nur ‚theoretisch‘ angesprochen.“ (Oser & Oelkers, 2001, S. 27) Dahingegen schätzen die Ausgebildeten sowohl die Bedeutung als auch die Anwendungswahrscheinlichkeit im Allgemeinen als hoch ein, wobei die schulbezogenen Standards (Gestaltung und Entwicklung von Schule) am schlechtesten abschneiden. Eine Erklärung hierfür könnte darin liegen, dass Absolventen und Berufsanfänger in nur geringem Maße die Möglichkeit erhalten, schon in der Schulentwicklung mitzuwirken. Dieses Argument kann insofern für sämtliche formulierte Standards gelten, als dass die Lehrerausbildung lediglich die *Basis* für kompetentes Lehrerhandeln schaffen kann und sich somit die Standards im Zuge des Weiterlernens im Beruf immer deutlicher ausprägen sollten.

Einschränkend muss zu dieser Studie bemerkt werden, dass nur Selbsteinschätzungsverfahren zum Einsatz kamen, welche – wie bereits dargelegt – zur Erfassung von Kompetenzen als eher ungeeignet gelten. Mit welchen Methoden kann eine Evaluation der Lehrerbildung tatsächlich kompetenzorientiert realisiert werden?

### Methoden der Kompetenzmessung

Nach Klieme und Hartig (2007) ist „die Verbindung von pädagogischen Konstrukten, psychologischen Kompetenzmodellen und Messverfahren ... eines der schwierigsten Probleme der Kompetenzforschung.“ (S. 24)

Die Ziele von Kompetenzmessungen bestehen zum einen darin, Bildungsprozesse zu diagnostizieren und auf Basis dieser Diagnostik Förder-, Platzierungs- und Auswahlentscheidungen treffen und rechtfertigen zu können (Jäger, 1999), zum anderen lassen sich durch Kompetenzmessungen Bildungsmaßnahmen und –institutionen evaluieren. Im Gegensatz zur Diagnostik klar umrissener Konstrukte (etwa Konzentrationsfähigkeit, mathematische oder sprachliche Kompetenz...) werden durch die Kompetenzdiagnostik komplexe, situations- und kontextabhängige Anforderungen erfasst. Für die Diagnostik von Kompetenzen reicht die Orientierung an kognitiven Fähigkeiten nicht aus. Während in der Pädagogischen Diagnostik oftmals konstruktneue Leistungs- und Persönlichkeitstests verwendet werden, die standardisiert und psychometrisch abgesichert bereits vorliegen und kontextunabhängig eingesetzt werden können,

müssen für Kompetenzmessungen zumeist Instrumente kontextspezifisch angepasst oder neu entwickelt werden, wobei das Ziel in der anforderungsbezogenen, ganzheitlichen Erfassung „relevanter Kompetenzfacetten“ (Schaper, Hilligus & Reinhold, 2009) besteht. Zudem beziehen sich Kompetenzmessungen eher auf Kriterien denn auf Normen (Schaper, 2009).

Die im Rahmen der Lehrerbildungsforschung bislang hauptsächlich eingesetzten Selbsteinschätzungsverfahren weisen im Bereich der Kompetenzeinschätzung lediglich begrenzte Validität auf, da beispielsweise einerseits soziale Erwünschtheit die Antworten verzerren könnte und andererseits eine von der konkreten Handlung abstrahierte Einschätzung eigener Kompetenzen nur eingeschränkt valide gelingt. Solcherlei Instrumente messen nach Hartig und Jude (2007) nicht die tatsächliche *Kompetenz*, sondern das *Selbstkonzept eigener Fähigkeiten*. Auch Tousignant und DesMarchais (2002) gelangen zu dem Schluss, dass die Selbsteinschätzung von Studierenden kaum mit ihrer tatsächlichen Leistung übereinstimmt.

Nach Helmke (2007) stellt die Videografie das Mittel der Wahl dar, um pädagogische Kompetenz im Feld analysieren zu können. Zwar ist die Unterrichtsaufzeichnung durch Kameras objektiv und hinsichtlich der Nutzung flexibel, jedoch ist die Anwendung der Methode sehr aufwendig und durch das Design der offenen Beobachtung störanfällig. Da sich Beobachtungsverfahren für große Stichproben (etwa für die Evaluation der Standards für die Lehrerbildung: Bildungswissenschaften) aufgrund des organisatorischen und zeitlichen Aufwands sowie der aufwendigen Auswertung nur eingeschränkt eignen, wird die Erfassung von Kompetenzen weiterhin größtenteils durch standardisierte, ökonomische Messverfahren (etwa mit Fragebögen) realisiert.

Terhart (2007) differenziert hinsichtlich den Instrumenten zur Kompetenzerhebung im pädagogischen Kontext zwischen

1. der Erfassung des schulpädagogisch-(fach)didaktischen Wissens und Urteilens von Lehrkräften
2. Selbsteinschätzungen der eigenen beruflichen Kompetenzen
3. der Einholung von Fremdbeurteilungen
4. der direkten Beobachtung und Beurteilung des beruflichen Handelns durch geschulte Beurteiler.

Zu 1.: *Lehrerwissen* bestimmt das Handeln nicht direkt (z. B. Neuweg, 2004). Trotzdem kann schulpädagogisches und (fach)didaktisches Wissen als wichtige professionelle Kompetenzdimension gelten, die mithilfe von Wissens- und Kenntnistests erfasst werden könnte. Die Erhebung der *Urteilskompetenz* könnte durch zu kommentierende und zu analysierende Videosequenzen, für die zusätzlich Handlungsempfehlungen ausgesprochen werden sollten, erfolgen.

Zu 2.: *Selbsteinschätzungen* der eigenen Kompetenz sind aufgrund bereits genannter Aspekte problematisch. Werden jedoch im Rahmen eines multiperspektivischen Designs Selbsteinschätzungen mit Fremdeinschätzungen (z. B. durch Vorgesetzte, KollegInnen, KundInnen oder SchülerInnen) kombiniert, stellen Selbsteinschätzungen eine Grundlage dafür dar, Diskrepanzen zwischen Selbst- und Fremdbild aufdecken und somit Veränderungsprozesse in Gang zu setzen zu können (Dries & Bovi, 2004).

Zu 3.: *Fremdbeurteilungen* der professionellen Kompetenz von Lehrpersonen können durch SchülerInnen, KollegInnen oder Vorgesetzte geleistet werden. SchülerInnen scheinen dabei recht zuverlässige BeurteilerInnen zu sein (z. B. Clausen, 2002). Die Beurteilung durch KollegInnen erfolgt bislang eher informell und unkontrolliert, wohingegen die Beurteilung von ReferendarInnen und aufstiegswilligen Mitgliedern des Lehrerkollegiums durch SchulleiterInnen aufgrund von Handreichungen, Formularen etc. formalisiert ist.

Zu 4.: Die direkte Erfassung professioneller Kompetenz von Lehrpersonen kann durch *Beobachtung* des Handelns im Unterricht realisiert werden. Hierbei sind eine sorgfältige Entwicklung des Instruments sowie die Überprüfung von dessen Reliabilität und eine Schulung der Beobachter vonnöten.

Universitäten und Hochschulen vermitteln den Lehramtsstudierenden fachwissenschaftliches, fachdidaktisches und erziehungswissenschaftliches Wissen. Dieses deklarative Wissen lässt sich durch *objektive Leistungstests* erfassen, bei denen freie, Multiple-Choice- oder Zuordnungs-Aufgabenformate vorgegeben werden. Diese Tests werden anhand von Codebooks, Musterlösungen für die freien Antwortformate und deren definierten Qualitätsstufen ausgewertet. Deklaratives Wissen wird außerdem durch die *Prüfungen* des Lehramtsstudiums abgefragt und stellt insofern eine Voraussetzung zur Ausübung des Lehrberufs dar.

Versteht man die Lehrerbildung ähnlich der Berufsbildung, da bei beiden Ausbildungen sowohl praktisch als auch theoretisch berufliche Kompetenzen vermittelt werden, so können die Bestimmungsstücke berufsbezogener Kompetenzdiagnostik von Achtenhagen und Baethge (2007) auf die Kompetenzmessung innerhalb der Lehrerbildung übertragen werden:

- „Messungen im Bereich der Berufsbildung sind immer *handlungs- und aktivitätsbezogen*.“ (S. 55) – Kompetenzen sollten folglich durch „state“-Variablen erhoben werden, da „trait“-Variablen Leistungsvariationen oder Kompetenzniveaus nicht ausreichend zu erklären in der Lage seien.
- „*Handlungen und Aktivitäten* im Bereich der Berufsbildung sind sehr spezifisch.“ (S. 55)
- „International bekannt gewordene Schemata zur Beschreibung beruflicher Kompetenzen ... berücksichtigen ... nicht hinreichend die Verschiedenheit von *Wissensdimensionen*.“ (S. 56) Achtenhagen und Baethge (2007) empfehlen hier die Unterscheidung deklarativen (Faktenwissen), prozeduralen (Wissensanwendung) und strategischen Wissens (inhaltsbasiertes Problemlösen und Planen).
- „Berufliche *Handlungen und Aktivitäten* erfordern Kompetenzen, die auf berufliche Aufgaben am Arbeitsplatz sowie auf die gegebenen sozialen Beziehungen hin ausgerichtet ist.“ (S. 56)

*Nicht-kognitive Teilkompetenzen* (z. B. motivationale, volitionale Orientierungen, sozial-kommunikative Fähigkeiten, Einstellungen und Überzeugungen) werden im Allgemeinen mit Selbst- oder Fremdeinschätzungsverfahren erfasst. Auch hier können allerdings sich Wahrnehmungsprobleme, Urteilstendenzen und Verzerrungen ergeben, wodurch die Validität dieser Verfahren ebenso eingeschränkt ist wie durch den meist mangelnden Kontextbezug.

*Beobachtungs-, Interview- oder Portfolioverfahren* können Kompetenzen zwar kontextnah erfassen, sind jedoch hinsichtlich der Testgütekriterien als problematisch zu werten. Gerade Portfoliotechniken sollten eher als Förderinstrumente denn als Kompetenzmessungen verstanden werden, da hier ein eher qualitativer und deskriptiver Zugang realisiert wird (Schaper, 2009).

Diehl (2005) fordert eine Kombination von qualitativen und quantitativen Methoden zur Erhebung pädagogischer Kompetenzen. Grundlegendes Wissen -

beispielsweise im Bereich Pädagogischer Diagnostik - könne durch Wissenstests erfasst werden, wohingegen die anwendungsorientierte Verknüpfung von Wissen und Handeln einerseits durch die Präsentation problemhaltiger pädagogischer Situationen, andererseits durch die Schilderung von Situationen aus der eigenen pädagogischen Praxis erhoben werden solle.

*Behavioral-Event-Interviews* (McClelland, 1998) als teilstrukturierte Verfahren erfassen Kompetenzen durch die Frage nach jeweils drei arbeitsbezogenen Episoden, in denen sehr erfolgreiches/nicht erfolgreiches Handeln erfolgte. Die ProbandInnen werden gebeten, zu jeder Episode zu berichten, was sie währenddessen taten oder sagten. Nach vereinbarten Kriterien wird anschließend ausgewertet, ob die einzelnen TeilnehmerInnen der Gruppe der „ExpertInnen“ oder der Gruppe der „Durchschnittskräfte“ angehören. Somit lassen sich folglich einerseits Kompetenzgrade erheben, andererseits lässt die Methode Schlüsse auf als von den Teilnehmenden relevant erachtete Kompetenzen zu.

Ein Kompromiss zwischen der Selbsteinschätzung im Rahmen einer ökonomischen paper-pencil-Erhebung und der ökologisch validen Unterrichtsbeobachtung lässt sich durch die Anwendung von *Situational Judgement Tests* (Weekley & Ployhart, 2006) schließen. Hierbei werden den ProbandInnen in Fragebogenform fiktive Szenarien vorgegeben, die einerseits analysiert werden und zu denen andererseits Handlungen zur Problemlösung generiert werden sollen. Pädagogische Kompetenzen werden mithin handlungsnah und auf der Ebene prozeduralen Wissens abgefragt. Die Auswertung von *Situational Judgement Tests* erfolgt meist aufgrund von Kriterien oder Musterlösungen (Ulbricht, Schaper & Hochholding, 2006). Dieses simulationsorientierte Verfahren wird in Kapitel 3.1 näher vorgestellt, da das im Rahmen der vorliegenden Studie entwickelte Instrument zur Methodenklasse der *Situational Judgement Tests* gehört.

Hinsichtlich einer *multimethodalen Kompetenzmessung* liegen zwar Ansätze innerhalb der Lehrerbildungsforschung vor (etwa Schmotz & Blömeke, 2009), Erkenntnisse bezüglich einer optimalen Kombination von Verfahren fehlen jedoch bislang ebenso wie kriterienbezogene Auswertungen und Interpretationen.

Die Erfassung pädagogischer Kompetenz lässt sich folglich im Rahmen verschiedener Erhebungsmethoden realisieren: (a) Self-Assessment-Verfahren focussieren auf die subjektive Einschätzung eigener Kompetenzen, (b)

Beobachtungsverfahren erheben die Performanz in natürlichen oder artifiziellen Settings, (c) biographische Methoden erheben Kompetenzen qualitativ beispielsweise durch Inhaltsanalyse oder Interviews, persönlichkeitsorientierte Ansätze (d) erheben etwa Motivstrukturen, Einstellungen und Überzeugungen, (e) Leistungstests erheben Kompetenzen im Sinne von zu lösenden Aufgaben und (f) simulationsorientierte Verfahren erlauben eine relative Annäherung an komplexe realitätsnahe Anforderungen (Seifert, Hilligus & Schaper, 2009).

Nach Schaper, Hilligus und Reinhold (2009, S. 3) sollten Verfahren der Kompetenzmessung innerhalb einer Domäne anforderungsbezogen und ganzheitlich die betreffenden Kompetenzfacetten erheben: „Für eine anforderungsbezogene Messung sind die Erhebungsverfahren daher in erster Linie gemäß ‚simulationsorientierter‘ eignungsdiagnostischer Ansätze (z. B. in Form von Arbeitsproben, situativen Übungen oder Situational Judgement Tests) zu gestalten.“ Aufgrund der Situations- und Kontextabhängigkeit von Kompetenzen (siehe Kap. 1.1) sollten sich Kompetenzmessungen idealerweise nicht auf Einzelbeobachtungen beschränken, sondern die Kompetenzausprägung eines Individuums im Rahmen verschiedener Aufgaben und Situationen erfassen.

Kritisch merkt Schaper (2009) jedoch an, dass solcherlei Instrumente häufig nicht auf empirischen Anforderungsanalysen beruhen, wodurch die inhaltliche Validität solcher Verfahren nur beschränkt gewährleistet wird.

### Die Stufen einer standardorientierten Evaluation

Die im Rahmen der genannten Methoden durchführbare Evaluation von Stufen der Erreichung vorgegebener Standards kann unterschiedlich feinkörnig gestaltet werden. Die im Folgenden dargestellten Vorgehensweisen bewegen sich entsprechend auf einem Kontinuum von „wenig aussagekräftig“ (Stufe 1) bis „sehr aussagekräftig“ (Stufe 4) (Terhart, 2002):

Stufe 1: *Selbsteinschätzung*. Die Befragten schätzen die eigene Erreichung von Standards subjektiv ein. Zwar ist dieses Verfahren leicht zu implementieren, zeigt sich jedoch äußerst störanfällig.

Stufe 2: *Testverfahren*. Wissen, Handlungs- sowie Reflexions- und Urteilskompetenzen werden hierbei durch kontextorientierte diagnostische Instrumente erfasst. Der Vorteil solcher Methoden besteht in der realitätsnäheren Erfassung der Standarderreichung, steht allerdings einem

hohen testdiagnostischen Vorbereitungs-, Durchführungs- und Auswertungsaufwand gegenüber. Zudem werden solcherlei Instrumente bislang in Deutschland kaum eingesetzt.

Stufe 3: *Beobachtung und Beurteilung*. Die (angehenden) Lehrpersonen werden in ihrem professionellen Handeln von Schulleitern, Ausbildern, Kollegen oder Ähnlichen beobachtet und hinsichtlich der Standarderreicherung beurteilt. Auch von den zu beurteilenden Personen erstellte Portfolios sollten in die Beurteilung einbezogen werden. Nachteilig wirken sich hier der große Aufwand hinsichtlich der Beobachterschulung, der Durchführung und Auswertung aus.

Stufe 4: *Lernleistung/Erfahrung der SchülerInnen*. Aufgrund von SchülerInnenbefragungen bzw. der Ermittlung von Schülerleistungen wird auf die Standarderreicherung der betreffenden (angehenden) Lehrperson rückgeschlossen. Problematisch ist hierbei die unsichere Wirkungskette zwischen Lehrerkompetenz sowie -handeln einerseits und Schülerlernen sowie -erfahrung andererseits.

In der vorliegenden Studie wird durch die Entwicklung eines Situational Judgement Tests die zweite Stufe der Feinkörnigkeit realisiert. Dieses simulationsorientierte Verfahren kann als ökonomischer Kompromiss zwischen Selbsteinschätzungsfragebögen und Beobachtungsverfahren gelten, da es einerseits eine relativ realitätsnahe Erhebung erlaubt und andererseits auch an großen Stichproben, wie sie sich im Rahmen einer Evaluation durch Lehrerbildungszentren ergeben, einsetzbar ist.

## 2. Fragestellung

In den bisherigen Ausführungen wurde der Bedarf an kontext- und standardbasierten Instrumenten der Kompetenzdiagnostik im Bereich der Lehrerbildung deutlich.

Ziel der vorliegenden Arbeit ist es, diese Lücke durch die Entwicklung eines adäquaten kontext- und standardbasierten Instruments zur Erfassung der professionellen pädagogischen Kompetenz Lehramtsstudierender in Form eines Situational Judgement Tests zu schließen.



Die Entscheidung für diese Methode fiel zum Einen aufgrund der Ökonomie des Verfahrens in Abgrenzung zu aufwendigeren handlungsnahen Instrumenten wie Beobachtungsverfahren, Assessment Centers oder videobasierten Tests. Entscheidend waren jedoch die vielversprechenden Forschungsergebnisse zur Validität von Situational Judgement Tests (McDaniel, Hartman, Whetzel & Grubb, 2007). So hat sich beispielsweise nach Mulder, Messmann und Gruber (2009) hinsichtlich der Erfassung professionellen Handelns insbesondere die individuelle Leistungsmessung beim komplexen Problemlösen in simulierten Szenarien (z. B. Planspiele, Simulationen) bewährt.

Die erste Fragestellung lautet folglich:

- Lässt sich entlang der Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004) ein inhaltsvalider Situational Judgement Test entwickeln?

Das zu entwickelnde Instrument basiert auf der durch die KMK geleistete Gliederung professioneller pädagogischer Kompetenz in vier Dimensionen bzw. elf Facetten. Aus dieser Basis leitet sich die zweite Fragestellung ab:

- Lassen sich die elf Unterkompetenzen der Standards für die Lehrerbildung: Bildungswissenschaften als Facetten der professionellen pädagogischen Kompetenz Lehramtsstudierender abbilden?

Für das im Zuge dieser Arbeit entwickelte Instrument stellt letztlich ein Professionalisierungsmodell im Sinne Galluzos und Craigs (1990, S. 603) den Rahmen: Ausgangspunkt dieses Modells sind die durch Lehrpersonen gesammelten Erfahrungen, welche über den aufgrund dieser Erfahrungen erfolgten Kompetenzgewinn auf Seiten des/der Lehrers/in zu Lernprozessen auf Schülerseite und schließlich zu Lernergebnissen führen sollen. Abbildung 2 stellt das Modell grafisch dar, wobei der grau hinterlegte Bereich die Variablen auf Lehrerseite kennzeichnet und sich somit auf die Wirkung der Lehrerbildung auf die Kompetenz des/der Lehrers/in bezieht.

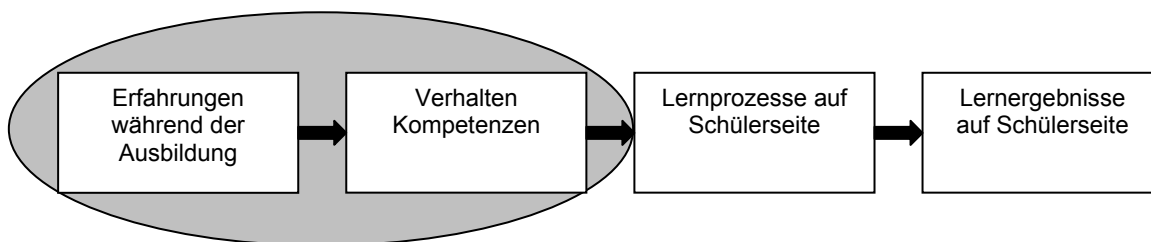


Abbildung 2. Das Wirkmodell der Lehrerbildung nach Galluzzo und Craig (1990), eigene Übersetzung

Dieser Teilbereich „Lehrerbildung“ des Modells wird in der vorliegenden Arbeit focussiert, wobei die latente Variable „Erfahrungen während der Ausbildung“ durch die manifesten Variablen „Anzahl besuchter Semester“, „Anzahl absolvierter Praktika“ sowie „Prüfungsleistungen“ operationalisiert wird. Die Variable „Verhalten, Kompetenzen“ wird durch die Items des Situational Judgement Tests erfasst, wobei die Kompetenzen durch die Standards der KMK (ebd.) konkretisiert werden.

### 3. Die Operationalisierung professioneller pädagogischer Kompetenz im Rahmen des diagnostischen Prozesses

Nach Jäger und Petermann (1999) ist psychologische Diagnostik „das systematische Sammeln und Aufbereiten von Informationen mit dem Ziel, Entscheidungen und daraus resultierende Handlungen zu begründen, zu kontrollieren und zu optimieren.“ (S. 11)

Unter dieser Prämisse kann der intendierte Situational Judgement Test – sofern sich der Test als valide zeigt – als diagnostisches Instrument gelten, weil hierdurch die Kompetenz angehender Lehrpersonen aufgrund eines Sammelns, Aufbereitens und Verarbeitens von Informationen diagnostiziert und gegebenenfalls adäquate Handlungen (wie etwa ein verpflichtendes zusätzliches Selbststudium) eingeleitet werden könnten.

Der diagnostische Informationsverarbeitungsprozess gestaltet sich äußerst komplex (ebd.): Zunächst muss eine Fragestellung ausgearbeitet werden, welche zugleich die Ziele des diagnostischen Prozesses impliziert. Die Fragestellung *der vorliegenden Studie* – „Lässt sich entlang der Standards für die Lehrerbildung:

Bildungswissenschaften (KMK, 2004) ein inhaltsvalider Situational Judgement Test entwickeln?“ wurde in Abschnitt 2 dargelegt, wobei bezogen auf den *diagnostischen Prozess* die Fragestellung lauten würde: „Weist Person X professionelle pädagogische Kompetenz im Sinne der durch die KMK (ebd.) formulierten Standards auf?“

Das Rahmen gebende *Ziel* der Studie besteht auf *Makroebene* darin, die professionelle pädagogische Kompetenz von Lehramtsstudierenden treffend diagnostizieren zu können und somit letztendlich im Sinne einer Selektionsstrategie eine Zuordnung von Personen zu definierten Gruppen (etwa im einfachsten Falle zu den Gruppen „Kompetent“ und „Nicht kompetent“) leisten zu können (Jäger & Petermann, 1999). Hierbei wird eine Statusdiagnostik realisiert, welche sowohl norm- als auch kriteriumsorientiert ausgestaltet werden kann. Auf *Mikroebene* müssen geeignete Items entwickelt werden, um die professionelle pädagogische Kompetenz von Lehramtsstudierenden adäquat operationalisieren zu können. Die folgenden Abschnitte legen nun den Prozess der Operationalisierung professioneller pädagogischer Kompetenz im Rahmen der Methodik eines Situational Judgement Tests dar.

### **3.1 Situational Judgement Tests**

Eine mögliche Methode der kontextbezogenen, realitätsnahen Kompetenzdiagnostik stellen Situational Judgement Tests dar. Bislang wurde dieses simulationsorientierte Verfahren vornehmlich im Rahmen der beruflichen Eignungsdiagnostik verwandt, und zudem zeigt sich die Forschungslage zu dem Verfahren gegenwärtig noch eher dünn (Weekley & Ployhart, 2006). Aufgrund der relativen Neuheit dieser Methode – gerade im Kontext der Evaluation von Lehrerbildung – soll dieses Verfahren in den nächsten Abschnitten vorgestellt werden.

#### **3.1.1 Definition: Situational Judgement Test**

Situational Judgement Tests basieren auf einem simulationsorientierten diagnostischen Ansatz. Kompetenzen werden hierbei anders als etwa bei Self-Assessment-Verfahren indirekt und situationsspezifisch erfasst (Bledow & Frese, 2009). Ziel ist es hierbei, eine „berufsrelevante Verhaltensstichprobe“ (Pixner & Schüpbach, 2008, S. 139) standardisiert zu erfassen.

Den ProbandInnen werden - meist in schriftlicher und standardisierter Form - Situationen vorgegeben, die im entsprechenden beruflichen Kontext üblicherweise auftreten können. Critical Incidents, Arbeits- und Anforderungsanalysen oder vorgegebene Kriterien (wie etwa geschaffen durch die Kompetenzbereiche der Bildungsstandards) steuern bei der Testentwicklung die Auswahl adäquater Situationen.

Die vorgegebenen Situationen werden durch die ProbandInnen „beurteilt“, indem diese beispielsweise die in einer Situation ihrer Meinung nach erfolg- oder mißerfolgversprechendste Handlungsalternative wählen oder hinsichtlich ihrer Effektivität einschätzen.

Zusammen mit den „Situationsvignetten“ werden den ProbandInnen mögliche Handlungsalternativen innerhalb der Situationen präsentiert, die von den TestteilnehmerInnen dergestalt „beurteilt“ werden, als dass sie entweder

- in einem Forced-Choice-Format die Handlung wählen, die sie am Ehesten (nicht) ausüben würden („Welche Handlung würden Sie wahrscheinlich (nicht) ausüben?“, vgl. Abbildung 3)

oder

- auf einer Likert-Skala einschätzen, wie effektiv mehrere Handlungsalternativen sind („Stufen Sie jede Handlungsalternative auf einer Skala ein von 1=sehr effektiv bis 5=sehr ineffektiv.“, vgl. Abbildung 4)

oder

- in einem freien Antwort-Format ihre favorisierte Handlungsalternative beschreiben („Beschreiben Sie in einem Satz, wie Sie in dieser Situation handeln würden.“).

Die Auswertung von Situational Judgement Tests erfolgt durch Antwortschlüssel oder Musterantworten, die aufgrund empirischer oder rationaler Erwägungen erarbeitet werden.

Einer Ihrer Angestellten berichtet Ihnen, dass er keine ausreichenden Ressourcen (Budget, Ausstattung u. Ä.) zur Erfüllung eines Auftrages habe.
<ul style="list-style-type: none"> <li>a) Sie weisen ihn an, wie er vorzugehen hat.</li> <li>b) Sie übertragen den Auftrag an einen anderen Angestellten, der keine solchen Einwände hat.</li> <li>c) Sie sagen zu dem Angestellten: „Tun Sie’s einfach.“</li> <li>d) Sie bitten den Angestellten um Alternativen und besprechen diese mit ihm.</li> <li>e) Sie statten den Angestellten mit mehr Ressourcen aus.</li> </ul>
Welche Antwort ist Ihrer Meinung nach die <i>beste</i> ?
Welche Antwort ist Ihrer Meinung nach die <i>schlechteste</i> ?

Abbildung 3: Beispielitem eines Situational Judgement Tests – Forced-Choice-Format (Weekley & Ployhart, 2006, S. 3; eigene Übersetzung)

Lange Zeit wurde im Bereich der Situational Judgement Tests lediglich sporadisch geforscht, bis sich zwischen 1999 und 2004 die Anzahl der betreffenden Forschungsarbeiten mehr als verdoppelte (Weekley & Ployhart, 2006).

Weekley und Ployhart (2006) führen diese wachsende Popularität auf Forschungsergebnisse zurück, die hinsichtlich der Validität von Situational Judgement Tests zuversichtlich stimmen (ausführlich zur Validität von Situational Judgement Tests siehe Kapitel 3.1.2).

### 3.1.2 Die Validität von Situational Judgement Tests

Die Diskussion über die Validität von Situational Judgement Tests begann mit Thorndikes und Steins (1937) Kritik am George Washington Social Intelligence Test (Moss, 1926, zit. nach McDaniel, Whetzel, Hartman, Nguyen & Grubb, 2006, S. 184). Dieser erste Test mit Items, in denen beispielsweise Situationen des Arbeitslebens beurteilt werden sollten, wurde als Test der sozialen Intelligenz konzipiert. Thorndike und Stein (1937) beanstandeten jedoch, dass der George Washington Social Intelligence Test nicht mit anderen Tests der *sozialen* Intelligenz, sondern mit Tests der *allgemeinen* Intelligenz korreliere.

Die Konstrukt- und kriteriale Validität von Situational Judgement Tests stehen seither im Blickpunkt. Nach McDaniel, Whetzel, Hartman, Nguyen und Grubb (2006) beeinflussen insbesondere der Grad der Fidelität, die kognitive Komplexität, der Inhalt der vorgegebenen Situationen sowie die Antwortinstruktionen eines

Situational Judgement Tests dessen Validität. Auf diese vier Aspekte soll daher im Folgenden näher eingegangen werden.

<p>Sie haben über mehrere Monate hinweg versucht, einen Termin bei einem wichtigen potentiellen Neukunden zu bekommen, aber Sie scheinen an seiner Sekretärin nicht vorbeizukommen. Die Sekretärin filtert alle Anrufe und sämtliche Post Ihres Vorgesetzten.</p>
<p>a) Versuchen Sie, vorbeizuschauen, wenn Sie in der Nähe sind, und sagen Sie, dass Sie auf den Chef warten werden.</p> <p>b) Sagen Sie der Sekretärin diplomatisch, dass Ihr Chef - eher als sie - die Entscheidung treffen sollte, ob er Sie sehen möchte.</p> <p>c) Schreiben Sie einen privaten/vertraulichen Brief an den potentiellen Neukunden, in welchen Sie die Situation erklären.</p> <p>d) Versuchen Sie, den potentiellen Neukunden früh morgens oder abends zu erreichen, wenn die Sekretärin nicht anwesend ist.</p> <p>e) Versuchen Sie, die Aufmerksamkeit des potentiellen Neukunden zu erhalten, indem Sie etwas Ungewöhnliches tun, wie etwa Blumen, Eintrittskarten oder ein „singendes Telegramm“ zu senden.</p>
<p>Bewerten Sie jede Antwortalternative anhand der folgenden Skala:</p> <p>6 = sehr effektiv</p> <p>5 = eher effektiv</p> <p>4 = etwas effektiv</p> <p>3 = etwas ineffektiv</p> <p>2 = eher ineffektiv</p> <p>1 = sehr ineffektiv</p>

Abbildung 4. Beispielitem eines Situational Judgement Tests – Likert-Skala (Weekley & Ployhart, 2006, S. 3; eigene Übersetzung)

## Fidelität

Im Rahmen von Situational Judgement Tests kann „Fidelität“ definiert werden als der Grad der Ähnlichkeit zwischen den innerhalb der Testitems formulierten fiktiven Situationen und den tatsächlich auftretenden Situationen. Die Fidelität videobasierter Tests wird hierbei höher eingeschätzt als die von paper-pencil-Verfahren – ebenso weisen Arbeitsproben eine höhere Fidelität auf als Tests der kognitiven Fähigkeit (McDaniel, Whetzel, Hartman, Nguyen & Grubb, 2006). Üblicherweise werden Situational Judgement Tests jedoch in der schriftlichen Form von paper-pencil- oder computergestützten Formaten eingesetzt, da deren

Entwicklung im Vergleich zu videobasierten Tests weniger kosten- und zeitaufwendig ist.

Bislang stehen eindeutige Forschungsergebnisse zur Beziehung zwischen der Fidelität und der Validität von Situational Judgement Tests noch aus (McDaniel, Whetzel, Hartman, Nguyen & Grubb, 2006), wobei die Annahme nahe liegt, dass die Fidelität sowohl deren Konstrukt- als auch von deren kriteriale Validität beeinflussen sollte.

### Kognitive Komplexität der Items

Situational Judgement Tests unterscheiden sich hinsichtlich der Komplexität der präsentierten Situations-Items. Wird die Situation mit nur einem kurzen Satz geschildert, liegt ein wenig komplexes Item vor, werden viele Sätze zur detaillierten Schilderung einer vielschichtigen Situation verwendet, so liegt ein hoch komplexes Item vor.

Da komplexe Situations-Items sowohl (im Falle von paper-pencil-Verfahren) eine höhere Lesefähigkeit als auch ein höheres Maß an schlussfolgerndem Denken erfordern, könnten solche Items höher mit kognitiver Fähigkeit korrelieren als Items mit weniger komplexen Situationsbeschreibungen. McDaniel und Nguyen (2001) gehen dabei davon aus, dass die Komplexität von Situationen mit der Länge des Situations-Items in Beziehung steht und dass es mehr Wörter erfordert, komplexe Situationen zu beschreiben als weniger komplexe Situationen.

McDaniel, Whetzel, Hartman, Nguyen und Grubb (2006) nehmen außerdem an, dass die Länge, Komplexität und Verständlichkeit von Situationsbeschreibungen korreliert sind und diese Faktoren die kognitive Belastung („cognitive load“; Sweller, 1988) bei der Bearbeitung der Situations-Items beeinflussen könnten. Zusätzlich erwarten die Autoren eine höhere kriteriale Validität derjenigen Situational Judgement Tests, welche einen höheren kognitiven „Load“ aufweisen, sofern es sich um Kriterien des Berufserfolgs bei kognitiv komplexen Tätigkeiten handelt, da die kognitive Fähigkeit selbstverständlich hierfür einen wichtigen Prädiktor darstellt (z. B. O'Connell, Hartman, McDaniel, Lee Grubb & Lawrence, 2007).

Situational Judgement Tests mit hohem kognitivem „Load“ könnten in Bezug auf Tests der kognitiven Fähigkeit nur wenig inkrementelle Validität aufweisen. Eine inkrementelle Validität eines Situational Judgement Tests kann folglich dann gegeben sein, wenn er als Test mit niedrigem kognitiven „Load“ Teil einer

Testbatterie mit Tests der kognitiven Fähigkeit ist, da der Situational Judgement Test dann zur Varianzerklärung jenseits der kognitiven Fähigkeit beitragen kann.

### Inhalt der vorgegebenen Situationen

Der Inhalt der von Situational Judgement Tests vorgegebenen Situationen variiert: So können zum Einen Situationen vorgegeben werden, die sich auf interpersonale Szenarien beziehen und somit kein berufsbezogenes Wissen erfordern. Zum Anderen können Situations-Items sowohl allgemeines (etwa Problemlösen) als auch spezifisches Wissen (z. B. pädagogisches Wissen) erfordern.

McDaniel, Whetzel, Hartman, Nguyen und Grubb (2006) postulieren einen Einfluss des Situationsinhaltes sowohl auf die Konstrukt- als auch auf die kriteriale Validität von Situational Judgement Tests. Die Beziehungen zwischen dem Inhalt von Items oder von Antwortoptionen und den zugrunde liegenden Konstrukten aufzudecken gestaltet sich den Autoren zufolge jedoch schwierig, da a) verschiedene Items verschiedene Konstrukte repräsentieren könnten und b) ein Item gleichzeitig mehrere Konstrukte repräsentieren könnte. Dies bedeutet, dass Items innerhalb von Situational Judgement Tests zum einen inhomogen bezüglich des zugrunde liegenden Konstrukts sein können, zum anderen Items nicht eindimensional sein können.

Auch Bledow und Frese (2009) vertreten die Ansicht, dass die Beantwortung eines Items aus der Interaktion zwischen Person und simulierter Situation resultiert. Die Autoren schlußfolgern: „This implies that situational aspects of each scenario influence responses, and therefore, ... items are not parallel measures of a unidimensional construct. Each scenario item ... measures a unique phenomenon.“ (S. 238) Situational Judgement Tests können demnach Kompetenzen meist nicht konstruktrein erfassen (Christian, Edwards & Bradley, 2010). Diese Aussage wird durch die Befundlage gestützt, dass der Metaanalyse von McDaniel und Whetzel (2005) zufolge keine veröffentlichte Faktorenanalyse eines solchen Testverfahrens mit interpretierbaren Faktoren vorliegt.

### Antwortinstruktionen

Üblicherweise werden in Situational Judgement Tests zweierlei Antwortinstruktionen eingesetzt: (1) Die TestteilnehmerInnen werden gebeten zu überlegen, wie sie in einer vorgegebenen Situation (nicht) handeln würden. (2) Die TestteilnehmerInnen werden angewiesen, die Effektivität verschiedener Antwortoptionen einzuschätzen.



Nach McDaniel und Nguyen (2001) lassen sich die Antworten auf Items mit wissensbasierten Instruktionen weniger leicht verfälschen als Items mit verhaltensbasierten Instruktionen. Dies lässt sich gut an einem Test mathematischen Wissens im Vergleich zu einem Persönlichkeitstest nachvollziehen: Sollen mathematische Aufgaben gelöst werden, so kann die Antwort eines/r Testteilnehmers/in lediglich aufgrund von dessen/deren Wissen (oder durch Raten) zustande kommen. Wird jedoch die Frage nach eigenem typischen Verhalten – wie beispielsweise in Persönlichkeitstest – gestellt, so können etwa soziale Erwünschtheit oder mangelnde Selbstreflexivität die Antwort verfälschen.

Antworten auf verhaltensbasierte Instruktionen sollten nach McDaniel, Whetzel, Hartman, Nguyen und Grubb (2006) mit Persönlichkeitsdispositionen korrelieren, wohingegen Antworten auf wissensbasierte Instruktionen mit kognitiver Fähigkeit korrelieren sollten.

Aufgrund der leichteren Verfälschbarkeit von Antworten auf verhaltensbasierte Instruktionen sollten Situational Judgement Tests mit solcherlei Instruktionen weniger valide als Situational Judgement Tests mit wissensbasierten Instruktionen sein. Zudem sollten Tests mit wissensbasierten Instruktionen aufgrund ihres höheren kognitiven „Loads“ und dessen Korrelation mit Berufserfolg eine höhere Validität bezüglich des Berufserfolgs aufweisen als Tests mit verhaltensbasierten Instruktionen.

McDaniel, Hartman und Grubb (2003a) gelangen in ihrer Metaanalyse bezüglich der Korrelationen von Situational Judgement Tests mit verhaltens- oder wissensbasierten Instruktionen und kognitiver Fähigkeit sowie den Big Five zu dem Ergebnis, dass Situational Judgement Tests mit verhaltensbasierten Instruktionen eher mit „emotionaler Stabilität“, „Verträglichkeit“ und „Gewissenhaftigkeit“ korreliert sind und Situational Judgement Tests mit wissensbasierten Instruktionen eher mit kognitiver Fähigkeit zusammenhängen.

### Zusammenfassung

Die kriterienbezogene Validität von Situational Judgement Tests zur Vorhersage professioneller Leistungen wird allgemein positiv beurteilt. So gelangten etwa McDaniel, Morgeson, Finnegan, Campion und Braverman (2001) in ihrer Metaanalyse zu einer mittleren Validität von  $\rho = .34$ . Zudem wird die hohe Augenscheinvalidität solcher Tests als geeignet angesehen, sowohl die Akzeptanz

unter den TestteilnehmerInnen zu erhöhen als auch berufsbezogene Verhaltensanforderungen an die TeilnehmerInnen zu kommunizieren, um somit Selbstselektionsprozesse in Gang zu setzen (Pixner & Schüpbach, 2008).

Da Situational Judgement Tests im Allgemeinen mehrere Konstrukte erfassen, lässt sich ein einzelner Test kaum einem einzelnen Konstrukt bei gleichzeitigem Ausschluss anderer Konstrukte zuordnen (McDaniel, Whetzel, Hartman, Nguyen & Grubb, 2006). Typischerweise korrelieren Situational Judgement Tests mit kognitiver Fähigkeit, Verträglichkeit, Gewissenhaftigkeit und emotionaler Stabilität, wobei Tests mit verhaltensbasierten Instruktionen höher mit Persönlichkeitsdispositionen und Tests mit wissensbasierten Instruktionen höher mit kognitiver Fähigkeit korrelieren. Tests mit wissensbasierten Instruktionen sind zudem weniger leicht verfälschbar.

Die Metaanalyse von McDaniel, Hartman und Grubb (2003b) zeigte eine höhere Korrelation von wissensbasierten Situational Judgement Tests mit kognitiver Fähigkeit und eine höhere Korrelation von verhaltensbasierten Situational Judgement Tests mit Persönlichkeitsdispositionen. Zudem gelangten die Autoren zu dem Ergebnis, dass Tests mit wissensbasierten Instruktionen eine etwas höhere kriteriale Validität als Tests mit verhaltensbasierten Instruktionen aufweisen.

Eine Vielzahl von Studien fasste die inkrementelle Validität von Situational Judgement Tests jenseits von Tests der kognitiven Fähigkeit in den Blick (z. B. Chan & Schmitt, 2002; McDaniel, Hartman & Grubb, 2003b). Alle Studien gelangen zu dem Schluss, dass Situational Judgement Tests hinsichtlich kognitiver Leistungstests inkrementelle Validität aufweisen. Der Grad der inkrementellen Validität ist dabei direkt abhängig mit dem kognitiven „Load“ des Situational Judgement Tests: Tests, die nicht-kognitive Konstrukte erfassen, weisen eine höhere inkrementelle Validität auf als Tests, die hoch mit kognitiver Fähigkeit korreliert sind. Hinsichtlich der inkrementellen Validität von Situational Judgement Tests jenseits von kognitiver Fähigkeit *und* Persönlichkeitsdispositionen steht weitere Forschung noch aus.

Bezüglich der Beziehung zwischen Situational Judgement Tests-Scores und berufsbezogenem Wissen nehmen McDaniel, Whetzel, Hartman, Nguyen und Grubb (2006) in einem heuristischen Modell einen direkten Einfluss von allgemeinem und spezifischem berufsbezogenem Wissen auf die individuell erreichten Test-

Scores an. Motowidlo, Hooper und Jackson (2006) vermuten in ihrem Modell des durch Situational Judgement Tests gemessenen prozeduralen Wissens, dass dieses Wissen einen kausalen Einfluss auf Berufserfolg habe. Dabei definieren die Autoren prozedurales Wissen als das Wissen über die Effektivität bestimmter Verhaltensweisen in berufsbezogenen Situationen, wie sie von Situational Judgement Tests vorgegeben werden. Werden nun Urteile über die Effektivität von Verhaltensweisen mit Antwortformaten wie der Wahl/Einschätzung einer Verhaltensweise als „am wahrscheinlichsten“/effektivsten oder „am unwahrscheinlichsten“/uneffektivsten operationalisiert, und wird prozedurales Wissen als die Übereinstimmung zwischen der (wahren) Effektivität jeder Antwortoption und der individuellen Einschätzung der Antwortoptionen operationalisiert, so kann prozedurales Wissen bewertet werden als die Summe der Übereinstimmungen zwischen den tatsächlich und gewählt effektivsten Alternativen minus der Summe der nicht übereinstimmenden Alternativen (tatsächlich effektiv/gewählt nicht effektiv).

In der vorliegenden Arbeit wird berufsbezogene Kompetenz dementsprechend als das prozedurale Wissen in einem bestimmten professionellen Kontext definiert und folglich ein Situational Judgement Test mit *wissensbasierten* Items generiert. Durch die Formulierung wissensbasierter Items wird zudem eine weniger leichte Verfälschungsanfälligkeit der Antworten angestrebt (siehe McDaniel, Whetzel, Hartman, Nguyen & Grubb, 2006).

Bislang besteht keine Klarheit darüber, wie Situational Judgement Tests idealerweise konstruiert, skaliert und bewertet werden sollten, um ein bestimmtes Konstrukt zu erfassen und ihre Reliabilität zu erhöhen. Zudem besteht bislang kein theoretischer Rahmen, innerhalb dessen Situational Judgement Tests verortet werden könnten. Theoretische Überlegungen anderer psychologischer Bereiche (wie etwa die Forschung zur Wahrnehmung von Situationen, zum Treffen von Entscheidungen oder Urteilen) könnten hilfreich sein, um Prinzipien zur Konstruktion von Situational Judgement Tests bereitstellen zu können (Ployhart & Weekley, 2006).

Auch hinsichtlich der Verfälschbarkeit, Trainierbarkeit, Kulturabhängigkeit oder der Nutzungswahrnehmung von Situational Judgement Tests durch die TestteilnehmerInnen besteht weiterhin Forschungsbedarf.

### 3.1.3 *Situational Judgement Tests in Abgrenzung zu anderen situationsorientierten Verfahren*

Situational Judgement Tests weisen offenkundige Parallelen zu anderen situations- und simulationsorientierten Verfahren auf (etwa Videovignetten oder Methoden der situativen Diagnostik), lassen sich jedoch ebenso deutlich von diesen abgrenzen. Exemplarisch seien im Folgenden das *Situationsinterview*, *Arbeitsproben* und *Assessment Centers* sowie *Videovignetten* vergleichend dargestellt.

#### Situationsinterviews

Im Rahmen eines Interviews werden dem/der Probanden/in Situationen vorgegeben, die sich bei der Ausübung des betreffenden Berufs ergeben können (Weekley & Ployhart, 2006). Die Antworten des/der Probanden/in werden durch den oder die InterviewleiterIn hinsichtlich ihrer Effektivität für eine erfolgreiche Bewältigung der Situation eingeschätzt, wobei zur Bewertung im Allgemeinen verhaltensbezogene Beurteilungsskalen eingesetzt werden.

Sowohl hinsichtlich der Struktur als auch der Validität ähneln sich Situationsinterviews und Situational Judgement Tests stark. Unterschiede zeigen sich bezüglich

- der Präsentation (verbal vs. schriftlich im Situational Judgement Test)
- der Beantwortung (verbal vs. Auswahl aus vorgegebenen Antwortalternativen im Situational Judgement Test)
- der Beurteilung (Interviewerurteil vs. Vergleich mit einem Auswertungsschlüssel im Situational Judgement Test).

#### Arbeitsprobe und Assessment Center

Arbeitsproben sowie Assessment Centers stellen Verfahren der Berufseignungsdiagnostik dar. Nach Schuler (1998) wird im Rahmen von Arbeitsproben Verhalten durch tätigkeits-spezifische, für die erfolgreiche Bewältigung eines bestimmten Berufs repräsentative Simulationsaufgaben erfasst. Ein Beispiel hierfür ist die „Drahtbiegeprobe“ (Lienert, 1967): Bei diesem Verfahren der handmotorischen Eignungsdiagnostik soll ein Draht entsprechend einer bestimmten Vorlage gebogen werden, wobei diese Leistung nach 29 Kriterien bewertet wird (z. B. Formrichtigkeit, Sauberkeit der Arbeit).

Assessment Centers sind nach Sarges (2001) eignungsdiagnostische Verfahren für den Managementbereich. Hierbei werden mehrere Kandidaten von geschulten Beobachtern in verschiedenen Beurteilungssituationen (z. B. Gruppendiskussionen oder Fallstudien) meist im Laufe mehrerer Tage nach explizierten Kriterien für den Berufserfolg (z. B. Initiative, Teamfähigkeit) beurteilt.

Aufgaben im Zuge von Assessment Centers wie Rollenspiele oder die „Postkorbübung“ sind ebenso wie Arbeitsproben situationsorientierte Methoden und weisen somit Parallelen zu Situational Judgement Tests auf. Arbeitsproben und Assessment Centers gehen jedoch über Situational Judgement Tests hinaus, indem sie die ProbandInnen zum Teil professionsbezogenen Situationen tatsächlich aussetzen. Hinsichtlich der ökologischen Validität lassen sich die drei Methoden auf einem Kontinuum anordnen: Situational Judgement Tests besitzen die geringste, Assessment Centers eine mittlere und Arbeitsproben die höchste Fidelität bezüglich tatsächlicher Anforderungen im Beruf.

Verhalten wird durch Arbeitsproben und Assessment Centers direkter erfasst als durch Situational Judgement Tests, wobei die ProbandInnen bei den ersten beiden Methoden durch Beobachtende beurteilt werden. Situational Judgement Tests weisen hier den Vorteil auf, automatisiert ausgewertet werden zu können. Im Vergleich zu Interviews, Arbeitsproben oder Assessment Centers lassen sich Situational Judgement Tests außerdem leichter in großen Stichproben implementieren und sind kostengünstiger.

### Videovignetten

Wird im Rahmen von Situational Judgement Tests der Itemstamm als textbasierte Beschreibung einer Situation dargeboten, so besteht bei der Testung mit Videovignetten der Itemstamm in einer Filmsequenz.

Nach Betrachtung der Videovignette werden zum Gesehenen Fragen gestellt, die in offenem (z. B.: „Welche Funktion hat die Sitzordnung der Schülerinnen und Schüler für den gezeigten Unterricht?“, König & Lebens, 2012) oder geschlossenem (z. B.: „Wann findet die Situation zeitlich betrachtet ungefähr statt? – Am Anfang/Im ersten Drittel/Im letzten Drittel/Am Ende der Unterrichtsstunde“, ebd.) Format beantwortet werden können.

Ein Beispiel für den Einsatz von Videovignetten im Kontext der Erfassung professioneller pädagogischer Kompetenz stellt die Studie COACTIV-Referendariat (z. B. Voss & Kunter, 2011) dar: Hier wurden zur Erfassung des pädagogisch-psychologischen Wissens – als Facette der professionellen pädagogischen Kompetenz – neben Multiple-Choice-Items und offenen Fragen auch videobasierte Items eingesetzt. Ziel letzterer war es, das Wissen über Klassenführung zu erheben – und dies mithilfe einer Methode, welche die „Komplexität von Unterrichtssituationen“ (ebd., S. 202) abzubilden erlaubt.

Die Vignetten bestehen aus 23- bis 78-sekündigen Filmsequenzen. Das filmische Material entstammt hierbei keiner tatsächlichen videographierten Unterrichtsstunde, sondern wurde von SchülerInnen der Klassen 5 bis 7 nachgespielt.

Das entwickelte Verfahren war in der Lage, statistisch signifikant zwischen den zwei Gruppen „LehramtskandidatInnen zu Beginn des Referendariats“ vs. „LehramtskandidatInnen am Anfang des zweiten Jahres des Referendariats“ zu diskriminieren, wobei allerdings die Effektstärke von  $d=.22$  auf eine eher geringe praktische Bedeutsamkeit hinweist.

Ein weiteres Beispiel für den Einsatz von Videovignetten bei der Testung der professionellen pädagogischen Kompetenz von Lehrkräften liefern König und Lebens (2012). Auch diese Autoren focussieren die Klassenführung (Classroom Management) und nähern sich ihr aus Perspektive kognitiver Anforderung, indem sie drei Dimensionen der „wissensgesteuerten Verarbeitung von Unterricht“ (ebd., S. 8) definieren: (1) Genauigkeit der Wahrnehmung, (2) Holistische Wahrnehmung sowie (3) Rechtfertigung einer Handlung.

König und Lebens (ebd.) entnahmen die verwendete Videovignette dem auf der Web-Site „Guter Unterricht“ frei verfügbaren Pool an Unterrichtsvideos (Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg). Die Filmsequenz dauert 81 Sekunden und wurde Lehramtsstudierenden des Hauptstudiums, ReferendarInnen sowie erfahrenen Lehrkräften dargeboten. Drei offene sowie eine geschlossene Frage mussten anschließend beantwortet werden.

Auch dieses Instrument deckte statistisch signifikante Unterschiede mittlerer praktischer Bedeutsamkeit zwischen den untersuchten Gruppen „Lehramtsstudierende“, „Referendare“ sowie „erfahrene Lehrkräfte“ auf.

Zuletzt sei der videobasierte Test „Observer“ (Seidel, Blomberg & Stürmer, 2010) erwähnt. Dieses Instrument wurde zur Erfassung der professionellen Wahrnehmung von Unterricht entwickelt und in Form eines computer- und onlinebasierten Anwendertools implementiert. Professionelle Wahrnehmung untergliedern die Autoren in zwei Komponenten: (1) noticing – Identifikation relevanter Situationen und Ereignisse im Unterrichtsgeschehen sowie (2) knowledge-based reasoning – wissensgesteuerte Verarbeitung identifizierter Situationen und Ereignisse.

Seidel et al. (ebd.) wählten aus 86 bestehenden Videosequenzen zwölf Clips aus. Für jeden Clip wurden drei standardisierte Rating-Items für die Komponente „knowledge-based reasoning“ entwickelt.

In der Pilotstudie wurde das Instrument ExpertInnen und 40 Lehramtsstudierenden vorgelegt, welche die Inhaltsvalidität der eingesetzten Videoclips (ExpertInnen und Studierende) sowie der Ratingsformate (ExpertInnen) einschätzten. Zudem wurden beide Gruppen im Rahmen eines Evaluationsbogens zum Instrument bezüglich der Akzeptanz der Items (z. B. „Die Fragen zu den Clips fand ich schwierig.“) und zum Gesamteindruck des Verfahrens befragt (z. B. „Das Programm ist ein geeignetes Mittel, um Unterricht zu analysieren.“).

Bei der Testapplikation erhielten die TeilnehmerInnen zunächst eine Instruktion mit den wichtigsten Hintergrundinformationen zur im Clip gezeigten Unterrichtssequenz. Dann wurde der Clip angesehen und hinsichtlich wahrgenommener Unterrichtskomponenten (Zielorientierung, Lernbegleitung und –atmosphäre) eingeschätzt. Anschließend wurde der Clip ein zweites Mal präsentiert und durch die Studierenden anhand von Rating-Items zum knowledge-based reasoning bewertet. Zuletzt bearbeiteten die TeilnehmerInnen einen Evaluationsbogen. Die gesamte Bearbeitungszeit des Instruments *Observer* betrug schließlich durchschnittlich 90 Minuten.

Die Inhaltsvalidität des Instruments sehen Seidel, Blomberg und Stürmer (2010) aufgrund eher übereinstimmender ExpertInnenurteile zur Repräsentativität der Videoclips als abgesichert. Allerdings schließen die Autoren aus einer Übereinstimmung der Repräsentativitätseinschätzungen zwischen 48,8% und 66,3%, dass die Darstellung komplexen Unterrichtsgeschehens durch standardisierte Instrumente schwierig sei.

Die Akzeptanz des Tests sehen Seidel et al. durch die positiven Einschätzungen im Rahmen des Evaluationsbogens als gegeben.

Die Erfassung professioneller pädagogischer Kompetenz mithilfe videobasierter Instrumente trägt der Forderung nach situations- und kontextorientierter Kompetenzdiagnostik Rechnung (z. B. König & Lebens, 2012). In der Literatur zur Diagnostik professioneller pädagogischer Kompetenz werden Tests mit Videovignetten deshalb häufig als „Mittel der Wahl“ zur realitätsnahen Erfassung von professioneller pädagogischer *Kompetenz* – in Abgrenzung zu *Wissen* – genannt (z. B. Borgmann, 2011; Voss & Kunter, 2011; König & Lebens, 2012). Situational Judgement Tests werden hingegen in einschlägigen Studien kaum erwähnt (einige Beispiele werden in Abschnitt 3.1.4 vorgestellt). Zur *Auswahl* Studierender werden Situational Judgement Tests freilich international bereits breiter eingesetzt:

Da der Einsatz kognitiver Studierfähigkeitstests in den USA in der Kritik steht, bestimmte Bevölkerungsgruppen zu benachteiligen (z. B. Sacco, Scheu, Ryan & Schmitt, 2000) und ethnische Gruppen sich bei den Ergebnissen kognitiver Tests stärker unterscheiden als bei den Ergebnissen in Situational Judgement Tests (z. B. Weekley & Ployhart, 2006), erlebt die Forschung zu dieser Methode der Studierendenauswahl hier einen starken Aufschwung. Zudem scheinen Situational Judgement Tests brauchbare Validitäten zur Vorhersage von Studienleistungen zu besitzen (Pixner & Schüpbach, 2008).

In Belgien wird ein Situational Judgement Test zur Auswahl von Studierenden der Medizin bereits realisiert (Lievens, Buyse & Sackett, 2005). Die inkrementelle Validität scheint dort hinsichtlich der Vorhersage von Studiennoten gegeben, sofern diese Noten nicht alleine akademisches Wissen spiegeln, sondern berufsrelevante Kompetenzen einbeziehen.

Dass der Einsatz von Videovignetten zur realitätsnahen Erfassung professioneller pädagogischer Kompetenz bislang auf recht ungeteiltes Interesse stößt, überrascht – besitzen diese Tests gegenüber Situational Judgement Tests doch einige Nach- oder zumindest keine Vorteile:

Dass Videovignetten gegenüber schriftlichen Situationsbeschreibungen bezüglich der Fidelität deutlich überlegen wären, kann kritisch gesehen werden: Auch im



Rahmen von filmbasierten Testitems kann lediglich ein kleiner Ausschnitt des Unterrichts präsentiert werden – in den oben genannten Studien bewegt sich der „Beobachtungszeitraum“ beispielsweise zwischen einer halben und eineinhalb Minuten – Helmke (2009) beschreibt eine Nutzung von kurzen Unterrichtsausschnitten freilich als „wenig informativ“ (S. 344) und verweist zusätzlich darauf, dass bei der Darbietung nur *einer* Perspektive (z. B. eine alleine auf die Lehrkraft gerichtete Kamera) ein Großteil des Unterrichtsverlaufs nicht rezipierbar sei. Um Situational Judgement Tests bezüglich der Fidelität zu übertreffen, müsste also entweder die Dauer der Videovignetten deutlich ausgedehnt werden oder aber es müssten schriftliche Zusatzinformationen (etwa hinsichtlich des Ziels der Unterrichtsstunde oder des gezeigten Zeitpunktes innerhalb einer Unterrichtssequenz) geliefert werden, was die Testdauer ebenso wie die Abhängigkeit von der Lesefähigkeit erhöhen würde. Außerdem müsste das Unterrichtsgeschehen multiperspektivisch aufgezeichnet werden, wodurch sich der Aufwand bei der Entwicklung von Videomaterial – Helmke (2009) folgend – erhöhen würde.

Die ökonomisch zwingende (sowohl hinsichtlich einer zumutbaren Testdauer als auch bezüglich des Entwicklungsaufwandes) Nutzung *kurzer* Videovignetten bei Tests der professionellen pädagogischen Kompetenz könnte somit gegenüber Situational Judgement Tests keinen Fidelitätsvorsprung aufweisen. Vergleichende Forschung steht allerdings noch aus.

Sollen Videovignetten eigens entwickelt werden, so gestaltet sich die Herstellung inzwischen zwar kostengünstiger als zu Beginn der Entwicklung von Aufzeichnungstechnologien (Helmke, 2009) – trotzdem ergeben sich höhere Kosten als bei der Entwicklung eines Situational Judgement Tests: Sollen tatsächlich gehaltene Unterrichtsstunden videographiert werden, so müssen ggf. große Mengen an Videos erstellt werden, um geeignetes Material (bestimmte Unterrichtssituationen) zu erhalten. Sollen bestimmte Unterrichtssituationen nachgestellt werden, so ist auch dies sehr zeitaufwendig: Skripte müssen erstellt, die „Schauspieler“ instruiert und schließlich die Situationen mehrmals dargestellt werden. Die Akzeptanz ebenso wie der Realitätsgehalt solcher Videovignetten ist zudem von der schauspielerischen Leistung der Darsteller abhängig. Im Falle solcher Spielsituationen könnte zudem die Fidelität durch den artifiziellen Charakter der nachgestellten Situationen leiden (vgl. Olson-Buchanan & Drasgow, 2006).

Probleme des Datenschutzes ergeben sich, sofern tatsächlich gehaltene Unterrichtsstunden aufgezeichnet werden sollen. Im Falle einer Nicht-Einwilligung von TeilnehmerInnen oder deren Eltern müssen nachträglich die betreffenden SchülerInnen nicht sichtbar und nicht hörbar gemacht werden, was entweder die Aufzeichnung erschwert (spezielle Kameraposition), zu nötigen Nachbearbeitungen führt oder eine zeitweise Platzierung der betreffenden SchülerInnen in Parallelklassen führt.

Auch kann sich die Unterrichtssituation durch das Vorhandensein einer Kamera verändern, sofern nicht ausreichend Zeit für eine Gewöhnung gelassen wird (Helmke, 2009).

Soll auf bereits existierendes Filmmaterial zurückgegriffen werden – inzwischen besteht eine große Anzahl auch frei verfügbarer Videos (z. B. TIMSS, Stigler, Gonzales, Kawanaka, Knoll & Serrano, 1999) – so muss eine Auswahl getroffen und das Material, falls möglich, weiter bearbeitet werden, um die Filme an das jeweils zu untersuchende Konstrukt anzupassen.

Ein weiterer Nachteil von videobasierten Tests ist die Schwierigkeit, sie auf große Stichproben anzuwenden, wie sie etwa bei der Evaluation der Lehrerbildung anfallen: Sofern den TestbearbeiterInnen die Möglichkeit der individuellen Bearbeitungsgeschwindigkeit und –modalität (erneutes Ansehen von Vignetten, Wahl der Lautstärke, „Zurück- oder Vorblättern“, ...) – wie bei schriftlichen Tests der Fall – erhalten bleiben soll, müssten Einzelarbeitsplätze mit Kopfhörern für die gleichzeitige Testung großer Stichproben zur Verfügung stehen – dies würde die Kapazität vieler Hochschulen übersteigen. Dieser Nachteil könnte durch die Präsentation der Vignetten auf Leinwänden verhindert werden – dies allerdings würde eine individuelle Bearbeitung verhindern, und zusätzlich müsste dafür gesorgt sein, dass sich Verständnis- oder Sichtprobleme nicht ergeben können. Stadler (2005) weist jedoch darauf hin, dass ein mehrmaliges Ansehen von Videosequenzen aufgrund der Fülle dargebotener Informationen nötig sei.

Die geschilderten Nachteile und Stolpersteine könnten den – möglicherweise auch nur vermeintlichen (s.o.) – Fidelitäts-Vorteil von Videovignetten gegenüber Situational Judgement Tests negieren. Allerdings bleibt dies Vermutung, da nach Kenntnis der Autorin keine Studien bestehen, welche Video- und textbasierte Vignetten direkt vergleichen.

Weitere Argumente gegen eine eingeschränkte Fidelität von Situational Judgement Tests liefern Stadler (2005) sowie Helmke, Helmke, Kleinbub, Nordheider, Schrader und Wagner (2007): Situational Judgement Tests können als Transkripte von Unterrichtssituationen verstanden werden. Stadler (ebd.) bezeichnet Transkripte als geeignet, um Interaktion und Sprache von Unterrichtssequenzen zu focussieren, und auch Helmke et al. (ebd.) sehen die Nutzung von Transkripten für die Einschätzung der Unterrichtsqualität als zweckdienlich.

Deutliche Vorteile von Situational Judgement Tests gegenüber Videovignetten zeigen sich bezüglich der Entwicklungs- und Anwendungsökonomie: Die oben geschilderte aufwendige Herstellung und/oder Auswahl von Filmmaterial entfällt ebenso wie schwierig zu realisierende Testszenarien – als paper-pencil-Verfahren können Situational Judgement Tests problemlos großen Stichproben dargeboten werden.

Ein Nachteil von Situational Judgement Tests liegt in der Abhängigkeit von der Lesefähigkeit der ProbandInnen. Ob letzteres schwerer wiegt als die Abhängigkeit vom Sprach- und Sprechverstehen beim Ansehen von Videovignetten, wäre zu untersuchen.

### *3.1.4 Situational Judgement Tests zur Erfassung professioneller pädagogischer Kompetenz*

Wie bereits oben geschildert, werden Situational Judgement Tests für die Erfassung professioneller pädagogischer Kompetenz bislang selten eingesetzt. Erste Ansätze bestehen jedoch inzwischen – ohne Anspruch auf Vollständigkeit werden im Folgenden einige Beispiele genannt.

#### Beratungskompetenz von Lehrkräften

Bruder, Keller, Kelava und Schmitz (2011) entwickelten entlang eines vierdimensionalen Modells der Beratungskompetenz von Lehrpersonen („Beratungskompetenz“, „diagnostisches/pädagogisches Wissen“, „Kollaboration/Perspektivenübernahme“ sowie „Coping“) einen Situational Judgement Test. Die dreizehn „Subskalen“ des Tests bestehen jeweils aus einem Item.

Die 13 Items wurden auf Basis des Modells der Beratungskompetenz bzw. basierend auf Erfahrungsberichten beratender Lehrender erstellt und anschließend durch ExpertInnen auf ihren Realitätsgehalt überprüft.

Der Test wurde als Multiple-Choice-Test mit vier Antwortalternativen konzipiert: In Bezug auf die beschriebene Situation wird gefragt, welches in den Antwortoptionen geschilderte Verhalten am wahrscheinlichsten und am unwahrscheinlichsten selbst gezeigt würde (Instruktion der bipolaren Verhaltenstendenz).

An einer Stichprobe von 78 GrundschullehrerInnen wurde die Itemanalyse durchgeführt. Die mittlere Itemschwierigkeit lag bei .70, wobei die Lösungswahrscheinlichkeit zwischen .53 und .84 schwankte.

An derselben Stichprobe wurde der Test theoriegeleitet an Außenkriterien (selbst eingeschätzte Praxisreflexion, Berufserfahrung, Anzahl an Beratungsgesprächen, Austausch mit Kollegen sowie Weiterbildung) validiert, wobei insbesondere das Kriterium „Praxisreflexion“ einen Prädiktor für den Gesamtttestscore darstellt. Die weiteren Ergebnisse der Validierung zeigen sich uneinheitlich.

### Pädagogische Kompetenzen in der universitären Lehrerbildung

Einen Test mit dem Ziel der Erfassung pädagogischer Kompetenz entwickelten Seifert, Hilligus und Schaper (2009). Ziel der Testentwicklung war es, den Kompetenzerwerb und die Kompetenzentwicklung während des Studiums nachzeichnen zu können und ein Instrument zur Studiengangs-Evaluation zur Verfügung zu stellen. Neben wissensbasierten Items wurden dabei auch Items in Form eines Situational Judgement Tests entwickelt.

Der Test basiert auf einem „Rahmenmodell zur Strukturierung pädagogischer Kompetenzen“ (ebd., S. 88), welches Inhaltsbereiche des Lehramtsstudiums der Paderborner Universität („Erziehung und Bildung“, „Unterricht und Allgemeine Didaktik“, „Schulentwicklung und Gesellschaft“) mit kognitiven Anforderungen („Wissen reproduzieren und verstehen“, „Reflektieren und anwenden“, „Urteilen, bewerten, entscheiden“) in einer Matrix verknüpft. Dabei wird zum Einen auf aktuelle Ansätze der Bildungsforschung wie diejenigen Brommes (1997) oder Kliemes et al. (2003) und auf einflussreiche Lehrerbildungsstudien wie COACTIV (ebd.) rekurriert, zum Anderen auf die Standards für die Lehrerbildung: Bildungswissenschaften der KMK (ebd.).

Für jede Zelle der aufgespannten Matrix wurden mehrere Items entwickelt, wobei hauptsächlich der Bereich „Urteilen, bewerten, entscheiden“ in Form eines Situational Judgement Tests operationalisiert wurde, da hier der Vorteil der Methode, komplexe Prozesse abbilden zu können, besonders zum Tragen komme (Seifert et al., 2009). Im Rahmen von Expertenworkshops entwickelten vier ExpertInnen für Didaktik, Lehrerbildungsforschung, Pädagogische Psychologie und diagnostische Forschungsmethoden die Testitems. Abbildung 5 zeigt ein Beispielitem.

<p>Im Rahmen von Lernerfolgskontrollen sollen bestimmte Gütekriterien berücksichtigt werden. Stellen Sie sich bitte folgende Situation vor:</p>
<p>Eine Deutschlehrerin hat die Klassenarbeit einer zehnten Klasse bewertet. Der Klassendurchschnitt liegt bei 2,1. Ihrem Mann, der auch Deutschlehrer ist, kommt das sonderbar vor. Er bewertet dieselbe Arbeit und kommt auf einen Schnitt von 3,5.</p>
<p>A In Bezug auf welches Gütekriterium liegt hier ein Problem vor? (Bitte antworten Sie in Stichpunkten)</p> <p>B Was kann die Lehrerin tun, damit andere Bewerter zu ähnlichen Ergebnissen wie sie kommen? (Bitte antworten Sie in Stichpunkten)</p>

Abbildung 5. Aufgabe EW20 (Seifert, Hilligus & Schaper, 2009, S. 93)

Da für die situationsbasierten Items keine Antwortoptionen vorgegeben, sondern ein offenes Antwortformat realisiert wurde, war durch die Autoren ein „Codebook“ zu erstellen. Das oben genannte ExpertInnenteam entwickelte Musterlösungen und mögliche Antwortbeispiele, welche mit Musterkodierungen versehen wurden, wobei verschiedene Qualitätsstufen definiert wurden.

Anhand der Daten aus einer Stichprobe von  $n=700$  wurden zunächst die Skalenkonsistenz und Itemkennwerte berechnet, wobei nicht zwischen Wissensitems und situationsbasierten Items differenziert wurde. Zur Einschätzung alleine des Situational Judgement Tests eignen sich die berichteten Ergebnisse also nur eingeschränkt.

Die Reliabilität des Gesamttests (Einzelskalen wurden nicht berichtet) beträgt  $\alpha=.868$ , die Schwierigkeiten aller Items liegen zwischen  $p=.04$  und  $p=.96$ .

Um erste Hinweise darauf zu erhalten, ob der Test in der Lage ist, Kompetenzentwicklungen während des Studiums aufzudecken, wurden t-Tests zwischen Studierenden des Grundstudiums und des Hauptstudiums gerechnet. Bei 40 von 111 Items schnitten erwartungsgemäß die Studierenden im Hauptstudium besser ab, bei 17 Items waren Studierende im Grundstudium besser. Seifert et al. (ebd.) schließen daraus, dass das Instrument zum Nachweis von Kompetenzzuwächsen nur eingeschränkt geeignet sei.

### Professionelle Handlungskompetenz (angehender) Physiklehrkräfte

Zur Operationalisierung und Messung der professionellen Handlungskompetenz Lehramtsstudierender entwickelten Riese und Reinhold (2008) eine Kombination aus situations- und anforderungsorientierten Wissenstests und Situational Judgement Tests. Ziel war auch hier der Nachweis von Effekten der Lehrerbildung.

Als theoretische Basis des Instrumentes fungiert zum Einen ein Kompetenzstrukturmodell aus den Facetten „Kognitive Fähigkeiten und Professionswissen“ (fachliches, fachdidaktisches sowie pädagogisches Wissen in Anlehnung an Shulman, 1986) sowie „Motivationale, volitionale und soziale Bereitschaften und Fähigkeiten“ (siehe Weinert, 2001), zum Anderen besteht ein Bezug zu den Kompetenzbereich „Unterrichten“ und „Beurteilen“ der KMK (ebd.). Zusätzlich wurden die Testinhalte inhaltlich auf den physikalischen Bereich der Mechanik beschränkt bzw. alleine der Anforderungsbereich des Experimentierens in den Blick genommen.

Die Entwicklung des Tests geschah literaturgeleitet. Die entwickelten Items wurden acht ExpertInnen (FachleiterInnen, erfahrene Lehrkräfte und FachdidaktikerInnen) zur Einschätzung der Vollständigkeit, Sinnhaftigkeit, inhaltlichen Relevanz und Repräsentativität vorgelegt. Das Antwortformat wurde offen gestaltet, sodass ein Codebook zur Auswertung der Tests zu erstellen war. Abbildung 6 zeigt ein Beispiel für eine „Unterrichtsvignette“ (Riese & Reinhold, 2008. S. 634). Die Aufgabe besteht im Beispiel daraus, suboptimales Verhalten des Lehrers zu analysieren und Miskonzeptionen von SchülerInnen zu erkennen.

Bei der Einführung des Prinzips „Actio=Reactio“ (9. Klasse) versucht der Lehrer, dies mit Hilfe einer Anordnung aus Feder und Gewicht zu demonstrieren. Es spielt sich folgende Szene ab:

Lehrer: Wenn ich das Gewicht an die Feder hänge, wird sie ein Stück ausgedehnt. Nehme ich das Gewicht weg und ziehe stattdessen mit einem Kraftmesse, dann muss ich mit etwa 10 N ziehen, damit die Feder genauso weit ausgelenkt wird. Das ist die Kraft, mit der das Gewicht an der Feder zieht. Wie ihr seht, muss ich mit derselben Kraft am Gewicht ziehen, damit es nicht nach unten fällt. Die Kraft, mit der die Feder am Gewicht zieht, ist genauso groß.

Schüler signalisieren Zustimmung.

Lehrer: Stellt euch jetzt einmal vor, ein Apfel hängt an einem Baum. Wo haben wir jetzt Actio und Reactio?

Schüler A: Na ist doch klar, der Apfel zieht am Ast und der Ast hält den Apfel oben.

Lehrer: Ja richtig – schön ihr habt es verstanden! Was ist denn dann, wenn der Apfel jetzt herunterfällt? Also während des Fallens, wo ist da Actio und Reactio?

Ein Gemurmel stellt sich ein.

Schüler B: Ja gilt das denn dann überhaupt noch? Ich meine, ist doch immer nur ideal, dass das gilt?!

Schüler A: Klar hast du noch Actio und Reactio, nur Actio wird halt immer größer, der Apfel wird ja schließlich schneller beim Fallen!

Schüler B: Ich dachte die müssen gleich sein? Wo willst du überhaupt Reactio haben, der fällt doch frei und wird nicht mehr gehalten?!

Schüler A: Hm. Na, Actio hast du auf jeden Fall schon mal, er bewegt sich ja. Und er wird ja auch nicht beliebig schnell, die Luftreibung bremst ihn ja. Das ist deine Reactio!

*Abbildung 6.* Beispiel einer Unterrichtsvignette (Riese & Reinhold, 2008, S. 634)

Das entwickelte Instrument wurde an 56 Lehramtsstudierenden der Physik pilotiert – die Reliabilität des Situational Judgement Tests lag bei .89, weitere Kennwerte werden nicht berichtet.

### Fachbezogenes Professionswissen von Mathematiklehrkräften

Krauss, Blum, Brunner, Neubrand, Baumert, Kunter, Besser, und Elsner (2011) entwickelten im Rahmen des COACTIV-Forschungsprogramms einen erneut auf Shulmans (ebd.) Wissenstaxonomie basierenden Test des fachdidaktischen Wissens von MathematiklehrerInnen. Für die Facetten „Zugänglichmachen

mathematischer Inhalte für SchülerInnen: Wissen über Erklären und Repräsentieren“ (12 Items) sowie „Schülerinnen und Schüler: Wissen über typische SchülerInnenfehler und -schwierigkeiten“ (7 Items) wurden Unterrichtsszenarien mit offenen Antwortformat entwickelt.

Allerdings können nur einige Items dieser Skalen als Situational Judgements im tatsächlichen Sinne gelten, da nicht alle Items der Skalen einen direkten Situationsbezug aufweisen. Abbildung 7 stellt zwei Beispiele gegenüber.

Item mit Situationsbezug	Item ohne direkten Situationsbezug
<p>Eine Schülerin sagt: Ich verstehe nicht warum <math>(-1) \cdot (-1) = 1</math> ist. Bitte skizzieren Sie kurz möglichst viele verschiedene Wege, mit denen Sie der Schülerin diesen Sachverhalt klarmachen könnten.</p>	<p>Die folgenden Formeln liefern alle den Flächeninhalt eines Trapezes. <math>(g_1 + g_2) \cdot h/2</math> <math>(g_1 \cdot h)/2 + (g_2 \cdot h)/2</math> <math>((g_1 + g_2) \cdot h)/2</math> <math>((g_1 + g_2)/2) \cdot h</math></p> <p>Welchen didaktische Nutzen kann die Betrachtung dieser einzelnen Formeln haben? Begründen Sie Ihre Antwort.</p>

Abbildung 7. Itembeispiele aus dem COACTIV-Test zur Erfassung fachdidaktischen Wissens (Krauss et al., 2011, S. 140)

Zur Auswertung wurde unter Zusammenarbeit von Fachdidaktikern und Mathematiklehrkräften bzw. unter Nutzung erster im Rahmen der Pilotierung entstandenen LehrerInnenantworten ein Kodierschema erstellt.

Da der entwickelte Test nicht als „reiner“ Situational Judgement Test gelten kann, soll auf eine weitere Darlegung an dieser Stelle verzichtet werden.

## Resümee

Dass inzwischen zumindest vereinzelt die Methode des Situational Judgement Tests zur Erfassung professioneller pädagogischer Kompetenz eingesetzt wird, weist auf ein steigendes Interesse an dem Verfahren hin.



Die oben beschriebenen ersten Testentwicklungen zeigen allerdings eine Reihe insbesondere methodischer Schwächen:

- *Bruder, Keller, Kelava und Schmitz* (ebd.) entwickelten für ihre 13 Skalen jeweils nur ein Item, wodurch sowohl die Reliabilität des Tests als auch die Varianz eingeschränkt sind.
- Der verwendete Antwortschlüssel (2 Punkte bei vollständig richtiger Antwort, 1 bei teilweise richtiger Antwort, 0 bei falscher Antwort) macht eine ökonomische, automatische Auswertung der Tests schwierig.
- Die entwickelten Items mit einer mittleren Itemschwierigkeit von .70 zu leicht und können somit nicht ausreichend zwischen niedriger und hoher Kompetenzausprägung diskriminieren.
- Zuletzt konnten Validitätsnachweise nicht ausreichend erbracht werden, wobei ein Außenkriterium in einem Selbsteinschätzungsverfahren (Praxisreflektion) bestand – dies ist problematisch, da Situational Judgement Tests und Self-Assessment-Verfahren verschiedene Konstrukte messen.
- Die in der Iteminstruktion durch *Seifert, Hilligus und Schaper* (2009) erfolgte Erwähnung des jeweils interessierenden Konzept (im Beispiel oben „Gütekriterien“) liefert den ProbandInnen unnötige Hinweise, wodurch die Schwierigkeit der Items herabgesetzt sein könnte.
- Das offene Antwortformat macht eine ökonomische, automatische Auswertung unmöglich.
- Auch hier wurden Validitätsnachweise nur unvollständig erbracht.
- Die Items des von *Riese und Reinhold* (ebd.) entwickelten Tests sind sehr textlastig. Die Testleistung könnte also mit der Leseleistung stark konfundiert sein.
- Die bislang untersuchten Stichproben sind teilweise zu klein, um eine Generalisierung der Ergebnisse zu erlauben (*Bruder, Keller, Kelava & Schmitz, 2011; Riese & Reinhold, 2008*).
- Häufig werden verschiedene Itemarten vermischt, sodass dann Aussagen über die Validität der Methode „Situational Judgement Test“ nicht möglich sind (*Krauss et al., 200; Riese & Reinhold, 2008*).

Nicht zuletzt ist keines der beschriebenen Instrumente dazu geeignet, eine Erfassung professioneller pädagogischer Kompetenzen sämtlicher eng entlang sämtlicher von der KMK (ebd.) formulierter Standards für die Lehrerbildung:

Bildungswissenschaften zu leisten, da zumeist nur Teilbereiche der Standards erfasst werden oder die Standards den Instrumenten lediglich implizit zugrunde liegen.

#### Zusammenfassung: Warum Situational Judgement Tests?

Die Methode der Situational Judgement Tests wurde den verwandten situationsorientierten Verfahren des Situationsinterviews, der Arbeitsprobe, des Assessment Centers sowie der Testung mit Videovignetten gegenübergestellt.

All diesen Methoden ist ein simulationsorientiertes Ansatz gemein – Kompetenzen sollen realitätsnah und kontextbezogen erfasst werden.

All diese Methoden erfüllen somit die Forderung nach einer Kompetenzdiagnostik, welche „besonders eng mit der tatsächlichen Bewältigung spezifischer beruflicher Anforderungen verknüpft ist“ (König & Lebens, 2012, S. 4).

Besteht das Ziel allerdings in der Erfassung professioneller pädagogischer Kompetenz als Evaluation der Lehrerbildung, so müssen bei der Wahl der Methode spezifische Rahmenbedingungen mit bedacht werden: Soll die Wirksamkeit der Ausbildung zukünftiger Lehrkräfte nachgewiesen werden, so fallen große Stichproben an. Situationsinterviews sowie Assessmentcenters lassen sich mit Studierendenkohorten aus ökonomischen Gründen weder durchführen noch auswerten. Die Testung mit Videovignetten zeigt sich ebenso wenig ökonomisch aufgrund der aufwendigen Entwicklung und Testdarbietung. Arbeitsproben werden im Rahmen von Praktikabeurteilungen sowie während des Referendariats zwar realisiert, doch sind diese Beurteilungen nicht ausreichend mit der universitären Bildung verknüpft (siehe die viel beklagte Theorie-Praxis-Schere).

Situational Judgement Tests jedoch weisen zum Einen die geforderte Kontextnähe sowie Kompetenzorientierung auf und gelten als valide Prädiktoren für Verhalten in verschiedenen Situationen (McDaniel, Morgeson, Finnegan, Campion & Braverman, 2001), zum Anderen können sie problemlos großen Stichproben dargeboten werden. Der Entwicklungs- und Auswertungsaufwand ist dabei deutlich geringer als bei Situationsinterviews, Assessmentcentern und Arbeitsproben.

Insbesondere mit videobasierten Tests stehen Situational Judgement Tests „in Konkurrenz“. Doch auch Videovignetten werden aufgrund eines geringeren

Entwicklungsaufwands bzw. der Einfachheit der Testdurchführung übertroffen. Dass die Fidelität von Situational Judgement Tests gegenüber Tests mit Videovignetten eine geringere sei, kann bezweifelt werden, sofern die Videovignetten eine für die Testung großer Stichproben nötige Kürze aufweisen – ein Hinweis darauf findet sich bei Seidel, Blomberg und Stürmer (2010): Die Autoren befragten im Rahmen ihrer Entwicklung eines videobasierten Testinstruments Lehramtsstudierende sowie ExpertInnen hinsichtlich dessen, ob beide Gruppen die Länge der eingesetzten Videosequenzen als zu kurz einschätzten. Sowohl Studierende als auch ExpertInnen empfanden die Länge im Mittel als eher zu kurz. Forschung zum direkten Vergleich der Methoden „Video- vs. textbasierte Vignetten“ steht allerdings noch aus.

Gegenüber Selbsteinschätzungsverfahren besitzen Situational Judgement Tests den deutlichen Vorteil des Situations- und Kontextbezugs. Zudem ist die Akzeptanz von Situational Judgement Tests auf Seiten der ProbandInnen im Vergleich zu Selbsteinschätzungsverfahren höher (z. B. Clevenger, Pereira, Wiechmann, Schmitt, & Schmidt-Harvey, 2001).

Einiges weist also darauf hin, dass Situational Judgement Tests eine valide und ökonomische Methode zur Erfassung professioneller pädagogischer Kompetenz darstellen könnten. Die vorliegende Arbeit möchte einen Beitrag zur Beantwortung der Frage leisten, ob sich Situational Judgement Tests tatsächlich als Instrument der Evaluation von Lehrerbildung eignen.

### **3.2 Generierung der Items**

Die im Rahmen der vorliegenden Arbeit vorgenommene Entwicklung des Situational Judgement Tests wurde unter Einbezug von ExpertInneneinschätzungen in vier Schritten realisiert: Zunächst wurden als Grundlage geeignete Items aus dem standardbasierten Instrument zur Evaluation der Lehrerbildung „KOSTA“ („Kompetenz- und Standardorientierung in der Lehrerbildung“; Weresch-Deperrois, Bodensohn & Jäger, 2009) des Zentrums für Lehrerbildung der Universität Koblenz-Landau, Campus Landau, ausgewählt.

Diese Items wurden von Aussagen (z. B. „Ich handle als Lehrkraft meiner Vorbildfunktion entsprechend.“) zu qualitativen Fragestellungen („In welchen

Situationen handeln Sie als Lehrkraft Ihrer Vorbildfunktion entsprechend (Beispiele)?“) modifiziert und Studierenden verschiedenen Semesters zur offenen Beantwortung vorgelegt, um zu realistischen Handlungsalternativen im Unterricht auf der Kompetenzstufe von „studentischen NovizInnen“ – also mithin zu realistischen „Manifestationen“ professioneller pädagogischer Kompetenz im unteren Kompetenzbereich – zu gelangen.

Zudem wurden Handlungsalternativen auf der Kompetenzstufe von „ExpertInnen des Lehrens“ gemeinsam mit ebensolchen ExpertInnen entwickelt. Das Vorgehen, NovizInnen und ExpertInnen in die Itementwicklung einzubeziehen, hatte zum Ziel, ein möglichst breites Kompetenzspektrum reflektieren zu können. Die auf diese Weise generierten Items des Situational Judgement Tests wurden schließlich von ExpertInnen hinsichtlich ihrer Eignung geratet.

Die vier Entwicklungsschritte sollen nachfolgend eingehender dargestellt werden.

### 3.2.1 1. Schritt: Auswahl geeigneter Items aus dem Instrument KOSTA

Als Basis für die Itementwicklung wurden die Items des Self-Assessment-Instruments KOSTA („Kompetenz- und Standardorientierung in der Lehrerbildung“, Weresch-Deperrois, Bodensohn & Jäger, 2009) genutzt:

Für das Zentrum für Lehrerbildung der Universität Koblenz-Landau, Campus Landau, wurde mit KOSTA (ebd.) entlang der elf Standards für die Lehrerbildung: Bildungswissenschaften (siehe Kapitel 1.3) ein Instrument entwickelt, welches zu jedem dieser elf Kompetenzbereiche zwischen fünf und acht konkrete Kompetenzen der Unterrichtsgestaltung zur mehrdimensionalen Einschätzung durch Lehramtsstudierende vorgibt. Dieser Fragebogen wird von den Studierenden nach jeder der durch das Lehramtsstudium geforderten Praxisphasen sowie dem Referendariat beantwortet, um somit in einem Längsschnitt-Design die Kompetenzentwicklung der Lehramtsstudierenden verfolgen zu können.

Die Itementwicklung wurde hierbei auf Basis einschlägiger Literatur der Pädagogischen Psychologie (z. B. Helmke, Helmke & Schrader, 2007; Rheinberg, Bromme, Minsel, Winteler & Weidenmann, 2001; Brophy, 1999; Grouws & Cebulla, 2000; Reinmann-Rothmeier & Mandl, 2001; Schröder, 2007; Rheinberg, 2006; Oser, 2006; Wild & Krapp, 2001; Tent, 2006; Schnabel, 2001; Perrez, Huber &

Geißler, 2001; Wild, Hofer & Pekrun, 2001) sowie im Rahmen von ExpertInnenrunden mit Diplom-PsychologInnen, Diplom-PädagogInnen sowie Studierenden der Erziehungswissenschaften und des Lehramtes der Universität Koblenz-Landau, Campus Landau, realisiert.

In jedem Item wird für die jeweils genannte Kompetenz aus der Sicht der Studierenden abgefragt,

- wie wichtig die genannte Kompetenz für den gehaltenen Unterricht war
- wie häufig die Kompetenz angewandt wurde
- wie schwer es fiel, die Kompetenz im Unterricht umzusetzen
- welcher Bedeutung der Kompetenz innerhalb der universitären Ausbildung zukommen soll
- wie gut die Universität auf die genannte Kompetenzausübung vorbereitet.

Zudem wird ein einem freien Antwortformat eine Begründung erfragt, falls die genannte Kompetenz in der Praxisphase nicht angewandt werden konnte. Abbildung 8 zeigt ein Beispiel.

<b>Kompetenzbereich 1:</b>		<b>Lehrerinnen und Lehrer planen Unterricht fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.</b>	
<i>1. Ich strukturiere meinen Unterricht klar.</i>			
a	Welcher Stellenwert kommt der Kompetenz aus meiner Sicht zu?	(1) sehr wichtig ... (6) völlig unwichtig	1 2 3 4 5 6
b	Wie häufig habe ich die Kompetenz angewendet?	(1) sehr oft ... (6) nie	1 2 3 4 5 6
c	Wie schwer ist es mir gefallen, die Kompetenz im Unterricht anzuwenden?	(1) sehr leicht ... (6) sehr schwer	1 2 3 4 5 6
d	Welche Bedeutung sollte die Kompetenz innerhalb der Ausbildung der Universität einnehmen?	(1) sehr große ... (6) sehr geringe	1 2 3 4 5 6
e	Wie gut bin ich von der Universität auf diese Kompetenz vorbereitet?	(1) sehr gut ... (6) gar nicht	1 2 3 4 5 6
Sie haben unter <b>b „6=NIE“</b> angekreuzt. Begründen Sie bitte, warum Sie die Kompetenz <b>NIE</b> anwenden konnten:			

Abbildung 8. Beispielitem KOSTA (Weresch-Deperrois, Bodensohn & Jäger, 2009)

Als Basis der Entwicklung von Items für den Situational Judgement Test wurden aus KOSTA (ebd.) diejenigen Items gewählt, welche innerhalb der elf Kompetenzbereiche auf der Dimension „Bedeutung“ die höchsten Trennschärfen aufwiesen. Diese Dimension wurde gewählt, da sie auf die eingeschätzte *Wichtigkeit* einzelner Kompetenzen rekurriert und somit die Relevanz einzelner Kompetenzen innerhalb eines Kompetenzbereiches in den Blickpunkt stellt.

Anhang D zeigt die gewählten Items innerhalb der verschiedenen Kompetenzbereiche bzw. deren Trennschärfen.

### 3.2.2 2. Schritt: Empirische Basis der Items mit negativer Polung

Die Güte eines Situational Judgement Tests hängt maßgeblich vom Realitätsgehalt der zu bewertenden Situationen ab. Dabei ist es insbesondere relevant, Situationsvignetten zu entwickeln, welche ein breites Spektrum an Kompetenzausprägungen reflektieren. Somit ist es nicht alleine angezeigt, bei der Itementwicklung ExpertInnen einzubeziehen, sondern ebenso, Situationen gemeinsam mit NovizInnen zu entwickeln (Bruder, 2011, Bledow & Frese, 2009; McDaniel & Nguyen, 2001; Motowidlo, Hanson, & Crafts, 1997). Dies galt in besonderem Maße für das im Rahmen dieser Arbeit entwickelte Instrument, da für jede Kompetenz im Sinne der Standards für die Lehrerbildung: Bildungswissenschaften sowohl eine Situation vom oberen Ende (positive Polung) als auch vom unteren Ende (negative Polung) des entsprechenden Kompetenzspektrums gestaltet werden sollte.

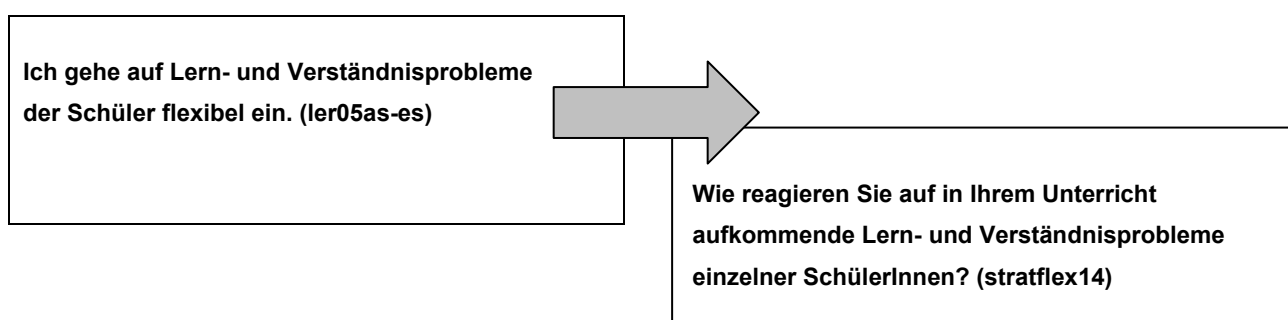
Lehramtsstudierende können als NovizInnen bezüglich professioneller pädagogischer Kompetenz gelten: Die Erlernbarkeit professionelle pädagogische Kompetenz im Rahmen des Lehramtsstudiums wird angenommen (vgl. Kunter, 2010), wobei der Ausbau der Kompetenz im Laufe der Lehrerbildung stattfindet und erst zum Ende des Studiums eine Kompetenzausprägung der Studierenden im Sinne der Berufsreife vorliegen sollte (Jurt, 2007). *Während* des Studiums sind Studierende somit NovizInnen, am *Ende* des Studiums sollten sie ExpertInnen sein – bezüglich der Standards für die Lehrerbildung: Bildungswissenschaften.

Ziel des zweiten Schritts war es deshalb, aufgrund von qualitativen Daten der professionellen pädagogischen Kompetenz von Lehramtsstudierenden zu realistischen Handlungsalternativen am unteren Ende des Kompetenzspektrums

(NovizInnen) zu gelangen, welche als Basis für die negativ gepolten Situationsvignetten des Situational Judgement Tests dienen sollten.

Um eine Erfassung der professionellen pädagogischen Kompetenzen Lehramtsstudierender jenseits einer Verhaltensbeobachtung realisieren zu können, wurde jedes aus dem Self-Assessment-Instrument KOSTA (ebd.) gewählte Item in eine offene Frage umgewandelt.

Abbildung 9 stellt dieses Vorgehen an einem Beispiel dar:



*Abbildung 9.* Schritt 2 der Itemgenerierung: Umwandlung gewählter Items in ein qualitatives Frageformat

Im Rahmen der Befragung sollten Studierende des Lehramts in einem offenen Antwortformat beantworten, wie sie Situationen des Lehralltags gestalten/gestalten würden. Das Ziel bestand darin, unter Annahme einer Novizenschaft der Studierenden zu realitätsnahen Situationsvignetten am unteren Ende des Kompetenzspektrums zu gelangen (negativ gepolte Items).

### Methode

55 offene Fragen zur konkreten Umsetzung professioneller pädagogischer Kompetenzen auf der Folie der KMK-Standards (2004) wurden von Lehramtsstudierenden jeweils stichwortartig oder in kurzen Sätzen beantwortet und von zwei unabhängigen RaterInnen nach Durchführung eines Codier-Trainings bewertet. Hierbei wurden die Antworten hinsichtlich ihrer „methodischen Grundentscheidung“ (Wuttke, 2009, S. 670) bzw. ihrer Konformität gemäß den KMK-Standards (2004) geratet: Eine „0“ wurde bei einer lehrerzentrierten/standardfernen Unterrichtsgestaltung vergeben, eine „2“ bei einer schülerzentrierten/konstruktivistisch geprägten/standardgemäßen Handlungsweise. Mit einer „1“ wurden Mischformen geratet. Nach Wuttke (2009) kann die Wahl eines

schülerzentrierten Settings als eine Kernkompetenz von Lehrenden gelten, die aus verschiedenen Facetten wie etwa Fachwissen, didaktischem Wissen, aber auch Einstellungen und Überzeugungen besteht. Für eine Implementierung dreier qualitativ verschiedener „Kompetenzgruppen“ lässt sich aufgrund Terharts (2007, S. 52) Definition von „qualitativ gestuften“ professionellen pädagogischen Kompetenzen im Lehrberuf plädieren.

Wurde bei einem Item nach mehreren Alternativen (z. B. „In welchen Situationen handeln Sie als Lehrkraft Ihrer Vorbildfunktion entsprechend?“) gefragt, so wurde die Angabe mehrerer Alternativen mit einer „2“, die Angabe einer Alternative mit „1“ sowie die Angabe keiner Alternative mit „0“ bewertet. Zusätzlich floss in dieses Urteil ein qualitatives Rating gemäß den oben genannten Kriterien ein. Abbildung 10 zeigt ein Beispielitem sowie eine erfolgte Antwort.

2
Nach welchen Überlegungen wählen Sie die Unterrichtsmethoden für Ihre Klassen aus?
Bsp.: Ich wähle die Methoden, mit denen ich die meisten Erfahrungen sammeln konnte.
<i>Methoden mit denen ich mich auskenne und die zu der Klasse passen.</i>

Abbildung 10. Beispielitem und –antwort aus dem Fragebogen „Qualitative Erfassung professioneller pädagogischer Kompetenzen Lehramtsstudierender“ der Vorstudie

### Durchführung der Studie

Zur Gewinnung der Stichprobe wurde über studentische Internet-Foren bzw. über den Allgemeinen Studierendenausschuß (AStA) verschiedener Pädagogischer Hochschulen und Universitäten Kontakt zu Studierenden aufgenommen. Außerdem wurden Dozenten in Lehramtsstudiengängen der Universität Koblenz-Landau per e-mail kontaktiert.

Zwei DozentInnen der Landauer Universität erklärten sich schließlich zur Teilnahme bereit und ermöglichten so eine Erhebung in zwei Veranstaltungen für das Lehramt. Die Bearbeitung des Fragebogens „Qualitative Erfassung professioneller pädagogischer Kompetenzen im Lehrberuf“ erfolgte freiwillig innerhalb der Veranstaltungszeit. Online erreichten Studierenden wurde der Fragebogen per e-mail zugesandt. Die Rückmeldung der Ergebnisse an die TeilnehmerInnen wurde



hierbei angeboten: Am Ende des Fragebogens konnte durch die TeilnehmerInnen eine persönliche e-mail-Adresse angegeben werden, falls eine Rückmeldung erwünscht war.

Die offenen Antworten der Lehramtsstudierenden wurden von zwei unabhängigen RaterInnen aus dem Fachbereich der Pädagogischen Psychologie beurteilt und in Form eines numerischen Codes mit der Statistiksoftware „PASW - Predictive Analytics Software - Statistics 18 von SPSS“ gespeichert sowie analysiert.

### Stichprobe

An der qualitativen Vorstudie nahmen 107 Studierende des Lehramts teil. 83,2% (89 Personen) der Studierenden waren weiblich, 16,8% (18 Personen) waren männlich.

Die Mehrzahl der TeilnehmerInnen studierten in Rheinland-Pfalz (95,3%, 102 Personen), drei Studierende (2,8%) in Baden-Württemberg und eine Person (0,9%) in Nordrhein-Westfalen.

Fast die Hälfte der Lehramtsstudierenden befand sich zum Zeitpunkt der Befragung im vierten Semester (48,6%, 52 Personen), 23,4% (25 Personen) waren im fünften, 11,2% (12 Personen) im sechsten Semester sowie 7,5% (8 Personen) im dritten und 5,6% (6 Personen) im zweiten Semester. Jeweils eine Person (0,9% der Stichprobe) war im siebten und achten Semester. Somit ergibt sich im Mittel eine Semesterzahl von 4,28.

Hinsichtlich des Berufsziels gab die Mehrheit der Studierenden die Grund- und Hauptschule an (91,6%, 98 Personen), sechs Personen (5,6%) möchten am Gymnasium unterrichten, 2 Personen an der Förderschule (1,9%). Eine Person (0,9%) machte keine Angaben zum Berufsziel.

49 Personen (45,8%) hatten zum Zeitpunkt der Befragung das dritte Orientierende Praktikum absolviert. 29,0 % (31 Personen) hatten das erste Vertiefende, 13,1% (14 Personen) das zweite Orientierende Praktikum hinter sich. 6,5% (7 Personen) bzw. 2,8% (3 Personen) hatten bereits das erste Orientierende oder das zweite Vertiefende Praktikum absolviert. Bislang noch kein Praktikum hatten zwei Personen (1,9%) durchlaufen, eine Person (0,9%) hatte noch im Rahmen der „alten“ Lehramtsstudiengänge an Praktika teilgenommen.

## Ergebnisse

### a) Interrater-Reliabilität

Bei einer objektiven Auswertung ist die Vergabe von Testpunkten für bestimmte Testantworten vom individuellen Testauswerter unabhängig (Bortz & Döring, 2003). Dies ist dann der Fall, wenn eine eindeutige Vorschreibung der Itembeantwortung sowie der Bewertung der Antworten (beispielsweise mithilfe eines Scoring-Schlüssels, also der Anweisung, für bestimmte Antworten eine bestimmte Anzahl an Punkten zu vergeben) vorliegt. Ein solcher Schlüssel wurde für das Instrument „Qualitative Erfassung professioneller pädagogischer Kompetenzen im Lehrberuf“ generiert. Die Beantwortungen der offenen Fragen wurden folgendermaßen quantifiziert:

0 = nicht standardkonform,

1 = teilweise standardkonform und

2 = standardkonform

Trotzdem ist die Auswertungsobjektivität bei qualitativen Studien zusätzlich zu überprüfen, da die Zuordnung qualitativer Antworten zu einer quantitativen Größe trotz des Vorliegens eines Scoring-Schlüssels durch eine subjektive Wahrnehmung beeinflusst wird.

Zur Bestimmung der Beurteilerreliabilität wurde auf Maße der Zuverlässigkeit für ordinalskalierte Ratings zurückgegriffen, da das Vorliegen einer Intervallskala bei drei Ratingstufen (2 = völlig standardkonform, 1 = teilweise standardkonform, 0 = nicht standardkonform) kaum als gegeben angesehen werden kann.

Zunächst wurde der Wilcoxon-Test für abhängige Stichproben gerechnet, um die Homogenität der Randverteilungen zu überprüfen. Nach Wirtz & Caspar (2002) wird dadurch analysiert, „ob die Zuverlässigkeit der Beurteilung unter Umständen dadurch verringert wird, dass die RaterInnen die möglichen Kategorien grundsätzlich unterschiedlich häufig wählen.“ (S. 129) Hierbei prüft der Wilcoxon-Test, ob die Beurteilungen beider RaterInnen im Durchschnitt systematisch voneinander abweichen.

Anschließend wurde Spearmans  $\rho$  als gebräuchlichstes Maß der Korrelation ordinalskalierter Urteile (Wirtz & Caspar, 2002) berechnet. Hierbei werden die Rangreihen der untersuchten Objekte für beide RaterInnen korreliert.

Bezüglich der Homogenität der Randverteilungen ist die Null-Hypothese „Die Randverteilungen beider RaterInnen unterscheiden sich nicht.“ die Wunschhypothese. Bei 53 Items wird die *Null-Hypothese aufgrund nicht-signifikanter Ergebnisse nicht abgelehnt*. Lediglich bei zwei Items („Förderung eigenverantwortlichen Urteilens“ und „Vertrauensvolles Klassenklima“) ergeben sich heterogene Randverteilungen. Somit weichen die Beurteilungen der RaterInnen mit Ausnahme von zwei Items nicht systematisch voneinander ab. Die vollständigen Ergebnisse lassen sich Anhang F entnehmen.

Die Korrelationen zwischen den Rangreihen beider RaterInnen zeigen sich durchweg hochsignifikant sowie hoch positiv. Es ergibt sich eine hohe mittlere Korrelation von .803 (Anhang F stellt die Ergebnisse vollständig dar). Somit zeigt sich auch hier eine hohe Homogenität der Urteile.

Zusätzlich wurden die Kreuztabellen der 55 von beiden RaterInnen beurteilten Items analysiert, um zu überprüfen, ob sich Diskrepanzen zwischen den Urteilen dergestalt ergaben, dass eine Antwort von Rater 1 als „völlig standardkonform“ und von Rater 2 als „nicht standardkonform“ beurteilt wurde. Bei den Variablen „Umgang mit Verständnisproblemen“, „Berücksichtigung von Prinzipien des Wissenserwerbs“, „Bereitstellung von Lernhilfen“, „Planung von Projekten“, „Informationsbasis bei der Planung von Projekten“, „Berücksichtigung von Subjektivität“, „transparente Konfliktlösung“, „nachvollziehbare Beurteilungskriterien“, „Reflektion eigener Erfahrungen“, „Evaluation von Unterricht“, „Fehlernutzung“, „Perspektivenwechsel“, „Weiterbildung“, „Einführung in neues Thema“ sowie „soziale Lernprozesse“ zeigten sich Diskrepanzen bei einzelnen Datensätzen, wobei insgesamt von 11.770 Urteilen (107 ProbandInnen x 55 Items x 2 RaterInnen) in 30 Fällen diskrepante Urteile resultierten. Dies entspricht einem prozentualen Anteil von 0.25%.

Aufgrund der Ergebnisse des Tests auf Homogenität der Randverteilungen sowie der Rangkorrelationen kann somit von einer hohen Interrater-Reliabilität und somit Auswerterobjektivität ausgegangen werden.

#### b) Deskriptive Untersuchungsergebnisse

Mit 86,0% (92 Personen), 90,7% (97 Personen) und 86,9% (93 Personen) wurden bei den Items „Schülerförderung“, „Nutzung von Informationsquellen zur Optimierung des eigenen Unterrichts“ sowie „Förderung von SchülerInnen mit Migrationshintergrund“ am häufigsten „standardkonforme“ Kompetenzen genannt.

Anhang F stellt die prozentualen Anteile „standardkonformer“, „teilweise standardkonformer“ und „nicht standardkonformer“ Antworten vollständig pro Item dar (dort wird zugunsten der Übersichtlichkeit auf die Darstellung der Anteile für Zwischenzuordnungen – 0,5 oder 1,5 – verzichtet).

Bezüglich nicht standardkonformer Nennungen zeigt sich folgendes Bild: Das Item „Evaluation von Unterricht“ zeigt mit 18,7% (20 Personen) den höchsten Anteil an „nicht standardkonformen“ Antworten. Die Items „Aktualisierung eigenen Wissens“, „Planung von Projekten“, „Berücksichtigung von Subjektivität“, „Mitwirkung an Schulentwicklung“, „Nutzung von Diagnostik zur Unterrichtsgestaltung“, „Evaluation von Unterricht“, „Nutzung von Ergebnissen der Bildungsforschung“, „rechtliche Rahmenbedingungen“, „Moralerziehung“ sowie „Zustandekommen von Arbeitsergebnissen“ weisen die weiteren höchsten Anteile an nicht standardkonformen Nennungen zwischen 9,3% („Nutzung von Diagnostik zur Unterrichtsgestaltung“, „Moralerziehung“, 10 Personen) und 16,8% („Planung von Projekten“, 20 Personen) auf.

Im Mittel wurden die Items zu 49,08% mit standardkonformen bzw. zu 5,18% mit nicht standardkonformen Nennungen beantwortet.

Somit kann für diese Stichprobe ohne Anspruch auf Generalisierbarkeit gelten, dass die genannten Kompetenzen den Standards der KMK (2004) vornehmlich entsprechen.

Die qualitative Vorstudie sollte als realitätsnahe Quelle für die negativ gepolten Situationsvignetten des Situational Judgement Tests dienen. Die Nennungen verschiedener Kompetenzausprägungen durch die Studierenden konnten hierfür trotz deutlicher Deckeneffekte genutzt werden, da bei allen Items mit Ausnahme von Item „Vorbildfunktion“ auch „nicht standardkonforme“ Kompetenzen genannt wurden, die als Vorlage für negativ gepolte Items fungieren konnten.

### 3.2.3 3. Schritt: Entwicklung der Items

#### Allgemeine Konstruktionskriterien von Situational Judgement Tests

Situational Judgement Tests stellen ein simulationsorientiertes Verfahren zur situationsspezifischen Erfassung multipler Konstrukte dar (Bledow & Frese, 2009). Doch wie muss ein solcher Test konstruiert werden, um zu einer objektiven,

reliablen und validen Kompetenzeinschätzung zu gelangen? Die folgenden Abschnitte geben einen Überblick über die Konstruktionskriterien eines Situational Judgement Tests, wobei bedacht werden muss, dass es sich hier um pragmatische, bislang empirisch noch nicht ausreichend fundierte Empfehlungen handelt. Zuletzt wird auf die konkrete Konstruktion der Items des Instrumentes „Simulationsorientierte Erfassung professioneller pädagogischer Kompetenz Lehramtsstudierender“ eingegangen.

#### a) Entwicklung des Item-Stammes

Die Items von Situational Judgement Tests werden durch die Situationsvignetten realisiert, die den ProbandInnen beispielsweise schriftlich oder videobasiert zur Beurteilung vorgegeben werden. Bei der Auswahl adäquater Situationen – dargestellt als Item-Stamm – müssen Entscheidungen hinsichtlich a) der Quelle, b) der Komplexität, c) der Fidelität und d) des Inhaltes des Item-Stammes getroffen werden.

- Die Quelle des Item-Stammes

Bei der Generierung von Items für Situational Judgement Tests lassen sich nach Weekley, Ployhart und Holtz (2006) zwei grundsätzliche Vorgehensweisen unterscheiden, die im Allgemeinen der *induktiven und deduktiven Methode* der Testentwicklung entsprechen.

Die *Critical-Incident-Technique* („Technik der kritischen Ereignisse“, Flanagan, 1954) zählt zu den *induktiven Ansätzen* der Item-Generierung und wird bei Weitem am häufigsten zur Entwicklung von Situational Judgement Tests eingesetzt. ExpertInnen des interessierenden Professionsbereiches werden hierbei gebeten, sich an sehr erfolgreiches und sehr wenig erfolgreiches Handeln zu erinnern, wobei meist ein dreiteiliges Format aus Antezedenz, Verhalten und Konsequenz verwendet wird. Die Antezedenzen liefern somit den Itemstamm (die Situationen), während die resultierenden Verhaltensweisen als Antwortalternativen genutzt werden können.

Kritische Situationen werden gemeinhin von Vorgesetzten oder deren Angestellten erfragt, wobei die Befragungen innerhalb von Aufgabenbereichen (z. B. Kundenbetreuung, Verwaltung, Entwicklung) gruppiert werden sollten, da sich je nach Tätigkeit unterschiedliche kritische Situationen ergeben (Schuler & Höft, 2004): So wird ein am Kundenschalte einer Bank arbeitender Angestellter eher auf

Situationen im Zuge der Kundenbetreuung focussieren, wohingegen ein Angestellter der Verwaltung eher administrative Aspekte herausstellen wird.

Eine weitere Quelle für kritische Ereignisse stellen schriftliche Aufzeichnungen – beispielsweise von Kundenbeschwerden oder auch Flugzeugunfällen (Hunter, 2003) – dar. Die Unabhängigkeit solcher Aufzeichnungen vom Befragungskontext ist insbesondere dann hoch, wenn sie nicht aufgrund einer Critical-Incident-Befragung zustande kommen, sondern aus obligatorischen Berichten (beispielsweise von Unfällen oder Polizeieinsätzen) bestehen.

Bislang liegen allerdings noch keine Forschungsergebnisse vor, die Aufschluss darüber liefern könnten, in welchem Ausmaße die Quellen der kritischen Ereignisse die Entwicklung von Situational Judgement Tests beeinflussen und welche Vertreter eines Berufsstandes zu welchen Kontexten befragt werden sollten.

Im Zuge des *deduktiven Ansatzes* werden die Items auf Basis eines theoretischen Modells, einer Arbeits- und Aufgabenanalyse, aufgrund von Reviews der einschlägigen Literatur oder eines Sets von empirisch bestätigten Kompetenzen generiert.

Zwei Beispiele für ein modellgeleitetes Vorgehen liefern die Studien von Stevens und Campion (1999) sowie Becker (2004): Stevens und Campion (ebd.) identifizierten beispielsweise fünf Dimensionen der Teamfähigkeit und entwickelten auf Basis ihres Modells Items ohne Ableitung aus der Critical-Incident-Technik. Becker (ebd.) realisierte einen Situational Judgement Test entlang seines Modells der Integrität mit den sieben Facetten „Gutmütigkeit, Ehrlichkeit, Unabhängigkeit, Gerechtigkeit, Produktivität, Stolz und Rationalität“.

Obwohl zahlreiche Situational Judgement Tests auf der Grundlage mehrdimensionaler theoretischer Modelle und Konstrukte entwickelt wurden, weisen nach Weekley, Ployhart und Holtz (2006) nur wenige entsprechende Subskalen auf (z. B. Bruder, 2011). Der Großteil der Testautoren implementiert lediglich einen Gesamtscore.

Ob dem induktiven oder deduktiven Ansatz der Vorzug zu geben ist, lässt sich bislang nicht eindeutig beantworten. Induktive Critical-Incident-Methoden erlauben zwar den Zugang zu reichhaltiger, realitätsnaher Information, sind jedoch in der

Entwicklung aufwendig und teuer. Zudem lassen sie den Bezug zu einer theoretischen Basis häufig vermissen. Ein rein deduktives Vorgehen könnte andererseits einen zu geringen praktischen Bezug aufweisen, und so plädieren Weekley, Ployhart und Holtz (2006) für die Kombination beider Ansätze: Zunächst solle aufgrund theoretischer Überlegungen oder Arbeitsanalysen ein Modell entwickelt werden, anschließend sollten Critical Incidents gesammelt werden, um das Modell zu bestätigen.

In der vorliegenden Arbeit wird eine solche Kombination des induktiven und deduktiven Vorgehens realisiert: Die Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004) fungieren als „theoretische Basis“ der Items und Skalen, wohingegen die situative Ausgestaltung der Items aufgrund der qualitativen Befragung Lehramtsstudierender bzw. den Einbezug von ExpertInnen erreicht wird.

- Die Komplexität des Item-Stammes

Verschiedene Situational Judgement Tests unterscheiden sich stark im Hinblick auf die Komplexität, Detailgenauigkeit und Spezifität ihrer Itemstämme (McDaniel & Nguyen, 2001).

Lievens, Buyse und Sackett (2005) postulieren einen Zusammenhang zwischen dem Entsprechungsgrad „Test-Items/Kriterienbereich“ und der Validität des Tests. McDaniel et al. (2001) nahmen an, dass eine größere Anzahl spezifischer Items zu höherer Validität führe, sahen diese Erwartung durch ihre Meta-Analyse jedoch nicht bestätigt: Detailliertere Situational Judgement Tests wiesen eine geringere Kriteriumsvalidität auf als relativ globale Situational Judgement Tests. Reynolds, Sydell, Scott und Winter (2000) sowie Reynolds, Winter und Scott (1999) gelangten allerdings zu einem konträren Schluss, indem sie für mehr und spezifischere Items sowohl eine bessere Validität als auch eine größere Kulturunabhängigkeit aufzeigten.

Sacco, Scheu, Ryan & Schmitt (2000) beleuchteten einen anderen Aspekt der Komplexität: In ihrer Studie war die Lesekompetenz signifikant korreliert mit Unterschieden der Leistung in einem Situational Judgement Test zwischen verschiedenen Subgruppen (gruppiert hinsichtlich ihres ethnischen Hintergrunds), wobei Effekte der Lesekompetenz alleine auf der Ebene der Situationsbeschreibungen – nicht auf der Ebene der Antwortoptionen – beobachtet

wurden. In einer Folgestudie von Sacco, Scheu, Ryan, Schmitt, Schmidt und Rogg (2000) wurden einfachere und kürzere Situationsbeschreibungen verwandt, wodurch die Unterschiede zwischen den Subgruppen sich stark verkleinerten und sich zudem nicht mehr auf Indizes der Lesekompetenz beziehen ließen.

Zwar scheint folglich die Lesekompetenz einen Teil des kognitiven „Loads“ von Situational Judgement Tests zu bedingen, inwieweit die Lesekompetenz die Kriteriumsvalidität beeinflusst, ist allerdings bislang ungeklärt. Zukünftige Forschung ist hier ebenso notwendig wie hinsichtlich der Beziehungen zwischen Item-Spezifität und -Komplexität, Lesekompetenz, Differenzen zwischen Subgruppen, kognitiver Fähigkeit und Validität.

- Die Fidelität des Item-Stammes

Der Begriff der Fidelität bezieht sich auf den Grad der realistischen Darstellung von Sachverhalten in Tests (vgl. Motowidlo, Hanson & Crafts, 1997). Im Kontext von Situational Judgement Tests beschreibt die Fidelität den Realitätsgehalt der vorgegebenen Situationen (Items).

Es lassen sich zwei Arten der Fidelität unterscheiden (Weekley, Ployhart & Holtz, 2006): Die *psychologische Fidelität* bezeichnet hierbei die Genauigkeit, mit der die Items eines Tests diejenigen psychologischen Konstrukte spiegeln, welche im Kriterienbereich benötigt werden (z. B.: „Erfordert der Test die gleichen kognitiven Operationen, die auch bei der Ausübung des Berufes benötigt werden?“). Items weisen *physikalische Fidelität* auf, wenn die vom Test erforderten Verhaltensweisen dem Verhalten entsprechen, welches in der tatsächlichen Anforderungssituation gezeigt wird. Durch die Art der Präsentation eines Tests (paper-pencil-Verfahren, Videoformat, PC-gestützte Erhebung usw.) können sowohl physikalische als auch psychologische Fidelität beeinflusst werden. In einer Studie von Chan und Schmitt (1997) ließen sich etwa Gruppendifferenzen zwischen TestteilnehmerInnen unterschiedlicher Ethnie durch einen Wechsel von einem paper-pencil- zu einem Videoformat reduzieren.

- Inhalt des Item-Stammes

Viele Situational Judgement Tests geben insbesondere interpersonelle Situationen vor (Weekley, Ployhart & Holtz, 2006). Situational Judgement Tests können aber ebenso zur Erhebung prozeduralen Wissens genutzt werden, wenn etwa Angestellten einer Firma zur IT-Beratung innerhalb eines Items Fehlermeldungen



eines PCs vorgegeben werden, zu welchen eine adäquate Reparaturprozedur gewählt werden muss. Die Konstruktvalidität eines Tests ist dabei direkt abhängig vom Inhalt seiner Items.

#### b) Die Generierung von Antwortoptionen

Die Zahl der vorgegebenen Antwortoptionen der meisten Situational Judgement Tests bewegt sich Weekley, Ployhart und Holtz (2006) zufolge zwischen drei und zwanzig. Bei der Entwicklung adäquater Antwortmöglichkeiten müssen Entscheidungen hinsichtlich a) der Quelle, b) gegebenenfalls des zugrundeliegenden Konstrukts, c) der Komplexität, d) der Fidelität und e) der Transparenz getroffen werden.

- Die Quelle der Antwortoptionen

Bislang liegen keine Hinweise darauf vor, ob mögliche Antworten idealerweise durch die Testautoren oder durch Praktiker entwickelt werden sollten (ebd.). Der Vorteil einer praxisorientierten Generierung könnte hierbei in der möglicherweise höheren ökologischen Validität liegen, wobei die Antwortentwicklung entlang theoretischer Erwägungen konstruktreiner gelingen könnte.

- Konstruktbasierte Antwortoptionen

Bei der konstruktbasierten Antwortgenerierung sollen durch die Antwortoptionen bestimmte Konstrukte operationalisiert werden. Das Vorgehen orientiert sich somit an theoretischen Annahmen über einen interessierenden Inhaltsbereich.

Im Gegensatz zur Strategie der „intuitiven Konstruktion“ (Jonkisz & Moosbrugger, 2007), welcher gefolgt wird, wenn der Kenntnisstandbezüglich eines Merkmals noch keine theoriebasierte Entwicklung erlaubt, stellt bei der konstruktbasierten, rationalen Konstruktion eine Theorie den Rahmen für die Entwicklung von Items und Antwortoptionen. Hierbei werden von einem definierten Konstrukt (z. B. „Extraversion“) Verhaltensindikatoren abgeleitet (z. B. „Ich gehe gerne auf Parties.“).

- Komplexität

Bislang liegt nach Weekley et al. (2006) keine Forschung zur Komplexität von Antwortoptionen vor, es lassen sich also zum jetzigen Zeitpunkt keine Aussagen darüber treffen, ob detaillierten und spezifischen oder globaleren Antwortoptionen der Vorzug zu geben ist.

Es scheint den Autoren zufolge allerdings vertretbar anzunehmen, dass detaillierte und spezifische Antwortoptionen eine größere Validität hinsichtlich spezifischer Anforderungen (z. B. Kundenorientierung im Beschwerdemanagement) aufweisen, während sich globalere Antwortoptionen (z. B. allgemeine Kundenorientierung) besser auf eine breite Anforderungsklasse generalisieren lassen müssten.

- Fidelität

Antwortoptionen können – analog zum Itemstamm – mehr oder minder realistisch sein und verschiedene Grade der psychologischen und physikalischen Fidelität aufweisen. Dabei kann das Format der Antwortoptionen sich durchaus von demjenigen des Itemstammes unterscheiden: Beispielsweise können Situationen im Videoformat vorgegeben werden, während die Wahl von Antwortoptionen als paper-pencil-Fragebogen realisiert wird. Bezüglich der Beziehung zwischen dem Format der Antwortoptionen und der Konstruktvalidität ergeben sich die gleichen Implikationen wie beim Itemstamm: paper-pencil-Formate scheinen höher mit allgemeiner kognitiver Fähigkeit zu korrelieren (McDaniel et al., 2006).

- Transparenz

Ein möglicher Nachteil von Situational Judgement Tests liegt in der potentiellen Durchschaubarkeit – und somit Verfälschbarkeit – der Antwortoptionen. Bislang liegen hierzu zwar nur wenige Forschungsergebnisse vor, eine Studie von Olson-Buchanan, Drasgow, Moberg, Mead, Keenan und Donovan (1998) liefert allerdings Hinweise darauf, dass Antworten dann weniger fälschungsanfällig sind, wenn lediglich sozial erwünschte Antwortoptionen vorgegeben werden: Die TestteilnehmerInnen wählten aus einem größeren Antwortpool, wenn die Antwortoptionen annähernd sozial erwünscht waren.

Im Allgemeinen lassen sich innerhalb von Situational Judgement Tests drei Arten der Antwort-Instruktion unterscheiden:

1. Die Frage „Was würden Sie tun?“ („would do“) erfragt *Verhaltenstendenzen bzw. Verhaltensabsichten* des/der Testteilnehmers/in bei gegebenem Situations-Item.
2. „Was sollte in dieser Situation getan werden?“ („should do“) bezieht sich hingegen auf das *Wissen* darüber, welche Handlungsoption bei gegebenem Situations-Item die beste/schlechteste sei.

3. Auch Likert-skalierte Antwortinstruktionen sind gebräuchlich. Hierbei wird von den ProbandInnen eine Einschätzung der Effektivität der Antwortoptionen gefordert („Wie effektiv ist das geschilderte Verhalten Ihrer Meinung nach?“).

Ployhart und Ehrhart (2003) kommen nach der Reliabilitäts- und Validitätsanalyse von sechs verschieden gestalteten Antwort-Instruktionen zu dem Schluss, dass unter „would do“-Bedingungen die kriteriumsbezogenen Validitäten günstiger ausfallen als unter einer „should do“-Bedingung. Die Autoren führen als mögliche Erklärung für die höhere Validität verhaltensbezogener Antwortinstruktionen die Verhaltenskonsistenz an. Lievens, Sackett und Buyse (2009, S.1096), schreiben diesbezüglich: „Intended behavior predicts future behavior“. Zudem zeigten sich relativ geringe Korrelationen zwischen „would do“- und „should do“-Bedingungen, obwohl die gleichen Items verwendet wurden.

Antworten lassen sich nach Nguyen, Biderman und McDaniel (2005) unter „should do“-Bedingungen weniger leicht verfälschen, als wenn nach Verhaltenstendenzen gefragt wird. In einer Metaanalyse von 41 Situational Judgement Tests mit wissensbezogenen und 21 Situational Judgement Tests mit verhaltensbezogenen Antwortinstruktionen fanden McDaniel, Hartman, Whetzel und Grubb (2007) jedoch höhere Validitäten für „should-do“-Bedingungen und höhere Korrelationen dieser Instruktionsform mit kognitiver Fähigkeit, wohingegen verhaltensbezogene Instruktionen eher mit den Persönlichkeitseigenschaften „Gewissenhaftigkeit, Verträglichkeit und emotionale Stabilität“ korrelierten.

Likert-skalierte Antworten können beispielsweise durch den Vergleich der Ratings von ProbandInnen mit den Ratings von ExpertInnen verglichen und bezüglich ihrer Übereinstimmungen beurteilt werden. Ein entsprechend gestalteter Situational Judgement Test von Chan und Schmitt (2002) zeigte sich inkrementell valide gegenüber der kognitiven Fähigkeit, Erfahrung und der Big Five-Persönlichkeitsfaktoren bezüglich einer Prädiktion der arbeitsbezogenen Performanz. Wagner und Sternberg (1985) konnten aufgrund eines ebenfalls Likert-skalierten Situational Judgement Tests und eines Summations-Antwortschlüssels signifikant zwischen ExpertInnen und NovizInnen diskriminieren. Anhand der Testergebnisse von ProbandInnen nach der Bearbeitung eines eines Situational Judgement Tests mit Likert-Skalierung war es folglich möglich, Aussagen bezüglich der Berufserfahrung der ProbandInnen zu machen.

### c) Antwortschlüssel

Die Generierung der Antwortschlüssel kann sowohl rational aufgrund theoretischer Überlegungen als auch aufgrund empirischer Erkenntnisse oder auf Basis von ExpertInnenbefragungen erfolgen. Nach Weekley und Jones (1999) sind diese Methoden hinsichtlich der Validität gleichwertig. Auch Paullin und Hanson (2001) fanden hinsichtlich der Validität keine Unterschiede zwischen empirisch und rational entwickelten Antwortschlüsseln.

Bergman, Drasgow, Donovan und Juraska (2003) verglichen aufgrund rationalen und theoretischen oder empirisch geleiteten Vorgehens entwickelte Antwortschlüssel: Sowohl auf Basis von ExpertInneneinschätzungen entwickelte als auch auf empirischen Erkenntnissen basierende Antwortschlüssel zeigten hierbei einen Zusammenhang mit der Performanz. Theoriebasierte Antwortschlüssel zeigten dagegen keine signifikanten Kriteriumsvaliditäten.

Zwar erweisen sich in den meisten betreffenden Studien empirisch und rational entwickelte Antwortschlüssel als ähnlich valide, allerdings korrelieren Testergebnisse in Situational Judgement Tests mit empirischen vs. rationalen Antwortschlüsseln meist nur gering (Weekley, Ployhart & Holtz, 2006). Gründe hierfür könnten darin bestehen, dass (1.) unterschiedliche Konstrukte erfasst werden, dass (2.) verschiedene Facetten eines Kriterienbereichs erfasst werden oder dass (3.) die unterschiedlichen Antwortschlüssel verschiedene – aber jeweils richtige – Antworten provozieren. Hier ist weitere Forschungsarbeit zu leisten.

Obwohl die meisten Antwort- und Bewertungsschlüssel von Situational Judgement Tests auf der Basis von ExpertInnenaussagen entwickelt werden, besteht nach Weekley, Ployhart und Holtz (ebd.) weiterer Forschungsbedarf hinsichtlich dessen, inwiefern verschiedene Subgruppen von ExpertInnen oder anderen betroffenen Personengruppen (wie etwa KundInnen) zu unterschiedlichen Antwortschlüsseln gelangen. So erscheint es beispielsweise plausibel anzunehmen, dass SchülerInnen die pädagogische Kompetenz von Lehrpersonen hinsichtlich anderer Kriterien bewerten als LehrerInnen selbst – beispielsweise könnten LehrerInnen das eigene fachwissenschaftliche Wissen als wichtiger einschätzen als dies SchülerInnen tun. Diese Heterogenität hinsichtlich als relevant erachteter Kriterien könnte jedoch bewusst genutzt werden, um einen Kriterienbereich breit und multiperspektivisch abdecken zu können.

Am seltensten wird zur Generierung von Antwortschlüsseln auf theoretische Überlegungen zurückgegriffen (Weekley, Ployhart und Holtz, 2006), und die Forschungslage bezüglich der Validität theoriegestützter Antwortschlüssel zeigt sich zudem uneinheitlich: Bei einem theoriebasierten Antwortschlüssel von Olson-Buchanan et al. (1998) zeigte sich zwar dessen Prädiktivität bezüglich konfliktfähigen Verhaltens, allerdings war diese geringer als diejenige eines empirisch gestützten Antwortschlüssels. In der Studie von Bergman, Drasgow, Donovan und Juraska (2003) zeigte sich gar, dass keiner von zwei auf Führungstheorien basierenden Antwortschlüsseln mit der Performanz korreliert war.

#### d) Scoring

Nachdem festgelegt wurde, welche Antwortoptionen als (nicht) effektiv gelten können, muss der Antwortschlüssel mit einem Beurteilungsschlüssel verknüpft werden, um die Testleistung eines/einer Probanden/in bestimmen zu können. Zweierlei Scoring-Methoden werden hierbei eingesetzt: Im Zuge von Forced-Choice-Methoden muss der/die Proband/in (beispielsweise) eine Antwortoption wählen, bei Likert-skalierten Antwortformaten schätzt der/die Proband/in (beispielsweise) die Effektivität von Antwortoptionen ein (Weekley, Ployhart & Holtz, 2006). Beide Methoden werden nachfolgend dargestellt.

- Forced-Choice-Methoden

Die zwingende Auswahl aus vorgegebenen Antwortoptionen wird im einfachsten Falle so gestaltet, dass eine der Antwortoptionen als „richtig“ bestimmt wird, wohingegen alle anderen Optionen als „falsch“ gelten. Der/die Proband/in erhält hierbei bei Wahl der richtigen Antwort einen Punkt, wählt er eine der falschen Optionen („Distraktoren“), wird kein Punkt vergeben. Dem Sachverhalt, dass nicht alle Distraktoren gleichermaßen „falsch“ sein könnten, kann dadurch Rechnung getragen werden, dass unterschiedliche Anzahlen von Punkten vergeben werden: Für die „richtige“ Antwort erhält der/die Testteilnehmer/in 2 Punkte, für die „falsche“ Antwort 0 Punkte. Wählt der/die Proband/in jedoch eine Antwort, die zwar nicht als die beste definiert wurde, jedoch besser ist als die „falsche“ Antwort, so erhält er einen Punkt.

Motowidlo, Dunnette und Carter (1990) differenzierten diesen Ansatz weiter, indem sie ProbandInnen die beste und schlechteste Antwortoption wählen ließen und wie folgt Scores pro Item errechneten: Wählte eine Person sowohl die als beste als auch die als schlechteste Option definierte Antwort, so erhielt sie +2 Punkte,

wohingegen -2 Punkte resultierten, wenn die Person die als beste definierte Option als die schlechteste wählte und umgekehrt. Die Punktzahlen +1 und -1 wurden vergeben, wenn der/die Proband/in eine der als „beste“ und „schlechteste“ Option definierten Antworten korrekt wählte (+1) oder wenn der Proband die „beste“ Option als die „schlechteste“ wählte oder umgekehrt (-1). 0 Punkte erhielt der/die Proband/in, wenn er/sie in beiden Fällen Distraktoren wählte.

Die Vorteile einer getrennten Erhebung der „besten/wahrscheinlichsten“ und der „schlechtesten/unwahrscheinlichsten“ Antwortoption liegen zum Einen in der erhöhten Varianz jedes Item-Scores, zum Anderen bestehen Hinweise darauf, dass sich die kriteriale Validität von Situational Judgement Tests hierdurch verbessern lässt (z. B. Cucina, Vasilopoulos & Leaman, 2003).

Eine weitere Variation der Forced-Choice-Methode besteht darin, die TestteilnehmerInnen die Antwortoptionen in eine Rangreihe bringen zu lassen. Die sich hierdurch ergebende inkrementelle Validität ist nach Weekley, Ployhart und Holtz (2006) im Vergleich zu „Wählen Sie die beste/schlechteste Antwort“-Formaten jedoch lediglich klein.

- Likert-Skalen

Werden Likert-Skalen in Situational Judgement Tests eingesetzt, so müssen die ProbandInnen jede Antwortoption auf einer vorgegebenen Skala (beispielsweise von 1=sehr ineffektiv bis 6= sehr effektiv) dahingehend bewerten, wie effektiv sie ihrer Meinung nach ist.

Nach McDaniel und Nguyen (2001) können sich durch den Einsatz von Likert-Skalen eine Reihe von Vorteilen und Problemen ergeben:

1. Die Antworten auf jede Option sind unabhängig: Bei Forced-Choice-Formaten können die TestteilnehmerInnen sämtliche Antwortoptionen direkt miteinander vergleichen, wodurch die Wahl einer Antwortoption nicht unabhängig von den Antwortalternativen vorgenommen wird. Beim Einsatz von Likert-Skalen stellt dagegen jede Antwortalternative ein eigenes Item dar, wodurch die Einschätzung hinsichtlich der Effektivität beispielsweise des geschilderten Verhaltens unabhängiger erfolgt, sofern allerdings die Items innerhalb des Tests nicht direkt hintereinander stehen.

2. Da jede Antwortoption einzeln bewertet wird (siehe 1.), ergibt sich eine größere Anzahl an Items im Vergleich zu einem entsprechenden „Wählen Sie die beste/schlechteste Antwort“-Format, wodurch sich eine erhöhte Reliabilität des Testverfahrens ergeben kann.
3. Problematisch ist es hingegen, dass sich Personen hinsichtlich der Ausschöpfung der vorgegebenen Skala unterscheiden – hierdurch können sich Verzerrungen aufgrund von Antwortmustern ergeben (z. B. die „Tendenz zur Mitte“ – hier vermeiden es ProbandInnen, die Extreme der Skala zu wählen, wodurch sich eine Varianzeinschränkung ergibt, welche unabhängig vom eigentlich zu erfassenden Merkmal ist.)

### e) Zusammenfassung

In den vorangegangenen Abschnitten wurden die Aspekte dargelegt, die im Zuge der Entwicklung eines Situational Judgement Tests berücksichtigt werden müssen. Hierbei wurde deutlich, dass bislang nur wenig systematische Forschung bezüglich der Generierung von Item-Stamm, Antwort-Optionen und –Instruktionen, des Antwortschlüssels sowie des Scorings vorliegt, wobei zusätzlich erschwerend zwischen den wenigen existierenden Studien weder die abhängigen Variablen (beispielsweise der Item-Stamm) noch die unabhängigen Variablen (beispielsweise die Validität) übereinstimmen.

Tabelle 2 gibt einen Überblick über die Bereiche, innerhalb derer Entscheidungen während der Entwicklung eines Situational Judgement Tests getroffen werden müssen, sowie über die Alternativen, die sich hierbei anbieten.

Tabelle 2. *Überblick über Entscheidungsalternativen bei der Entwicklung von Situational Judgement Tests (Weekley, Ployhart & Holtz, 2006, S. 158, eigene Übersetzung)*

	<b>Item-Stamm</b>	<b>Antwort-Optionen</b>	<b>Antwort-Instruktionen</b>	<b>Antwort-schlüssel</b>	<b>Scoring</b>
<b>Entscheidungsalternativen</b>	a) Quelle	a) Quelle	a) „Would do“	a) ExpertInnen	a) Forced-Choice
	b) Komplexität	b) Konstrukt- basiertheit	b) „Should do“	b) Empirisch	b) Kontinuierlich
	c) Fidelität	c) Komplexität	c) Effektivität	c) Theorie- basiertheit	
	d) Inhalt	d) Fidelität e) Verfälschbarkeit		d) Hybride	

Folgende Entscheidungen fielen im Rahmen der vorliegenden Entwicklung eines Situational Judgement Tests: Bezüglich des Item-Stammes wird eine Kombination *induktiven und deduktiven Vorgehens* realisiert. Die Komplexität des Itemstammes wird *niedrig* gehalten bei gleichzeitiger *hoher psychologischer und physikalischer Fidelität*. Die Itemstämme werden sowohl durch die *Testautorin* als auch durch *ExpertInnen und PraktikerInnen* generiert, wobei die Eignung der Antwortoptionen durch *ExpertInnenratings* bewertet wird. Die Antwortoptionen operationalisieren das standardorientierte „*Konstrukt*“ professioneller pädagogischer Kompetenz Lehramtsstudierender bei psychologischer und physikalischer *Fidelität*. Im Sinne einer durch die Darbietung von Antwortalternativen jeweils als Einzelitems (im Gegensatz zur parallelen Darbietung im Rahmen eines Forced-Choice-Formats) zu erreichenden lokalen stochastischen Unabhängigkeit, einer zu erwartenden größeren Varianz sowie Reliabilität und Validität (McDaniel & Nguyen, 2001) wird eine *Likert-skalierte Effektivitätseinschätzung* realisiert.



## Entwicklung der Items zur Erfassung professioneller pädagogischer Kompetenz

Auf Basis der Ergebnisse der qualitativen Vorstudie bzw. aufgrund theoretischer Überlegungen zu Einzelkompetenzen innerhalb der KMK-Standards (2004) wurden jeweils zwei gepolte Items für jedes der aus KOSTA (ebd.) ausgewählten Items generiert: Hierbei wurde ein Item für den Pol „standardkonform“ bzw. ein weiteres Item für den Pol „nicht standardkonform“ realisiert. Es resultierten also für jede Einzelkompetenz zwei Situationen, wobei ein/e fiktive/r Lehrer/in in *Situation 1 standardkonforme* Handlungen oder Überlegungen durchführt und in *Situation 2 nicht standardkonform* handelt.

Diese bipolare Operationalisierung basiert auf der im Rahmen von Situational Judgement Tests häufig praktiziertem Methodik, Situationen sowohl danach beurteilen zu lassen, wie wahrscheinlich es wäre, das beschriebene Verhalten selbst zu zeigen, als auch danach, wie wahrscheinlich es wäre, das beschriebene Verhalten selbst *nicht* zu zeigen (Motowidlo, Dunnette & Carter, 1990). Weekley, Ployhart und Holtz (2006) zufolge erhöhen sich durch dieses Vorgehen die Varianz sowie die Validität des Tests. Auch Cucina, Vasilopoulos und Leaman (2003) fanden Hinweise darauf, dass sich die kriteriale Validität von Situational Judgement Tests durch bipolare Items verbessern lässt.

Die Entwicklung der positiv gepolten Items (Vorliegen des Zielkriteriums gemäß der Standards für die Lehrerbildung: Bildungswissenschaften) wurde hierbei in Zusammenarbeit mit ExpertInnen aus dem Bereich der Pädagogischen Psychologie vorgenommen. Die negativ gepolten Items entstanden, indem aus den durch die im Rahmen der Vorstudie von den Studierenden genannten „nicht standardkonformen“ Kompetenzen diejenigen Nennungen als Basis für Situationsvignetten gewählt wurden, die sich zur Schilderung einer Situation eigneten, prägnant auf die entsprechende zugrundeliegende Kompetenz bzw. den KMK-Standard hinwiesen und gleichzeitig nicht zu augenfällig – und somit zu leicht für ein Item - erschienen. Beispiele für ungeeignete Items sind hier etwa die erfolgten Antworten „Kinder sind zum Transfer gar nicht in der Lage.“ oder „Ich denke, Kinder können das oder nicht.“ auf das offene Item „Wie erreichen Sie Transferleistungen Ihrer SchülerInnen (Beispiele)?“

Bei der Gestaltung der Items/Situationsvignetten sowie des Antwortformats wurden folgende Aspekte berücksichtigt:

a) Items/Situationsvignetten

Nach Terhart (2007) soll die Entwicklung eines Instrumentes der Erfassung professioneller pädagogischer Kompetenz im Lehrberuf einer „Beurteilung des Entwicklungsstandes der beruflichen Kompetenzen insgesamt ..., nicht eine[r] Beurteilung spezieller Teilkompetenzen für spezifische Gruppen von Lehrern“ (Terhart, 2007, S. 48) dienen. Terhart (ebd.) plädiert also für eine *fachübergreifende* Diagnostik professioneller pädagogischer Kompetenz, und auch Schmidt et al. (2007) zufolge stellt das Zurverfügungstellen fachübergreifender Lerngelegenheiten einen wichtigen Anteil der Lehrerbildung. Zudem sehen sich Lehrkräfte Anforderungen ausgesetzt, die zwar fachspezifisch ausgestaltet werden müssen, trotzdem aber für alle Fächer gelten (z. B. Klassenführung).

Da außerdem eine standardorientierte Erfassung professioneller pädagogischer Kompetenz für die Lehramtsstudierenden sämtlicher Fächer und Schulformen gleichermaßen zu leisten ist, wurde der Situational Judgement Test so gestaltet, dass in den Items fachspezifische Inhalte weitgehend ausgeklammert werden und somit eine fachübergreifende Evaluation ermöglicht wird. Hiermit wird dem Ansatz Osers und Oelkers (2001) gefolgt: Die Autoren beziehen die fachbezogene Kompetenz in ihre Standardformulierungen bewusst nicht ein.

Der Kontextgehalt der einzuschätzenden Situationen sollte innerhalb des Situational Judgement Tests folglich gering gehalten werden, um zu Items gelangen zu können, die für Lehramtsstudierende sämtlicher Fächer und Schulformen aussagekräftig sind. Außerdem sollte durch eine kurze, wenig komplexe Situationsschilderung sowohl der Einfluss der Lesefähigkeit als der kognitive „Load“ gering gehalten werden (McDaniel, Whetzel, Hartman, Nguyen und Grubb, 2006). Zuletzt wurde eine Eindimensionalität der Situationsvignetten hinsichtlich zugrunde liegender Kompetenzen angestrebt.

b) Antwortformat

Erste Vorstudien mit als Multiple-Choice-Fragen gestalteten Items mit drei Antwortalternativen zeigten eine zu geringe Schwierigkeit. Als Grund hierfür wurde die direkte Gegenüberstellung der richtigen und falschen Antwort sowie das Forced-Choice-Format gesehen, welches zur Wahl einer Alternative „zwingt“ und somit

nicht zur Klärung der Frage geeignet ist, ob eine Antwortalternative auch dann gewählt oder als effektiv eingeschätzt würde, wenn kein direkter Vergleich zu anderen Alternativen möglich ist.

Von einer Gestaltung der Items im Multiple-Choice-Format wurde folglich zugunsten einer bipolaren Likert-skalierten Effektivitätseinschätzung Abstand genommen. Die Polung der Items wurde gewählt, um über eine Beantwortung der Frage „Werden die standardorientierten Handlungen/Überlegungen als effektiv eingeschätzt?“ hinaus auch überprüfen zu können, ob die Studierenden die nicht standardkonformen Handlungen/Überlegungen als solche identifizieren können oder diese als (ebenso) effektiv einschätzen. Die Vorteile einer bipolaren Erhebung der „effektivsten“ und der „ineffektivsten“ Antwortoption liegen zum Einen in der erhöhten Varianz der Testergebnisse, zum Anderen scheint sich die Kriteriumsvalidität hierdurch verbessern zu lassen.

McDaniel, Hartmann, Whetzel und Grubb (2007) zufolge zählen Situational Judgement Tests mit Antwortinstruktionen in Form von Effektivitätseinschätzungen zu den wissensbasierten Messungen. Die Autoren postulieren, dass solcherlei Situational Judgement Tests zu den Messungen maximalen Verhaltens zählen, da sie eine Einschätzung hinsichtlich der (maximalen) Effektivität erfordern. Die Vorteile dieses Formats liegen in der relativen Unabhängigkeit von individuellen Differenzen in der Selbstwahrnehmung und vom „impression management“ (Palhaus, 1984). Zudem scheinen Antwortinstruktionen mit Effektivitätseinschätzungen zu höheren internen Konsistenzen zu führen (Ployhart & Ehrhart, 2003) und weniger verfälschbar zu sein (Hooper, Cullen & Sackett, 2006). Ein Grund hierfür könnte darin liegen, dass mit solchen Antwortinstruktionen tatsächliches eigenes Verhalten nicht direkt angesprochen wird (anders als in Antwortinstruktionen mit der Frage „Was würden Sie tun?“) und somit Verfälschungstendenzen weniger ins Gewicht fallen.

Die Entscheidung gegen ein Multiple-Choice-Format mit eindeutig richtigen und falschen Lösungen fiel zusätzlich auch aus dem Grunde, dass beim Handeln in realen Situationen selten nur *eine* eindeutig richtige Alternative (im Gegensatz zu weiteren eindeutig falschen Handlungsoptionen) denkbar ist (Hedlund & Sternberg, 2000). Ein gestuftes Antwortformat (Einschätzung der Effektivität der dargestellten Handlung) vermeidet die erzwungene Wahl einer Handlungsalternative und trägt so der Mehrdeutigkeit realer Situationen eher Rechnung.

Und zuletzt fiel die Entscheidung für Effektivitätseinschätzungen, weil eine Beurteilung hinsichtlich dessen geleistet werden sollte, ob das in der Situationsvignette geschilderte Verhalten bezüglich der verschiedenen, jeweils durch die Vignetten implizierten, Zielkriterien von Unterricht (z. B. Fachwissen und „Schlüsselkompetenzen“, Wertorientierungen, soziale Kompetenz,...; z. B. Weinert, 2000) geeignet ist. Der Begriff der Unterrichtseffektivität wird in einschlägiger Literatur vielfach verwandt (z. B. Kunter, Kleickmann, Klusmann & Richter, 2011; Klieme et al., 2007; Hartig, Klieme, Jude, Jurecka, Kröhne, Maag-Merki, Reef, & Wirt, 2007; Borich, 2007; Seashore, Louis und Leithwood, 1998), sodass davon ausgegangen werden kann, dass der Begriff Studierenden des Lehramts geläufig ist.

Absichtlich wurde dieses Itemformat allgemein gehalten und auf eine spezifische Abfrage bezüglich des Vorliegens bestimmter Kompetenzen innerhalb der Situationsvignetten verzichtet, um nicht bereits durch die Fragestellung kognitive Anker zu setzen und die Aufmerksamkeit zu lenken. Zudem sind allgemeine Effektivitätseinschätzungen im Rahmen von Situational Judgement Tests gut erprobt (z. B. für den Bereich der pädagogischen Kompetenz: Bruder, 2011; Chan & Schmitt, 2002).

Von einer möglichen multi-dimensionalen Einschätzung der Situationsvignetten wurde Abstand genommen, um das mit 102 Items bereits sehr ausführliche Instrument nicht weiter zu überfrachten und die Wahrscheinlichkeit des Auftretens von Ermüdungsfaktoren bei der Bearbeitung zu erhöhen.

Eine weitere Option wäre gewesen, bei jeder Vignette nach dem „Erfüllungsgrad“ des zugehörigen Standardbereichs zu fragen („Wird Kompetenz X in der geschilderten Situation gezeigt?“). Dann wäre die Zielsetzung jedoch eine andere gewesen: Nicht der Rückschluss von der Itembeantwortung auf eigenes *Verhalten*, sondern das *Erkennen* der Manifestation von Standards wäre in den Focus gerückt.

Abbildung 11 zeigt beispielhaft die zusammengehörigen Items „Problemorientiertes Lernen und Lehren“ (positiver Pol) sowie „Problemorientiertes Lernen und Lehren“ (negativer Pol) im Kompetenzbereich 2 („Lehrerinnen und Lehrer unterstützen durch die Gestaltung von Lernsituationen das Lernen von Schülerinnen und Schülern. Sie

motivieren Schüler und befähigen sie, Zusammenhänge herzustellen und Gelerntes zu verknüpfen“).

Für wie effektiv halten Sie das Verhalten/die Überlegungen der jeweiligen Lehrperson?							
1=sehr ineffektiv 2= ineffektiv 3=eher ineffektiv 4=eher effektiv 5= effektiv 6=sehr effektiv Bitte kreuzen Sie die Ziffer an, die Ihrer Einschätzung entspricht.							
29.	Lehrer K möchte im Physik-Unterricht die Elektrik-Kenntnisse seiner SchülerInnen vertiefen. Lehrer K sagt: „Ich erkläre euch nun, wie ein Radio funktioniert.“	1	2	3	4	5	6
96.	Lehrer K möchte im Physik-Unterricht die Elektrik-Kenntnisse seiner SchülerInnen vertiefen. Lehrer K sagt: „Wir wollen zusammen ein Radio bauen. Was brauchen wir dafür?“	1	2	3	4	5	6

Abbildung 11. Beispielitems „Problemorientiertes Lernen und Lehren“

Um lokale stochastische Unabhängigkeit der Items zu erreichen, wurde die Reihenfolge der Items per Zufall bestimmt und zusätzlich darauf geachtet, dass sich zum Einen zwei zusammengehörige Items nicht direkt hintereinander befinden und zum Anderen Items desselben Kompetenzbereichs nicht aufeinander folgen.

Abbildung 12 stellt die ersten drei Schritte der Itemgenerierung dar.

### 3.2.4 4. Schritt: Eignungseinschätzung der generierten Items

Nach der Auswahl geeigneter Items aus dem Self-Assessment-Instrument KOSTA (ebd.) sowie der qualitativen Erfassung professionellen pädagogischen Handelns angehender Lehrpersonen und der darauf basierenden Itementwicklung wurden die aus diesen Vorarbeiten resultierenden Items einer unabhängigen Gruppe aus ExpertInnen zur Eignungseinschätzung der Items vorgelegt. Die Methode des Expertenratings wird im Rahmen der Entwicklung kontextbasierter Instrumente erfolgreich eingesetzt, um die inhaltliche Validität der Items abzusichern (z. B. Bruder et al., 2011; Lievens, Sackett & Buyse, 2009).

Die ExpertInnen wurden in Internet-Foren zu Themen aus Pädagogik, Schule und Lehramt rekrutiert, um ExpertInnen aus dem gesamten Bundesgebiet zu erreichen. Hierbei wurden als „ExpertInnen“ insbesondere Personen verstanden, die auf langjährige praktische Erfahrung im Lehrberuf verfügen oder aber sich forschend mit dem Gebiet der Pädagogik oder der Pädagogischen Psychologie auseinandersetzen.

Die ExpertInnengruppe wurde interdisziplinär zusammengestellt, da der Ertrag einer ExpertInnen-Befragung dann erhöht wird, „wenn sich die Expertenrunde durch Interdisziplinarität auszeichnet und nicht allein durch Vielfalt in gleicher Weise spezialisierter Experten.“ (Hasse, 1999, zit. nach Häder, 2009, S. 101). Nach Häder (ebd.) lassen sich auch mit einer kleinen Anzahl an ExpertInnen mehrere Perspektiven abdecken, da sich die Wahrnehmung bestimmter Sachverhalte vor allem als abhängig von der professionellen Perspektive der ExpertInnen erweise.

An der Befragung nahmen somit ExpertInnen aus den Bereichen Erziehungswissenschaft, Pädagogik sowie Lehramt teil. Sechs ExpertInnen beantworteten quantitative und qualitative Fragen zu den Items des Situational Judgement Tests (zu den Ergebnissen siehe Abschnitt 4.1.2).

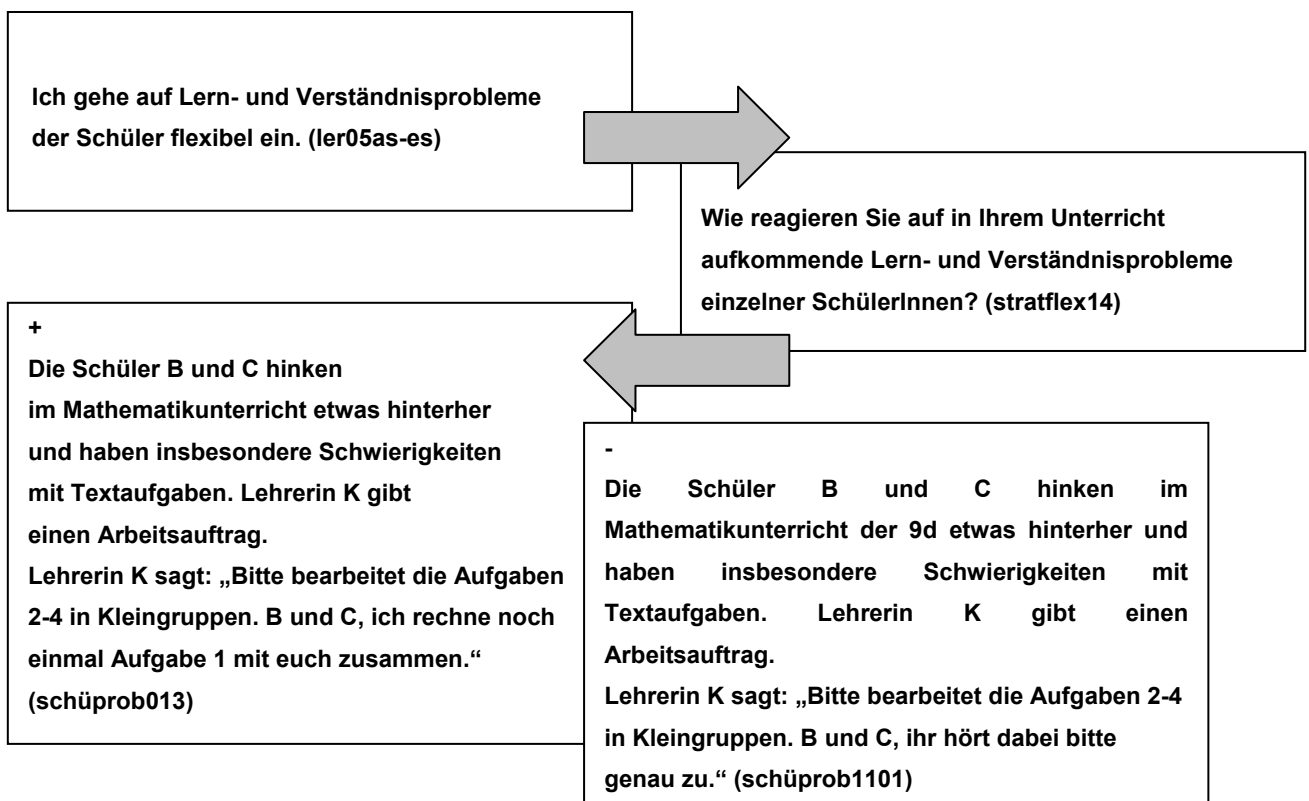


Abbildung 12. Schritt 3 der Itemgenerierung: Entwicklung positiv und negativ gepolter Situationen

Für die ExpertInnen wurde eine anonymisierte e-mail-Adresse eingerichtet, über welche sie mit der Autorin kommunizierten bzw. einen elektronischen Fragebogen zur Eignungsbeurteilung der generierten Items erhielten und zurücksendeten.

Der elektronischer Fragebogen sollte in zwei Wellen ausgegeben werden, wobei bei der zweiten Welle im Fragebogen die Ergebnisse der ersten Befragungswelle integriert, d. h. modifizierte Items vorgegeben werden sollten.

### Das Befragungsinstrument

Zu jedem Item des Fragebogens wurden die ExpertInnen um ihre Einschätzung gebeten, ob das betreffende Item eine bestimmte Kompetenz bzw. das Nichtvorliegen einer bestimmten Kompetenz repräsentiert. Im Falle eines positiv gepolten Items („standardkonformes Verhalten“, +) wurde beispielsweise gefragt: „Weist das genannte Verhalten darauf hin, dass der Unterricht individuellen Lernvoraussetzungen angepasst wird?“. Sollte ein negativ gepoltes Item („nicht standardkonformes Verhalten“, –) eingeschätzt werden, so lautete die Frage etwa: „Weist das genannte Verhalten darauf hin, dass der Unterricht *nicht* individuellen Lernvoraussetzungen angepasst wird?“.

Beide Fragen konnten mit ja oder nein beantwortet werden. Im Falle der Bewertung eines Items als ungeeigneter Indikator wurde um alternative Vorschläge gebeten (siehe die Abbildungen 13 und 14).

Die Schüler B und C hinken im Mathematikunterricht etwas hinterher und haben insbesondere Schwierigkeiten mit Textaufgaben. Lehrerin K gibt einen Arbeitsauftrag. Lehrerin K sagt: „Bitte bearbeitet die Aufgaben 2-4 in Kleingruppen. B und C, ich rechne noch einmal Aufgabe 1 mit euch zusammen.“		<b>+</b>
Weist das genannte Verhalten darauf hin, dass der <b>Unterricht individuellen Lernvoraussetzungen angepasst</b> wird?	Ja	Nein
Falls nein: Meine Vorschläge für ein Item, das darauf hinweist, dass der Unterricht individuellen Lernvoraussetzungen angepasst wird:		

Abbildung 13. Beispielitem der ExpertInnen-Befragung „Simulationsorientierte Erfassung professioneller Kompetenzen im Lehrberuf“/positiver Pol

<p>Die Schüler B und C hinken im Mathematikunterricht der 9d etwas hinterher und haben insbesondere Schwierigkeiten mit Textaufgaben. Lehrerin K gibt einen Arbeitsauftrag. Lehrerin K sagt: „Bitte bearbeitet die Aufgaben 2-4 in Kleingruppen. B und C, ihr hört dabei bitte genau zu.“</p>	-
<p>Weist das genannte Verhalten darauf hin, dass der Unterricht <b>nicht</b> individuellen Lernvoraussetzungen angepasst wird?</p>	Ja
<p>Falls nein:          Meine Vorschläge für ein Item, das darauf hinweist, dass der Unterricht <b>nicht</b> individuellen Lernvoraussetzungen angepasst wird:</p>	

*Abbildung 14.* Beispielitem der ExpertInnen-Befragung „Simulationsorientierte Erfassung professioneller Kompetenzen im Lehrberuf“/negativer Pol

Die aufgrund der ExpertInneneinschätzungen als geeignet befundenen und gegebenenfalls modifizierten Items wurden schließlich für den zu entwickelnden Situational Judgement Test ausgewählt. Der Prozess der Operationalisierung ist folglich nach dem vierten Schritt abgeschlossen. Abbildung 15 stellt die vier Schritte der Itementwicklung vollständig dar.



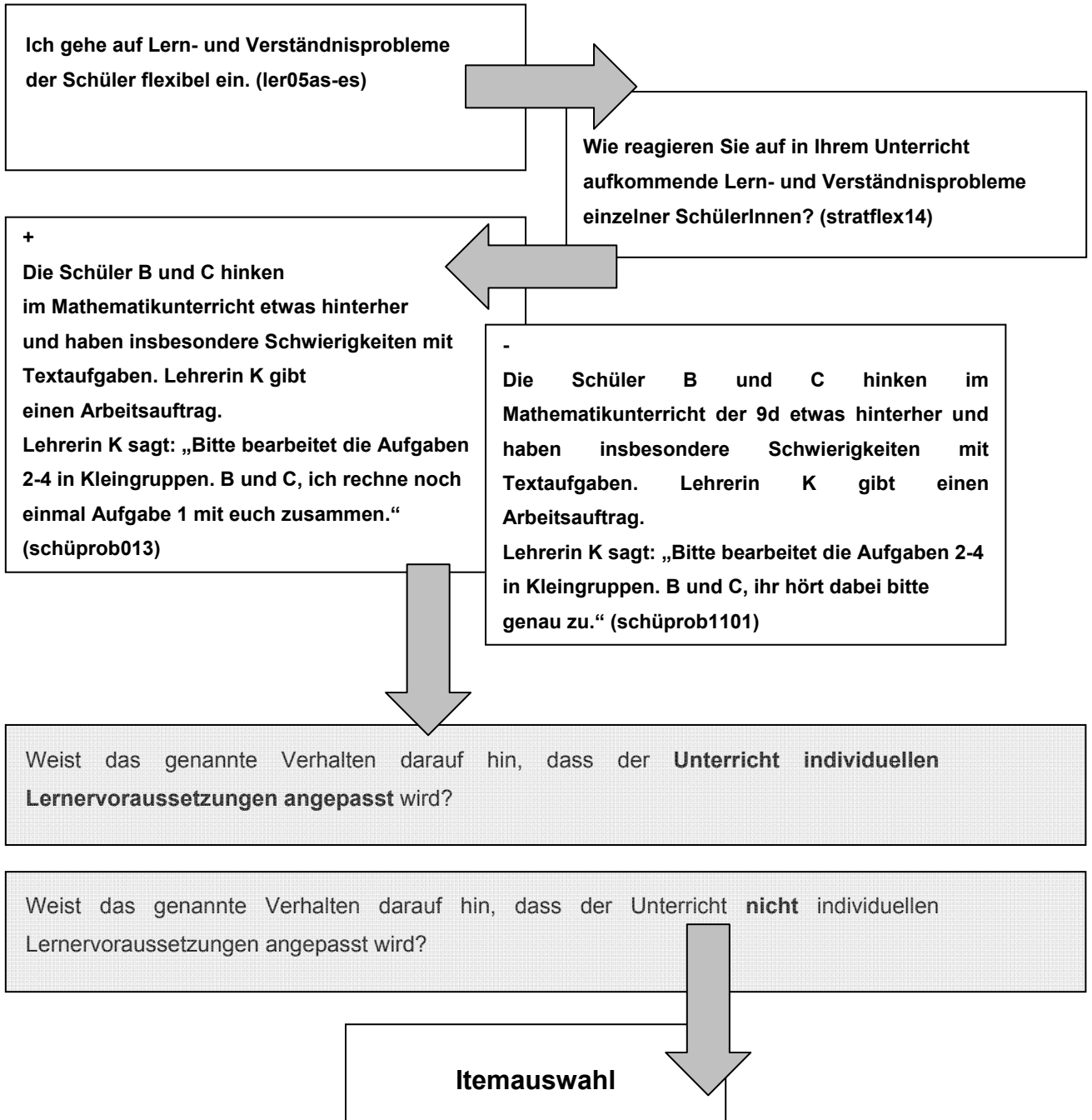


Abbildung 15: Vollständiger Prozess der Itemgenerierung

## 4. Ergebnisse

Die Ergebnisse der Eignungseinschätzung der entwickelten Items sowie der explorativen Validierungsstudie werden in den folgenden Abschnitten dargelegt und diskutiert. Hierbei wird jeweils zunächst die angewandte Methode vorgestellt und die Durchführung der Untersuchung sowie die Stichproben beschrieben, bevor die deskriptiven Daten und ggf. die Ergebnisse der Itemanalysen, der Überprüfung der Gütekriterien Objektivität, Reliabilität und Validität sowie der Nebengütekriterien aufgezeigt werden.

### 4.1 Eignungseinschätzung der Items durch ExpertInnen

Die gemeinsam mit ExpertInnen und auf Basis von studentischen Kompetenznennungen generierten Items sollten hinsichtlich ihrer Tauglichkeit von „ExpertInnen des Lehrens“ eingeschätzt werden. Die Bewertung der Eignung entwickelter Items, das Vorliegen oder Nichtvorliegen von – durch die Lehrerbildungs-Standards vorgegebenen – professionellen pädagogischen Kompetenzen zu markieren, erfolgte im Rahmen einer ExpertInnen-Befragung mit sechs ExpertInnen aus Lehramt, Erziehungswissenschaft und Pädagogik.

#### 4.1.1 *Die ExpertInnengruppe*

Zwei der sechs ExpertInnen gaben an, seit über 15 Jahren im Schuldienst tätig zu sein, ein Experte war zwischen 10 und 15 Jahren dort tätig, zwei weitere ExpertInnen unterrichteten zwischen einem und fünf Jahren und ein letzter Experte wies keine Erfahrung im Schuldienst auf. Bezüglich der Schulart wurden Grund-, Haupt-, Förder- und Realschule sowie Gymnasium je einmal genannt, zwei ExpertInnen unterrichteten an einer „anderen“ Schulart (Mehrfachnennungen waren möglich). Zwei ExpertInnen unterrichteten in Rheinland-Pfalz und jeweils ein Experte in Baden-Württemberg und Bayern. Ein Experte lehrt an einer deutschen Schule in Belgien.

Folgende Fächer unterrichteten die ExpertInnen: Deutsch, ästhetische Erziehung, Pädagogik der Lernbehinderten sowie Geistigbehinderten, Mathematik, Haushalt, Textiles Werken, Kunst, Englisch und Geschichte, Sachunterricht und Religion.

Vier ExpertInnen haben ein Lehramtsstudium absolviert, ein Experte arbeitet nach einem Studium der Erziehungswissenschaft (M.A.) als Lehrer, ein Experte ist Diplom-Pädagoge.

#### 4.1.2 *Die Bewertung der generierten Items*

Als Maß für die Übereinstimmung von Eignungseinschätzungen ist die Angabe von Prozentwerten geläufig (Seidel, Blomberg & Stürmer, 2010, auch: Phillips, 1992) und wurde somit auch in der vorliegenden Studie berechnet.

Hinsichtlich der prozentualen Übereinstimmung bei der Eignungseinschätzung der Items ergibt sich folgendes Bild: Bei 15 Items zeigte sich eine Übereinstimmung von 66,7%. 41 Items wurden mit einer Übereinstimmung von 83,3% geratet, und bei den restlichen 46 Items ergab sich eine prozentuale Interrater-Übereinstimmung von 100%. Sämtliche Items sind somit nach Einschätzung der ExpertInnengruppe geeignet, das Vorliegen/Nichtvorliegen standardorientierter pädagogischer Kompetenzen zu markieren. Im Falle von Nicht-Übereinstimmungen bezog sich der Dissens meist auf Formulierungen – nicht den Inhalt – der Items. Wurden bezüglich der Formulierung von mehr als einem/r Experten/in gleichartige Verbesserungsvorschläge angegeben, erfolgte eine entsprechende Veränderung des Wortlauts.

Anhang G stellt die detaillierten Ergebnisse der ExpertInnen-Befragung dar.

## 4.2 Explorative Datenanalyse und Gütekriterien

In den folgenden Abschnitten werden die Untersuchungsmethode, die Durchführung der Studie sowie deren deskriptive und inferenzstatistische Ergebnisse dargelegt und schließlich diskutiert. Das Ziel der Hauptuntersuchung besteht in der Klärung der Fragen:

- Lässt sich entlang der Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004) ein inhaltsvalider Situational Judgement Test entwickeln?

- Lassen sich die elf Kompetenzbereiche der Standards für die Lehrerbildung: Bildungswissenschaften als Facetten der professionellen pädagogischen Kompetenz Lehramtsstudierender abbilden?

#### 4.2.1 *Methode*

Die Entwicklung des Situational Judgement hat eine valide, kontextnahe Erfassung der Kompetenzen Lehramtsstudierender entlang der von der KMK (2004) formulierten Standards zum Ziel.

Die Operationalisierung der durch die KMK (ebd.) formulierten elf Kompetenzbereiche erfolgt pro Einzelkompetenz durch Situationsvignetten für die Pole „standardkonform“ bzw. „nicht standardkonform“. Die Situationen wurden hierbei gemeinsam mit ExpertInnen auf Basis der Itemanalyse des Instrumentes KOSTA (Weresch-Deperrois et al., 2009) sowie der Ergebnisse der qualitativen Vorstudie generiert.

Jede Vignette beschreibt hierbei kontextarm das Handeln einer fiktiven Lehrperson in einer für den Lehrberuf typischen Situation. Hierbei wird durch jede Vignette lediglich *eine* Kompetenz (z. B. die Diagnose von Lernvoraussetzungen) dargestellt. Dieses fiktive Handeln soll hinsichtlich seiner Effektivität auf einer bipolaren Likert-Skala mit den Antwortoptionen 1 = *sehr ineffektiv*, 2 = *ineffektiv*, 3 = *eher ineffektiv*, 4 = *eher effektiv*, 5 = *effektiv* sowie 6 = *sehr effektiv* durch die TestteilnehmerInnen bewertet werden. Die nähere Darstellung der Items und deren Entwicklung wurde in Abschnitt 3 beschrieben.

Die Situationsvignetten wurden in einem schriftlich zu bearbeitenden Fragebogen präsentiert, wobei nach ersten Voruntersuchungen eine Bearbeitungszeit von etwa 30 Minuten erwartet wurde.

#### 4.2.2 *Durchführung der Untersuchung*

Zur Gewinnung der Gesamtstichprobe eines angestrebten Umfanges von  $n > 300$  wurden DozentInnen in Lehramtsstudiengängen bzw. technischen Studiengängen an fünf Universitäten bzw. Pädagogischen Hochschulen in Rheinland-Pfalz und Baden-Württemberg per e-mail kontaktiert. Hieraus resultierte die Teilnahme der Universität Koblenz-Landau, Campus Landau, der Hochschule für Forstwirtschaft, Rottenburg, sowie der Hochschule für Wirtschaft, Technik und Soziales in Esslingen.

Die Ausgabe des Situational Judgement Tests „Simulationsorientierte Erfassung professioneller pädagogischer Kompetenzen Lehramtstudierender“ erfolgte in zufällig ausgewählten Lehrveranstaltungen. Die anwesenden Studierenden wurden dort um eine freiwillige Teilnahme gebeten.

Die Bearbeitungszeit wurde während den Erhebungen nicht begrenzt, sodass „speed-Effekte“ ausgeschlossen werden können. Die aufgrund von Vorabanalysen geschätzte Bearbeitungszeit von 30 Minuten wurde im Allgemeinen nicht überschritten – nur sehr vereinzelt benötigten Studierende bis zu 40 Minuten.

Die geringe Anzahl von  $n=6$  Fragebögen wurde nicht oder nicht vollständig bearbeitet. Diese Bögen wurden in die Auswertung nicht mit einbezogen.

Nach Erhebung der Daten wurden diese mit der Statistiksoftware „PASW - Predictive Analytics Software - Statistics 18 von SPSS“ analysiert.

#### 4.2.3 *Stichprobe*

Der Situational Judgement Test „Simulationsorientierte Erfassung professioneller pädagogischer Kompetenz Lehramtsstudierender“ wurde von 332 Personen vollständig bearbeitet. Hiervon studierten 79,2% (263 Personen) in Lehramtsstudiengängen, 20,8% (69 Personen) studierten naturwissenschaftliche oder technische Fächer. Die Stichproben der Lehramtsstudierenden sowie der Studierenden naturwissenschaftlicher oder technischer Fächer werden in den nächsten beiden Abschnitten hinsichtlich deskriptiver Größen vorgestellt.

##### Studierende des Lehramts

70,0% der Lehramtsstudierenden waren weiblich (184 Personen), 28,9% (76 Personen) männlich.

52,1% (137 Personen) der Lehramtsstudierenden waren zum Zeitpunkt der Befragung im zweiten Semester, 22,4% (59 Personen) studierten im ersten, 14,4% (38 Personen) im dritten Semester und 4,2% (11 Personen) im sechsten Semester. Die restlichen 6,9% verteilten sich auf die Semester vier (2,7%, 7 Personen), fünf (1,9%, 5 Personen) sowie sieben bis zehn (0,4 - 0,8%, 1-2 Personen). Somit sind Studierende im zweiten Semester deutlich überrepräsentiert. Im Mittel besuchten die Lehramtsstudierenden zum Zeitpunkt der Erhebung eine Semesterzahl von 2,35.

Sämtliche Lehramtsstudierende (263 Personen) studieren im Bundesland Rheinland-Pfalz.

Bezüglich der Orientierenden und Vertiefenden Praktika gaben 54,8% (144 Personen) an, das erste Orientierende Praktikum absolviert zu haben, 12,2% (32 Personen) hatten bereits das zweite Orientierende Praktikum hinter sich, 3,4% (9 Personen) hatten das dritte Orientierende Praktikum, 5,3% (14 Personen) das erste Vertiefende Praktikum sowie 1,1% (3 Personen) das zweite Vertiefende Praktikum absolviert. 23,3 % (61 Personen) hatten bislang noch kein Orientierendes oder Vertiefendes Praktikum durchlaufen.

Die Mehrheit der Lehramtsstudierenden (12,2%, 32 Personen), die bereits eine Prüfung abgelegt hatten, erreichte in ihrer letzten Prüfung im Lehramtsstudium die Note 2,3. Durchschnittlich ergab sich ein Mittelwert der letzten Prüfungsnote von 2,46 (SD=0,814). 28,5% (75 Personen) hatten im Lehramtsstudium noch keine Prüfung abgelegt. 49,4% (130 Personen) wurden zuletzt in einzelnen Fachwissenschaften geprüft, 14,8% (39 Personen) in bildungswissenschaftlichen Fächern. 5,3% (14 Personen) erhielten Noten in anderen Fächern (z. B. praktische Prüfung in Sport).

Hinsichtlich ihres Berufsziels gab die Mehrheit der Lehramtsstudierenden (41,8%, 110 Personen) die Grund- und Hauptschule an. Es folgt das Gymnasium mit 27,0% (71 Personen), die Förderschule mit 13,7% (36 Personen) und schließlich die Realschule mit 12,2% (32 Personen). 5,3% (14 Personen) der Lehramtsstudierenden hatten sich hinsichtlich ihres Berufsziels noch nicht entschieden.

Im Abitur erhielt die Mehrheit der Lehramtsstudierenden eine Benotung von 2,8 (10,6%, 28 Personen). Der Mittelwert der Abiturnoten beträgt 2,7 (SD=,521). 6 Personen machten bezüglich ihrer Abiturnote keine Angaben.

#### Studierende naturwissenschaftlicher und technischer Studiengänge

Die Mehrheit der Studierenden naturwissenschaftlicher und technischer Studiengänge (40,6%, 28 Personen) studierte Forstwirtschaft. Chemieingenieurwesen und Biotechnologie folgen mit jeweils 2,3% (16 Personen). Fünf Personen (7,2%) studierten Bioenergie, jeweils zwei Personen (2,9%) Informatik und Sensorik.

72,5% (50 Personen) der Studierenden naturwissenschaftlicher und technischer Studiengänge waren männlich, 26,1% (18 Personen) weiblich. Eine Person (1,4%) machte hinsichtlich des Geschlechts keine Angaben.

43,5% (30 Personen) studierten zum Zeitpunkt der Erhebung im ersten, 40,6% (28 Personen) im zweiten Semester. Jeweils 4,3% (3 Personen) waren im dritten und sechsten, 7,2% (5 Personen) im vierten Semester.

Sämtliche Studierende technischer und naturwissenschaftlicher Studiengänge (69 Personen) studieren im Bundesland Baden-Württemberg.

Die Mehrheit der Studierenden erhielt in der letzten Studienprüfung die Note 2,0 (17,4%, 12 Personen). Darauf folgen die Note 1,3 mit 14,5% (10 Personen) bzw. die Noten 1,0 und 2,3 mit jeweils 10,1% (7 Personen). Im Mittel erreichten die Studierenden eine Prüfungsnote von 2,05 (SD=.999). Sämtliche Prüfungsinhalte bezogen sich dabei auf fachwissenschaftliche Thematiken (z. B. Jura, Betriebswirtschaftslehre oder Sensorik).

Bezüglich der Abiturnote der Studierenden naturwissenschaftlicher und technischer Studiengänge ergibt sich ein Mittelwert von 2,49 (SD=.487).

#### 4.2.4 *Explorative Datenanalyse und Gütekriterien*

Zunächst wird die Verteilungsform der 102 Variablen, ihre Schwierigkeit sowie Trennschärfe in den Blick genommen, um Aussagen zu erhalten bezüglich

- der Normalverteilung der Daten als Voraussetzung für die vorzunehmenden statistischen Analysen,
- der Bearbeitungsschwierigkeit der Items sowie
- der Eignung der entwickelten Items, zwischen kompetenten und wenig kompetenten Personen zu diskriminieren.

Zur Überprüfung der Verteilungsform der 102 Variablen wurde zum Einen der Test auf Normalverteilung nach Kolmogorov-Smirnov gerechnet, zum Anderen wurden graphische Analysen der Q-Q-Plots sowie der Histogramme der 102 Variablen vorgenommen.

Der Test auf Normalverteilung nach Kolmogorov-Smirnov (K-S) mit Lilliefors-Korrektur ergab in allen Fällen signifikante Ergebnisse – somit kann aufgrund des K-S-Tests nicht für eine Normalverteilung der Daten plädiert werden. Da der K-S-Test bei großen Stichproben bereits bei relativ unbedeutenden Abweichungen von der Normalverteilung signifikant wird, sollen zwei graphische Überprüfungen der Normalverteilungsannahme die Ergebnisse des Verfahrens ergänzen:

Bei einer Überprüfung der Normalverteilung anhand des Quantile-Quantile-Plots wird jeder beobachtete Wert mit seinem unter Normalverteilung erwarteten Wert gepaart. Im Falle einer Normalverteilung liegen im Diagramm die Werte auf einer Geraden. Abbildung 16 stellt beispielhaft das Q-Q-Diagramm für die negativ gepolte Variable „Auswahl von Methoden und Medien“ vor. Deutlich wird die hinreichende Anpassung der beobachteten Werte an die Gerade. Mit Ausnahme von 11 Variablen zeigt sich nach der Analyse der Q-Q-Plots eine hinreichende Anpassung an die jeweilige Gerade.

Hinsichtlich der Analyse der Histogramme ergeben sich für den größten Teil der Variablen hinreichende Anpassungen an die Normalverteilungskurve. Abbildung 17 zeigt ein Beispiel. 20 der 102 Variablen zeigen sich allerdings stark linksschief und weisen somit auf Deckeneffekte hin. Die Variablen „Reflektion eigener Erwartungen +“ und „Anregung zum Perspektivenwechsel –“ weisen eine zweigipflige Verteilung auf.

Aufgrund der Ergebnisse der graphischen Analysen sowie des Sachverhalts, dass der Kolmogorov-Smirnov-Test bei einer Stichprobengröße von  $n > 200$  äußerst sensitiv ist und der Test dann bereits bei relativ unbedeutenden Abweichungen von der Normalverteilung signifikant wird, fällt die Entscheidung für die Annahme einer Normalverteilung des Großteils der Variablen.

### Schwierigkeit und Trennschärfe der 102 Items

Zur weiteren Analyse der Daten wurden sämtliche *negativ gepolten Items umkodiert*, da bei diesen Items ein niedriger Wert (z. B. 1 = sehr ineffektiv) einer hohen Kompetenzausprägung entspricht. Bei den positiv gepolten Items spiegelt sich dieser Sachverhalt direkt in einem hohen Scoring (z. B. 6 = sehr effektiv).



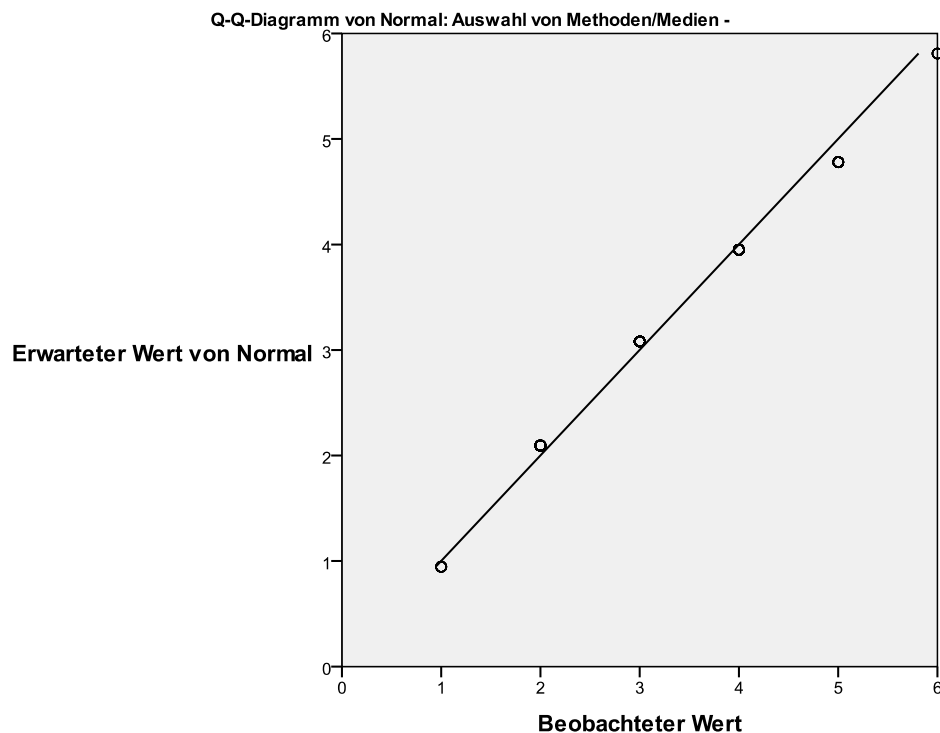


Abbildung 16. Beispielhafte Darstellung eines Q-Q-Plots anhand der Variablen „Auswahl von Methoden und Medien“

Die Itemschwierigkeit macht Aussagen über den Anteil derjenigen ProbandInnen, die ein Item richtig gelöst oder bejaht haben (Bortz & Döring, 2003). Dabei diskriminieren Items mit einer mittleren Schwierigkeit besonders gut zwischen ProbandInnen mit verschiedenen Eigenschaftsausprägungen. Ein Schwierigkeitsindex von  $p=.500$  weist eine solche mittlere Schwierigkeit auf: 50% der ProbandInnen konnten dieses Item lösen. Bei einer Testkonstruktion ist Bortz und Döring (2003) zufolge darauf zu achten, dass die Items hinsichtlich ihrer Schwierigkeit eine breite Streuung aufweisen. Vorgeschlagen wird hier, Itemschwierigkeiten zwischen .2 und .8 zuzulassen.

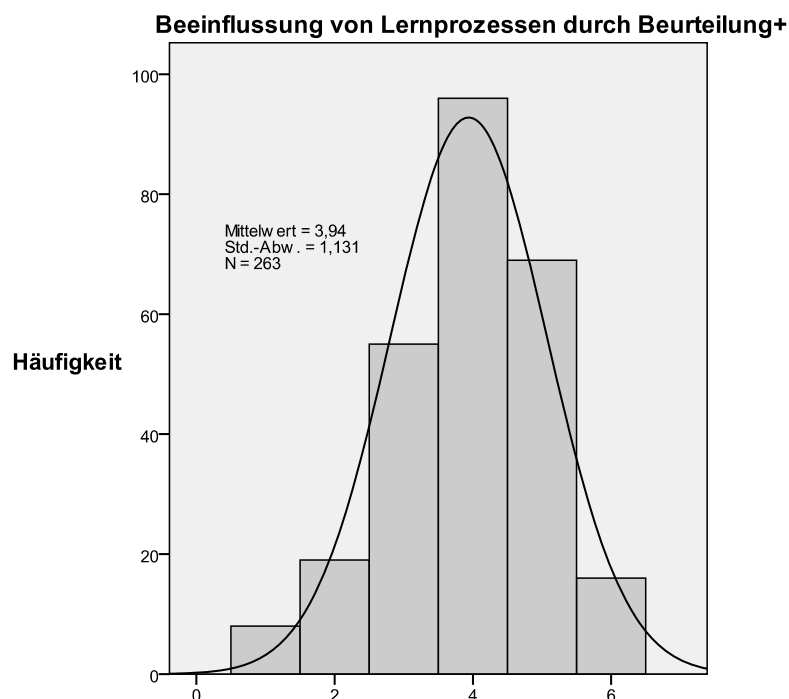


Abbildung 17. Beispielhafte Darstellung eines Histogramms anhand der Variablen „Beeinflussung von Lernprozessen durch Beurteilung +“

Zur Berechnung der Schwierigkeitsindices mehrstufiger Ratingskalen wurde die Formel

$$\bar{X} - 1 / k - 1$$

(bei  $k$  = Anzahl der Ratingsstufen und einer Skala von 1 bis 6) verwandt.

Die Schwierigkeitsindices des Situational Judgement Tests „Simulationsorientierte Erfassung professioneller pädagogischer Kompetenz Lehramtsstudierender“ liegen im Mittel bei  $\bar{X} = p = .673$ , wobei das leichteste Item einen Schwierigkeitsindex von  $p = .870$  aufweist und das schwierigste Item einen Schwierigkeitsindex von  $p = .298$  zeigt.

Items mit einem Schwierigkeitsindex unter  $.2$  finden sich im Test nicht, 15 Items weisen einen Schwierigkeitsindex über  $.8$  auf und sollten folglich aus dem Test ausgeschlossen werden, da sie kaum informativ sind und nur unzureichend zwischen Personen mit verschiedenen Kompetenzausprägungen diskriminieren. Alle Items werden gemeinsam mit ihren Schwierigkeitsindices in Anhang I vorgestellt.

Bei sämtlichen Items wurde die Likert-Skala von 1= sehr ineffektiv bis 6=sehr effektiv vollständig ausgeschöpft.

Nach Bortz und Döring (2003) lässt die Trennschärfe eines Items Rückschlüsse darüber zu, wie gut das Gesamtest- oder Skalenergebnis aufgrund der Beantwortung eines einzelnen Items voraussagbar ist. Berechnet wird hierfür die Korrelation zwischen der Itembeantwortung und dem Skalen- oder Gesamtestwert.

Die skalenbezogenen Trennschärfen der Items des überprüften Instruments liegen im Bereich zwischen  $-.402$  und  $.629$ . Im Mittel ergibt sich eine skalenbezogene Trennschärfe von  $.305$ . Negative Trennschärfen bedeuten hierbei, dass weniger kompetente TestteilnehmerInnen bei dem betreffenden Item entgegen der Erwartung besser abschneiden als kompetentere ProbandInnen. Positive Werte zwischen  $.3$  und  $.5$  gelten als mittelmäßig, Werte jenseits von  $.5$  als groß (Bortz & Döring, 2003). Geringe Trennschärfen sind bei Items in Situational Judgement Tests keine Seltenheit (z. B. McDaniel, Hartman, Whetzel & Grubb, 2007), und auch im vorliegenden Instrument weisen 50 Items eine geringe Trennschärfe  $<.3$  auf.

Tabelle 3 stellt die Ranges der Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Instruments dar. Zusätzlich wird für die auf den Kompetenzbereichen der KMK (2004) basierenden Skalen Cronbachs Alpha (Cronbach, 1951) als Maß der internen Konsistenz berichtet. Die vollständige Tabelle mit den Kenngrößen für jedes Item findet sich in Anhang I.

Tabelle 3. *Ranges der Itemkennwerte*

Skala (Cronbachs Alpha)	Beispielitem (Polung)	<i>M</i> (Range)*	<i>SD</i> (Range)	<i>r<sub>it</sub></i> (Range)	<i>p</i> (Range)
<b>1: Lehrerinnen und Lehrer planen Unterricht fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.</b>  ( $\alpha = .435$ )	Lehrerin B beginnt den Englisch-Unterricht in ihrer Klasse.				
	Lehrerin B sagt: „Wir behandeln heute das ‚future perfect‘. Wir werden erst den Stoff von gestern wiederholen, dann gemeinsam arbeiten und zuletzt die Stunde zusammenfassen.“ (+)	3,23- 4,86	0,912- 1,182	.028- .256	.446- .772
<b>2: Lehrerinnen und Lehrer unterstützen durch die Gestaltung von Lernsituationen das Lernen von Schülerinnen und Schülern. Sie motivieren Schüler und befähigen sie, Zusammenhänge herzustellen und Gelerntes zu verknüpfen.</b>  ( $\alpha = .548$ )	Lehrerin A möchte, dass ihre Schüler das Prinzip des Dividierens gut verstehen.				
	Lehrerin A sagt: „Jonas hat einen Freund zum Kuchenessen eingeladen. Vier Kuchenstücke hat er noch. Wie viele Kuchenstücke bekommt jeder der beiden?“ (+)	3,77- 5,10	0,983- 1,177	.034- .570	.554- .848

Tabelle 3. *Ranges der Itemkennwerte*

Skala (Cronbachs Alpha)	Beispielitem (Polung)	<i>M</i> (Range)*	<i>SD</i> (Range)	<i>r<sub>it</sub></i> (Range)	<i>p</i> (Range)
<b>3: Lehrerinnen und Lehrer fördern die Fähigkeiten von Schülerinnen und Schülern zum selbstbestimmten Lernen und Arbeiten.</b>  ( $\alpha = .273$ )	Ein Schüler hat bei einer Klassenarbeit schlecht abgeschnitten. Lehrer H spricht mit ihm über das Ergebnis.	2,68-	0,922-	-.402-	.336-
	Lehrer H sagt: „Schade, dass du nicht gut abgeschnitten hast. Dir liegt das Fach vielleicht einfach nicht.“ (-)	4,88	1,194	.416	.776
<b>4: Lehrerinnen und Lehrer kennen die sozialen und kulturellen Lebensbedingungen von Schülerinnen und Schülern und nehmen im Rahmen der Schule Einfluss auf deren individuelle Entwicklung.</b>  ( $\alpha = .373$ )	Schülerin J spricht in den Pausen häufig begeistert von der Kfz-Werkstatt ihres Onkels. Lehrer B möchte dieses Interesse fördern.	2,98-	1,038-	-.139-	.640-
	Lehrer B sagt: „Möchtest du an der Technik-AG teilnehmen?“ (+)	5,00	1,342	.359	.800
<b>5: Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern.</b>  ( $\alpha = .709$ )	Lehrer O möchte Umgangsregeln in seiner Klasse einrichten. Lehrer O sagt: „Ich habe einen Regelkatalog ausgearbeitet, an den ihr euch halten sollt.“	3,69-	0,996-	.021-	.538-
	(-)	5,03	1,200	.567	.806

Tabelle 3. *Ranges der Itemkennwerte*

Skala (Cronbachs Alpha)	Beispielitem (Polung)	<i>M</i> (Range)*	<i>SD</i> (Range)	<i>r<sub>it</sub></i> (Range)	<i>p</i> (Range)
<b>6: Lehrerinnen und Lehrer finden Lösungsansätze für Schwierigkeiten und Konflikte in Schule und Unterricht.</b>  ( $\alpha = .603$ )	In Lehrer Zs Klasse haben sich zwei Schülergruppen gebildet, die sich feindlich gegenüber stehen. Lehrer Z thematisiert die Situation. Lehrer Z sagt: „In der Klasse haben sich zwei Lager gebildet, die sich bekämpfen. Lasst uns darüber sprechen, wie wir das Problem lösen können.“ (+)	3,90- 5,35	0,904- 1,106	.174- .594	.610- .870
	<b>7: Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern.</b>  ( $\alpha = .632$ )	Lehrerin As Klasse schneidet in einem standardisierten Rechentest eher schlecht ab. Lehrerin A überlegt, was sie tun sollte. Lehrerin A denkt: „Der Test stimmt mit meinem Eindruck von der Klasse nicht überein, deshalb sehe ich keinen Handlungsbedarf.“ (-)	3,85- 4,90	0,818- 1,176	.206- .460

Tabelle 3. *Ranges der Itemkennwerte*

Skala (Cronbachs Alpha)	Beispielitem (Polung)	<i>M</i> (Range)*	<i>SD</i> (Range)	<i>r<sub>it</sub></i> (Range)	<i>p</i> (Range)
<b>8: Lehrerinnen und Lehrer erfassen Leistungen von Schülerinnen und Schülern auf der Grundlage transparenter Beurteilungsmaßstäbe.</b>  ( $\alpha = .386$ )	Lehrerin U kündigt eine Deutsch-Klassenarbeit mit dem Thema „Erörterung“ an.	3,15-	0,961-	-.103-	.500-
	Lehrerin U sagt: „Ich werde vor allem darauf achten, ob ihre Gedanken klar und differenziert darstellt.“ (+)	5,12	1,335	.324	.842
<b>9: Lehrerinnen und Lehrer sind sich der besonderen Anforderungen des Lehrerberufs bewusst. Sie verstehen ihren Beruf als ein öffentliches Amt mit besonderer Verantwortung und Verpflichtung.</b>  ( $\alpha = .749$ )	Die LehrerInnen einer Schule diskutieren über die rechtlichen Rahmenbedingungen des Lehrberufs.	3,54-	0,933-	.074-	.508-
	Lehrerin W sagt: „Ich kenne die Rahmenbedingungen nicht genau, da es selbstverständlich ist, dass ich mich korrekt verhalte.“ (-)	5,22	1,256	.664	.844
<b>10: Lehrerinnen und Lehrer verstehen ihren Beruf als ständige Lernaufgabe.</b>  ( $\alpha = .805$ )	Lehrer C hat einen wissenschaftlichen Artikel über effektive allgemeine Unterrichtsmethoden gelesen.	3,93-	0,855-	.319-	.586-
	Lehrer C denkt: „Dies lässt sich in meine Praxis nicht übertragen.“ (-)	5,20	1,101	.607	.840

Tabelle 3. *Ranges der Itemkennwerte*

Skala (Cronbachs Alpha)	Beispielitem (Polung)	<i>M</i> (Range)*	<i>SD</i> (Range)	<i>r<sub>it</sub></i> (Range)	<i>p</i> (Range)
<b>11: Lehrerinnen und Lehrer beteiligen sich an der Planung und Umsetzung schulischer Projekte und Vorhaben.</b>  ( $\alpha = .718$ )	Einige Schüler von Lehrer Ks zehnter Klasse möchten eine Klassenzeitung gründen und berichten Lehrer K davon.	2,48- 5,22	1,005- 1,184	-.407- .634	.296- .844
	Lehrer K sagt: „Ich berate euch bei eurer Arbeit.“  (+)				

### Gütekriterien

#### a) Objektivität

Der entwickelte Situational Judgement Test wurde als Paper-and-Pencil-Instrument realisiert, ließe sich jedoch auch als Online-Version umsetzen. Beide Verfahren sind hinsichtlich der *Durchführung* der Erhebung objektiv, da das Instrument mit einer schriftlichen Bearbeitungsinstruktion in standardisierter Form ausgegeben wird. Auch die Objektivität der *Auswertung* ist aufgrund standardisierter Likert-Skalen gewährt: Zur Auswertung müssen lediglich die Markierungen der Skalenstufen abgelesen werden.

Die Objektivität der *Ergebnisinterpretation* ergibt sich aufgrund der numerischen Datensätze, muss allerdings durch die Bereitstellung von Normen oder Kriterien konkretisiert werden: Bestimmte Scores können dann das Vorliegen von Kompetenz oder das Nicht-Vorliegen von Kompetenz in Bezug auf die Testwerte-Verteilung einer Referenzpopulation (normorientiertes Vorgehen) oder in Bezug auf definierte Kriterien (kriterienorientiertes Vorgehen) markieren. Die Generierung von Normen und Kriterien kann erst nach der Testanalyse sinnvoll stattfinden, da zunächst die grundsätzliche Güte des Instruments überprüft werden muss, die Definition von Kriterien bedarf außerdem einer Diskussion unter ExpertInnen der Domäne.

#### b) Reliabilität

Im Folgenden soll für die Items der elf den von der KMK (2004) formulierten Kompetenzbereichen entsprechenden Skalen Cronbachs Alpha (Cronbach, 1951)



als Maß der internen Konsistenz berichtet werden. Tabelle 4 stellt die Befunde im Überblick dar.

Die interne Konsistenz der elf Skalen gemäß Cronbachs Alpha liegt folglich zwischen .273 und .805, wobei die Skalen 1, 3 und 4 lediglich interne Konsistenzen zwischen .273 und .435 aufweisen. Drei Skalen (Skala 2, 6 und 7) des Situational Judgement Tests „Simulationsorientierte Erfassung professioneller pädagogischer Kompetenz Lehramtsstudierender“ erreichen ein Cronbachs Alpha zwischen .500 und .700 und Skala 5, 9, 10 sowie 11 überschreiten den allgemein anerkannten Richtwert von .700 (Zinnbauer & Eberl, 2004).

Für den Gesamttest mit sämtlichen 102 Items ergibt sich ein Cronbachs Alpha von .948.

Tabelle 4. *Cronbachs Alpha der elf Subskalen des Instrumentes*

Skala	Item- anzahl	$\alpha$	Cronbachs Alpha bei entferntem Item
Skala 1: Lehrerinnen und Lehrer planen Unterricht fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.	8	.435	<b>.472</b> Item: „Auswahl von Methoden und Medien“ (meme087)
Skala 2: Lehrerinnen und Lehrer unterstützen durch die Gestaltung von Lernsituationen das Lernen von Schülerinnen und Schülern. Sie motivieren Schüler und befähigen sie, Zusammenhänge herzustellen und Gelerntes zu verknüpfen.	10	.548	<b>.570</b> Item: „Prinzipien des Wissenserwerbs“ (wisserw139)
Skala 3: Lehrerinnen und Lehrer fördern die Fähigkeiten von Schülerinnen und Schülern zum selbstbestimmten Lernen und Arbeiten.	10	.273	<b>.460</b> Item: „Zustandekommen von Arbeitsergebnissen“ (ergeb113)

Tabelle 4. Cronbachs Alpha der elf Subskalen des Instrumentes

Skala	Item- anzahl	$\alpha$	Cronbachs Alpha bei entferntem Item
Skala 4: Lehrerinnen und Lehrer kennen die sozialen und kulturellen Lebensbedingungen von Schülerinnen und Schülern und nehmen im Rahmen der Schule Einfluss auf deren individuelle Entwicklung.	6	.373	<b>.504</b> Item: „Förderung von Schülerinteressen“ (schüinte040)
Skala 5: Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern.	10	.709	<b>.742</b> Item: „Vertrauensvolles Klassenklima“ (klima016)
Skala 6: Lehrerinnen und Lehrer finden Lösungsansätze für Schwierigkeiten und Konflikte in Schule und Unterricht.	8	.603	<b>.642</b> Item: „Perspektivenwechs el“ (pers041)
Skala 7: Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern.	10	.632	Keine Verbesserung durch Streichung von Items möglich
Skala 8: Lehrerinnen und Lehrer erfassen Leistungen von Schülerinnen und Schülern auf der Grundlage transparenter Beurteilungsmaßstäbe.	8	.386	<b>.444</b> Item: „Orientierung an Lernzielen“ (oribeur143)
Skala 9: Lehrerinnen und Lehrer sind sich der besonderen Anforderungen des Lehrerberufs bewusst. Sie verstehen ihren Beruf als ein öffentliches Amt mit besonderer Verantwortung und Verpflichtung.	10	.749	<b>.785</b> Item: „Vorbildfunktion“ (vorbild124)

Tabelle 4. Cronbachs Alpha der elf Subskalen des Instrumentes

Skala	Item- anzahl	$\alpha$	Cronbachs Alpha bei entferntem Item
Skala 10: Lehrerinnen und Lehrer verstehen ihren Beruf als ständige Lernaufgabe.	10	.805	<b>.806</b> <b>Item:</b> <b>„Aktualisierung von Wissen“</b> <b>(aktwiss076)</b>
Skala 11: Lehrerinnen und Lehrer beteiligen sich an der Planung und Umsetzung schulischer Projekte und Vorhaben.	10	.718	<b>.808</b> <b>Item:</b> <b>„Wissensbasis der Projektplanung“</b> <b>(infoplan148)</b>

### c) Validität

Die folgenden Abschnitte behandeln das Gütekriterium der Validität und gehen somit folgender Frage nach: Misst das Verfahren „Simulationsorientierte Erfassung professioneller pädagogischer Kompetenz Lehramtsstudierender“ das, was es zu messen vorgibt? (Bortz & Döring, 2003). Hierbei werden die Aspekte der Inhalts-, der Kriteriums- sowie der Konstruktvalidität beleuchtet.

#### *Inhaltsvalidität*

Ein Instrument ist dann inhaltsvalide, wenn seine Items den interessierenden Merkmals- oder Verhaltensbereich erschöpfend erfassen (Bortz & Döring, 2003). Die Inhaltsvalidität kann nicht aufgrund statistischer Analysen erfasst werden, sondern wird im Rahmen subjektiver ExpertInneneinschätzungen beurteilt (ebd.).

Die Inhaltsvalidität des vorliegenden Situational Judgement Tests wurde durch eine multidisziplinäre ExpertInnengruppe gerated beziehungsweise durch die Basis aus von ExpertInnen generierten Items des Instrumentes KOSTA (ebd.) gewährt. Sämtliche Items wurden dabei als geeignet beurteilt. Zur genaueren Darstellung sei auf Abschnitt 4.1.2 verwiesen.

#### *Kriteriumsvalidität*

Ein Instrument gilt dann als kriteriumsvalide, wenn dessen Resultate der Merkmalsmessung mit einem korrespondierenden Kriterium korrelieren (Bortz & Döring, 2003).

Das vorliegende Instrument soll anhand konvergenter und diskriminanter Kriterien im Rahmen einer konkurrenten Validierung auf seine kriteriale Validität überprüft werden. Dabei stützen sich die Analysen auf explorativ erfasste Daten zur Abiturnote, zur Anzahl im Rahmen des Studiums besuchter Praktika (Praxiserfahrung), zur Semesteranzahl (Studiendauer), den Prüfungsnoten in bildungswissenschaftlichen Fächern, zum Berufsziel sowie zum Studienfach.

Für die 11 durch die KMK (ebd.) benannten Kompetenzbereiche bzw. für den Gesamttest wurde hierbei jeweils ein Skalenmittelwert (Score) berechnet.

Für die Variablen *Abiturnote*, *Praxiserfahrung*, *Studiendauer*, *bildungswissenschaftliche Prüfungsnote* sowie *Berufsziel* und die *Testscores* wurden schrittweise multiple Regressionen gerechnet, um aufdecken zu können, ob die genannten Variablen auf die Testleistung in den 11 Skalenscores sowie im Gesamtscore Einfluss besitzen.

Im Ergebnis besitzen alleine die Ausprägungen der Variablen „Berufsziel“ überzufälligen Einfluss auf die Testscores: Für Score 1, 2, 3, 5, 7, 10, 11 und den Gesamtscore wurden Modelle mit Varianzaufklärungen zwischen 51,3% (Score 5: „Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern.“) und 32,3% (Gesamtscore) signifikant. Die Wahl des Berufsziels „Realschule“ beeinflusste hierbei die meisten Testscores (Score 2, 3, 5 – gemeinsam mit dem Berufsziel „Gymnasium“, 10, 11 sowie Gesamtscore), gefolgt vom Berufsziel „Gymnasium“ (Score 1 und 5 – gemeinsam mit dem Berufsziel „Realschule“) sowie „Grund- und Hauptschule“ (Score 7). Für die Scores 4, 6, 8 und 9 wurde kein Modell signifikant. Tabelle 5 stellt die Ergebnisse zusammenfassend dar.

Tabelle 5. Ergebnisse der multiplen Regression zur Vorhersage der Testleistungen

Score	Koeffizienten	Größe	$p$	Modell
				R-Quadrat (Signifikanz)
1	Semester	-,192	.223	<b>Modell 1:</b> <b>.344</b> <b>(.032)</b>
	Grund-/Hauptschule	,019	.915	
	Realschule	-,188	.460	
	<b>Gymnasium</b>	<b>-,449</b>	<b>.032</b>	
	Förderschule	-,044	.790	
	Praktikum	-,139	.376	
	Bilwiss. Prüfungsnote	-,089	.574	
	Abiturnote	,158	.313	
2	Semester	-,041	.814	<b>Modell 1:</b> <b>.440</b> <b>(.005)</b>
	Grund-/Hauptschule	,057	.715	
	<b>Realschule</b>	<b>-,910</b>	<b>.005</b>	
	Gymnasium	-,181	.232	
	Förderschule	,131	.391	
	Praktikum	-,115	.470	
	Bilwiss. Prüfungsnote	-,007	.964	
	Abiturnote	,095	.537	
3	Semester	-,003	.988	<b>Modell 1:</b> <b>.337</b> <b>(.036)</b>
	Grund-/Hauptschule	,086	.600	
	<b>Realschule</b>	<b>-,495</b>	<b>.036</b>	
	Gymnasium	-,128	.424	
	Förderschule	,010	.949	
	Praktikum	-,088	.601	
	Bilwiss. Prüfungsnote	-,010	.949	
	Abiturnote	,117	.471	
4	Semester	,092	.510	.155 (.772)
	Grund-/Hauptschule	-,145	.868	
	Realschule	-,889	.363	
	Gymnasium	-,684	.437	
	Förderschule	-,426	.629	
	Praktikum	-,148	.512	
	Bilwiss. Prüfungsnote	,092	.615	
	Abiturnote	-,063	.808	

Tabelle 5. Ergebnisse der multiplen Regression zur Vorhersage der Testleistungen

Score	Koeffizienten	Größe	$p$	Modell
				R-Quadrat (Signifikanz)
5	Semester	-,042	.800	<b>Modell 2:</b> .513 (.011)
	Grund-/Hauptschule	-,044	.807	
	<b>Realschule</b>	<b>-,956</b>	<b>.011</b>	
	<b>Gymnasium</b>	<b>-,728</b>	<b>.007</b>	
	Förderschule	-,073	.645	
	Praktikum	-,036	.816	
	Bilwiss. Prüfungsnote	,049	.741	
	Abiturnote	,088	.558	
6	Semester	,092	,510	.190 (.545)
	Grund-/Hauptschule	-,145	,868	
	Realschule	-,889	,363	
	Gymnasium	-,684	,437	
	Förderschule	-,426	,629	
	Praktikum	-,148	,512	
	Bilwiss. Prüfungsnote	,092	,615	
	Abiturnote	-,063	,808	
7	Semester	-,161	.306	<b>Modell 1:</b> .335 (.037)
	<b>Grund-/Hauptschule</b>	<b>,454</b>	<b>.037</b>	
	Realschule	-,180	.271	
	Gymnasium	-,100	.572	
	Förderschule	-,182	.302	
	Praktikum	-,092	.562	
	Bilwiss. Prüfungsnote	,143	.368	
	Abiturnote	,106	.504	
8	Semester	,021	,819	.092 (.932)
	Grund-/Hauptschule	-,520	,381	
	Realschule	-,938	,161	
	Gymnasium	-,505	,398	
	Förderschule	-,529	,378	
	Praktikum	,005	,971	
	Bilwiss. Prüfungsnote	-,112	,370	
	Abiturnote	,008	,963	

Tabelle 5. Ergebnisse der multiplen Regression zur Vorhersage der Testleistungen

Score	Koeffizienten	Größe	p	Modell
				R-Quadrat (Signifikanz)
9	Semester	-,032	,789	.185 (.565)
	Grund-/Hauptschule	-,343	,654	
	Realschule	-,974	,258	
	Gymnasium	-,866	,264	
	Förderschule	-,311	,687	
	Praktikum	,022	,910	
	Bilwiss. Prüfungsnote	-,050	,753	
	Abiturnote	-,023	,919	
10	Semester	-,083	.644	<b>Modell 1:</b> <b>.359</b> <b>(.025)</b>
	Grund-/Hauptschule	,092	.572	
	<b>Realschule</b>	<b>-,932</b>	<b>.025</b>	
	Gymnasium	-,225	.153	
	Förderschule	,103	.518	
	Praktikum	-,068	.684	
	Bilwiss. Prüfungsnote	-,003	.985	
	Abiturnote	,045	.781	
11	Semester	-,080	.651	<b>Modell 1:</b> <b>.376</b> <b>(.018)</b>
	Grund-/Hauptschule	,112	.490	
	<b>Realschule</b>	<b>-,830</b>	<b>.018</b>	
	Gymnasium	-,284	.066	
	Förderschule	,134	.393	
	Praktikum	-,104	.528	
	Bilwiss. Prüfungsnote	-,001	.995	
	Abiturnote	,009	.954	
GESAMT	Semester	-,043	.814	<b>Modell 1:</b> <b>.323</b> <b>(.045)</b>
	Grund-/Hauptschule	,169	.304	
	<b>Realschule</b>	<b>-,607</b>	<b>.045</b>	
	Gymnasium	-,302	.055	
	Förderschule	-,081	.613	
	Praktikum	-,066	.689	
	Bilwiss. Prüfungsnote	,011	.944	
	Abiturnote	,056	.731	

Welche Ergebnisse zeigen sich, wenn die Score-Variablen im Sinne einer Gruppierung in „kompetente“ und „nicht-kompetente“ Studierende unterteilt werden? Unter einer normorientierten Perspektive könnten die Skalenscores beispielsweise mithilfe des Mediansplits dichotomisiert werden. Hieraus würden zwei Gruppen aus „kompetenten“ und „nicht-kompetenten“ Studierenden resultieren, wobei die Stichprobe aller TeilnehmerInnen als Referenzpopulation fungierte. Dieses Vorgehen wurde in einem nächsten Schritt realisiert.

Für die Variablen *Abiturnote*, *Praxiserfahrung*, *Studiendauer*, *bildungswissenschaftliche Prüfungsnote* sowie *Berufsziel* und die *Testscores* wurden schrittweise binäre logistische Regressionen gerechnet, um aufdecken zu können, ob die Wahrscheinlichkeit des Eintreffens der Ereignisse „kompetent“ oder „nicht kompetent“ von den oben genannten Variablen abhängt.

Modelle für die Scores 1 („Lehrerinnen und Lehrer planen Unterricht fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.“), 5 („Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern.“), 7 („Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern.“) und 9 („Lehrerinnen und Lehrer sind sich der besonderen Anforderungen des Lehrerberufs bewusst. Sie verstehen ihren Beruf als ein öffentliches Amt mit besonderer Verantwortung und Verpflichtung.“) wurden signifikant.

Hierbei besitzen wiederum insbesondere Ausprägungen der Variablen „Schulart“ Einfluss auf die Testscores: Eine Wahl des Berufsziels Förderschule (B-Koeffizient: 3,312,  $p=.022$ ) beeinflusst gemeinsam mit der Variable „Abiturnote“ (B-Koeffizient: 1,922,  $p=.030$ ) die Wahrscheinlichkeit, auf Score 1 ein Testergebnis über dem Median zu erreichen (Modell: R-Quadrat nach Nagelkerke (1991)=.339,  $p=.004$ ). Gleiches gilt für die Variablen Gymnasium (B-Koeffizient=-1,946,  $p=.029$ ) und Score 5 (Modell: R-Quadrat nach Nagelkerke=.182,  $p=.017$ ), Grund-/Hauptschule und Score 7 (B-Koeffizient=1,925,  $p=.009$ , Modell: R-Quadrat nach Nagelkerke=.241,  $p=.005$ ) sowie Realschule (B-Koeffizient=-2,534,  $p=.044$ ), Gymnasium (B-Koeffizient=-2,688,  $p=.004$ ) und Score 9 (Modell: R-Quadrat nach Nagelkerke=.372,  $p=.002$ ).



Somit klären die Modelle zwischen 37,2% (Realschule/Gymnasium, Score 9) und 18,2% (Gymnasium, Score 5) der jeweiligen Varianz auf.

Tabelle 6 stellt die Ergebnisse vollständig dar.

Tabelle 6. *Ergebnisse der binären logistischen Regression zur Vorhersage dichotomisierter Testleistungen*

Score	Koeffizienten	Größe	$p$	Nagelkerkes R-Quadrat (Signifikanz)
1	Semester			
	Grund-/Hauptschule	,015	,903	
	Realschule	,650	,420	
	Gymnasium	,319	,572	
	<b>Modell 2: Förderschule</b>	<b>,853</b>	<b>,356</b>	<b>Modell 2: .339 (.004)</b>
	Praktikum	<b>3,312</b>	<b>,022</b>	
	Bilwiss. Prüfungsnote	,128	,720	
	<b>Modell 2: Abiturnote</b>	<b>,001</b>	<b>,974</b>	
2	Semester			
	Grund-/Hauptschule	,915	,339	
	Realschule	1,325	,250	
	Gymnasium	2,229	,135	
	Förderschule	,650	,420	Konstante (.041)
	Praktikum	,650	,420	
	Bilwiss. Prüfungsnote	1,018	,313	
	Abiturnote	,628	,428	
3	Semester			
	Grund-/Hauptschule	,007	,934	
	Realschule	,000	,983	
	Gymnasium	,018	,894	
	Förderschule	,003	,957	(.873)
	Praktikum	,086	,770	
	Bilwiss. Prüfungsnote	,086	,770	
	Abiturnote	,185	,667	
		,791	,374	
		1,076	,300	

Tabelle 6. *Ergebnisse der binären logistischen Regression zur Vorhersage dichotomisierter Testleistungen*

Score	Koeffizienten	Größe	$p$	Nagelkerkes R-Quadrat (Signifikanz)
4	Semester	,022	,883	Konstante (.041)
	Grund-/Hauptschule	2,596	,107	
	Realschule	,557	,455	
	Gymnasium	2,600	,107	
	Förderschule	,000	1,000	
	Praktikum	,016	,900	
	Bilwiss. Prüfungsnote	,439	,508	
	Abiturnote	,356	,551	
5	Semester	,415	,519	<b>.182</b> <b>(.017)</b>
	Grund-/Hauptschule	1,071	,301	
	Realschule	3,606	,058	
	<b>Gymnasium</b>	<b>-1,946</b>	<b>,029</b>	
	Förderschule	,000	1,000	
	Praktikum	,180	,672	
	Bilwiss. Prüfungsnote	,186	,666	
	Abiturnote	,050	,824	
6	Semester	1,128	,288	(.873)
	Grund-/Hauptschule	,616	,433	
	Realschule	1,004	,316	
	Gymnasium	3,288	,070	
	Förderschule	1,509	,219	
	Praktikum	,005	,942	
	Bilwiss. Prüfungsnote	,365	,546	
	Abiturnote	,129	,719	
7	Semester	,188	,664	<b>.241</b> <b>(.005)</b>
	<b>Grund-/Hauptschule</b>	<b>1,925</b>	<b>,009</b>	
	Realschule	,068	,795	
	Gymnasium	,471	,493	
	Förderschule	,059	,809	
	Praktikum	,153	,696	
	Bilwiss. Prüfungsnote	,701	,403	
	Abiturnote	,683	,409	

Tabelle 6. *Ergebnisse der binären logistischen Regression zur Vorhersage dichotomisierter Testleistungen*

Score	Koeffizienten	Größe	$p$	Nagelkerkes R-Quadrat (Signifikanz)
8	Semester	1,853	,173	(.425)
	Grund-/Hauptschule	,000	,987	
	Realschule	,626	,429	
	Gymnasium	,003	,953	
	Förderschule	,003	,953	
	Praktikum	2,043	,153	
	Bilwiss. Prüfungsnote	2,056	,152	
	Abiturnote	,164	,686	
9	Semester	,027	,870	<b>Modell 2:</b> <b>.372</b> <b>(.002)</b>
	Grund-/Hauptschule	1,099	,295	
	<b>Modell 2: Realschule</b>	<b>-2,534</b>	<b>,044</b>	
	<b>Modell 2: Gymnasium</b>	<b>-2,688</b>	<b>,004</b>	
	Förderschule	,130	,719	
	Praktikum	,078	,780	
	Bilwiss. Prüfungsnote	,149	,700	
	Abiturnote	,100	,751	
10	Semester	2,327	,127	(.631)
	Grund-/Hauptschule	,161	,688	
	Realschule	3,820	,051	
	Gymnasium	,774	,379	
	Förderschule	1,981	,159	
	Praktikum	,334	,563	
	Bilwiss. Prüfungsnote	,020	,887	
	Abiturnote	,815	,367	
11	Semester	1,733	,188	(.631)
	Grund-/Hauptschule	,161	,688	
	Realschule	3,820	,051	
	Gymnasium	,774	,379	
	Förderschule	1,981	,159	
	Praktikum	,876	,349	
	Bilwiss. Prüfungsnote	,303	,582	
	Abiturnote	,021	,886	

Tabelle 6. *Ergebnisse der binären logistischen Regression zur Vorhersage dichotomisierter Testleistungen*

Score	Koeffizienten	Größe	$p$	Nagelkerkes R-Quadrat (Signifikanz)
GESAMT	Semester			(.265)
	Grund-/Hauptschule	,331	,565	
	Realschule	,139	,709	
	Gymnasium	2,126	,145	
	Förderschule	1,021	,312	
	Praktikum	1,710	,191	
	Bilwiss. Prüfungsnote	,780	,377	
	Abiturnote	,302	,583	
		1,042	,307	

Studierende des *Lehramts* sollten bei der Bearbeitung des Situational Judgement Tests bessere Testergebnisse im Sinne höherer Scores erreichen als Studierende *technischer* und *naturwissenschaftlicher* Studiengänge, da nur erstere eine pädagogische Ausbildung genießen.

Um dies überprüfen zu können, wurden t-Tests für unabhängige Stichproben gerechnet.

Der Levene-Test zur Gleichheit der Varianzen wird für die Scores 2, 3, 4, 6, 9, 10 sowie den Gesamtscore signifikant: Die Gruppen der Lehramtsstudierenden und der Studierenden technischer oder naturwissenschaftlicher Studiengänge weisen folglich in diesen Scores heterogene Varianzen auf, die Freiheitsgrade für den t-Wert müssen somit korrigiert werden (Bühner & Ziegler, 2009).

Hinsichtlich der Variablen „Score 1-11“ und „Gesamtscore“ ergeben sich erwartungsgemäß (hohe) signifikante Gruppenunterschiede. Mit Ausnahme der Variable „Score 4“ ( $p=.025$ ,  $df=95,265$ , bei Ungleichheit der Varianzen und einseitigem Signifikanztest) mit einem lediglich *signifikanten* Ergebnis zeigen sich bei allen Skalenscores sowie beim Gesamtscore *hochsignifikante* Unterschiede ( $p=.000$ ) zwischen den Gruppen „Lehramt“ und „Technik“. Die Score-Mittelwerte der Lehramtsstudierenden liegen hierbei jeweils über den Mittelwerten der Studierenden technischer und naturwissenschaftlicher Studiengänge. Die exakten Daten können Tabelle 7 entnommen werden.

Zur Überprüfung der praktischen Bedeutsamkeit der signifikanten Ergebnisse wurde die für t-Tests mit unabhängigen Stichproben adäquate Effektgröße  $d$  berechnet (Bortz & Döring, 2003). Die Effektstärke  $d$  normiert die Unterschiede zwischen den experimentellen Gruppen auf die Streuung der Testwerte und lautet für t-Tests mit unabhängigen Stichproben:

$$d = \mu_A - \mu_B / \sigma$$

Die Effektstärken der Mittelwertsunterschiede aller 11 Skalenscores sowie des Gesamtscores bewegen sich zwischen 0.44 (Score 4) und 1.81 (Score 3). Mit Ausnahme von Score 4 (kleine bis mittlere Bedeutsamkeit) weisen alle Scores eine nach der Klassifikation von Bortz und Döring (2003) große Effektstärke auf (siehe Tabelle 7).

#### *Konstruktvalidität: Diskriminanzanalyse*

Bei einer Überprüfung der Konstruktvalidität eines Instruments werden Hypothesen bezüglich des Konstrukts bzw. dessen Relationen zu anderen Variablen formuliert (Bortz & Döring, 2003). Fallen bei der Hypothesenprüfung die Testwerte erwartungsgemäß aus, ist ein Instrument oder ein Modell wahrscheinlich konstruktvalide.

Die zum Nachweis der Konstruktvalidität angewandte Faktorenanalyse überprüft Korrelationen zwischen manifesten Variablen, welche durch latente Faktoren erklärbar sein sollen (Rost, 2004). Diese Methode gilt für Situational Judgement Tests als ungeeignet, da die Antworten/Antwortratings neben einer Funktion der Kompetenz auch eine Funktion der vorgegebenen Situation darstellen und somit solcherlei Tests selten zu interpretierbaren Faktorenstrukturen führen (McDaniel & Nguyen, 2001). Nach McDaniel und Whetzel (2005, auch: Schmitt & Chan, 2006) besteht sogar *gar keine* bekannte Faktorenanalyse eines Situational Judgement Tests mit interpretierbaren Faktoren. Bledow und Frese (2009) behaupten folgerichtig, dass jedes situationsbasierte Item ein unabhängiges Phänomen messe.

Stemler und Sternberg (2006) schlagen anstelle einer Faktorenanalyse im Falle eines Situational Judgement Tests die Diskriminanzanalyse zur Datenanalyse vor. Ziel einer Diskriminanzanalyse ist es, eine Diskriminanzfunktion

$$d = b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + \dots + b_n \cdot x_n + a$$

dergestalt zu ermitteln, dass die Werte der Funktion vorab definierte Gruppen möglichst gut trennen. Die Diskriminanzanalyse ermittelt hierbei Gewichte, welche die Bedeutsamkeit der unabhängigen Variablen für die Unterscheidung der Gruppen markieren.

Tabelle 7. *Ergebnisse der t-Tests für unabhängige Stichproben gemeinsam mit den entsprechenden Effektstärken (\*\*  $p = .000$ , \*  $p < .05$ )*

Score	<i>T</i>	<i>df</i>	<i>p</i>	Fach	<i>n</i>	<i>M</i>	<i>SD</i>	<i>d</i>
SCORE1	3,852	330	,000**	Lehramt	263	4,17	,49	0.96
				Technik	69	3,90	,55	
SCORE2	5,102	89,050	,000**	Lehramt	263	4,29	,49	0.88
				Technik	69	3,86	,66	
SCORE3	6,234	86,721	,000**	Lehramt	263	4,10	,39	1.81
				Technik	69	3,62	,56	
SCORE4	1,997	95,265	,049*	Lehramt	263	4,34	,58	0.44
				Technik	69	4,16	,68	
SCORE5	4,067	330	,000**	Lehramt	263	4,51	,57	0.95
				Technik	69	4,20	,58	
SCORE6	4,614	88,260	,000**	Lehramt	263	4,47	,56	0.96
				Technik	69	4,02	,77	
SCORE7	4,984	330	,000**	Lehramt	263	4,38	,51	1.21
				Technik	69	4,03	,57	
SCORE8	6,044	330	,000**	Lehramt	263	4,09	,44	1.74
				Technik	69	3,73	,48	
SCORE9	3,928	92,723	,000**	Lehramt	263	4,54	,59	0.82
				Technik	69	4,17	,73	
SCORE10	4,106	92,533	,000**	Lehramt	263	4,73	,61	0.84
				Technik	69	4,32	,76	
SCORE11	3,945	330	,000**	Lehramt	263	4,59	,56	0.91
				Technik	69	4,29	,59	
GESAMT	5,544	95,932	,000**	Lehramt	263	4,38	,43	1.68
				Technik	69	4,02	,50	

Zur Berechnung einer Diskriminanzanalyse wurde zunächst eine Gruppentrennung im Sinne eines normorientierten Vorgehens für die Variable „Gesamtscore“ mithilfe des Mediansplits vorgenommen. Somit resultierte eine Gruppe, deren Gesamtscore *über* dem Median der Gesamtstichprobe lag, und eine zweite Gruppe, deren Gesamtscore *unter* dem Median der Gesamtstichprobe lag.

Die Voraussetzung der Intervallskalierung für die Durchführung einer Diskriminanzanalyse erfüllen die Prädiktoren „Semester“, „Prüfungsnote“, „Abiturnote“ sowie „Anzahl abgeleiteter Praktika“. Für diese Prädiktoren soll die Gewichtung bzw. Signifikanz hinsichtlich ihrer Bedeutsamkeit bei der Gruppentrennung bestimmt werden.

Im Ergebnis fällt der Korrelationskoeffizient zwischen den Werten der Diskriminanzfunktion und der Gruppenzugehörigkeit – als Maß für die gelingende Trennung zwischen den beiden Gruppen „Gesamtscore < Median“ und „Gesamtscore > Median“ – mit .096 sehr gering aus, und auch der Eigenwert – entsprechend dem F-Wert der Varianzanalyse – zeigt sich mit .009 äußerst gering. Zudem unterscheiden sich die mittleren Werte der Diskriminanzfunktion beider Gruppen nicht signifikant ( $p=.773$ ).

Dies bedeutet, dass eine Trennung der Gruppen „Gesamtscore < Median“ und „Gesamtscore > Median“ aufgrund der Variablen „Semester“, „Prüfungsnote“, „Abiturnote“ sowie „Anzahl abgeleiteter Praktika“ nicht gelingt.

#### *Konstruktvalidität: ExpertInnenbefragung*

Zur qualitativen Analyse der Dimensionalität des Konstruktes „professionelle pädagogische Kompetenz gemäß der KMK-Standards“ wurde anschließend erneut eine unabhängige und interdisziplinäre ExpertInnengruppe konsultiert. Es sollte überprüft werden, ob sich die generierten Items den durch die KMK (ebd.) postulierten elf Kompetenzen inhaltlich zuordnen lassen.

Befragt wurden acht ExpertInnen der Fachbereiche Lehramt, Pädagogik und Psychologie.

Drei ExpertInnen waren hierbei LehrerInnen, ein/e ExpertIn war Pädagoge/in sowie vier ExpertInnen waren PsychologInnen. Drei LehrerInnen unterrichteten an Sonderschulen, ein/e LehrerIn am Gymnasium. Bezüglich der Berufspraxis gaben jeweils drei Personen Arbeitsdauern unter einem Jahr bzw. zwischen einem und fünf Jahren an. Zwei Personen arbeiteten bereits zwischen sechs und zehn Jahren.

Die Befragung erfolgte anonymisiert sowohl online als auch mit paper-pencil-Fragebögen. Hierbei wurden die ExpertInnen gebeten, jedes Item der ihrer Meinung nach passenden durch die KMK formulierten Kompetenz (ebd.) zuzuordnen, wobei

die Angabe mehrere Bereiche möglich war. Die Kompetenzbeschreibungen lagen den ExpertInnen dabei vor.

Erneut wurden als Maß für die Übereinstimmung Prozentwerte angesetzt:

Bei 23 Items wurde eine Übereinstimmung der Zuordnung von 100% (8 Personen) erreicht. Für 27 Items zeigte sich eine Übereinstimmung von 87,5% (7 Personen), bei 19 Items lag eine Übereinstimmung von 75% (6 Personen) vor, und eine Zuordnungsübereinstimmung von 62,5% (5 Personen) bestand bei 15 Items. Lediglich 18 Items wiesen eine Übereinstimmung von  $\leq 50\%$  auf. Somit wurden 84 von 102 Items mehrheitlich übereinstimmend einer KMK-Kompetenz zugeordnet. Von der Möglichkeit, pro Item mehrere Kompetenzbereiche anzugeben, wurde nur vereinzelt Gebrauch gemacht.

48 Items wurden durch die ExpertInnen *mehrheitlich* der bei der Itementwicklung *intendierten* Kompetenz zugeordnet. 10 Items streuten bezüglich der Zuordnung zu einem bestimmten Kompetenzbereich deutlich, ließen sich also nicht eindeutig einer KMK-Kompetenz zugliedern. 18 Items wurden von *keinem/r* Experten/in der *intendierten* Kompetenz zugeordnet. Somit scheinen sich zwar die Items thematisch „clustern“ zu lassen – allerdings scheint die Zuordnung zu bestimmten Kompetenzbereichen nicht immer eindeutig zu sein.

Am häufigsten wurden bei der Zuordnung Kompetenzbereich 2 („Lehrerinnen und Lehrer unterstützen durch die Gestaltung von Lernsituationen das Lernen von Schülerinnen und Schülern. Sie motivieren Schüler und befähigen sie, Zusammenhänge herzustellen und Gelerntes zu verknüpfen.“ - 102 Nennungen) und 10 („Lehrerinnen und Lehrer verstehen ihren Beruf als ständige Lernaufgabe.“ - 97 Nennungen) genannt. Kompetenzbereich 4 („Lehrerinnen und Lehrer kennen die sozialen und kulturellen Lebensbedingungen von Schülerinnen und Schülern und nehmen im Rahmen der Schule Einfluss auf deren individuelle Entwicklung.“ - 42 Nennungen) und 11 („Lehrerinnen und Lehrer beteiligen sich an der Planung und Umsetzung schulischer Projekte und Vorhaben.“ - 41 Nennungen) wurden am seltensten angegeben. Tabelle 8 stellt die Häufigkeiten vollständig dar.

In Anhang J finden sich die detaillierten Ergebnisse bezüglich der durch die ExpertInnen erfolgten Zuordnung der entwickelten Items zu den durch die KMK formulierten Kompetenzen.



Tabelle 8. *Häufigkeit der Nennung von Kompetenzbereichen durch ExpertInnen bei der Itemzuordnung*

Kompetenzbereich	Häufigkeit der Nennungen
1: Lehrerinnen und Lehrer planen Unterricht fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.	91
2: Lehrerinnen und Lehrer unterstützen durch die Gestaltung von Lernsituationen das Lernen von Schülerinnen und Schülern. Sie motivieren Schüler und befähigen sie, Zusammenhänge herzustellen und Gelerntes zu verknüpfen.	102
3: Lehrerinnen und Lehrer fördern die Fähigkeiten von Schülerinnen und Schülern zum selbstbestimmten Lernen und Arbeiten.	57
4: Lehrerinnen und Lehrer kennen die sozialen und kulturellen Lebensbedingungen von Schülerinnen und Schülern und nehmen im Rahmen der Schule Einfluss auf deren individuelle Entwicklung.	42
5: Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern.	81
6: Lehrerinnen und Lehrer finden Lösungsansätze für Schwierigkeiten und Konflikte in Schule und Unterricht.	67
7: Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern.	77
8: Lehrerinnen und Lehrer erfassen Leistungen von Schülerinnen und Schülern auf der Grundlage transparenter Beurteilungsmaßstäbe.	85
9: Lehrerinnen und Lehrer sind sich der besonderen Anforderungen des Lehrerberufs bewusst. Sie verstehen ihren Beruf als ein öffentliches Amt mit besonderer Verantwortung und Verpflichtung.	44
10: Lehrerinnen und Lehrer verstehen ihren Beruf als ständige Lernaufgabe.	97
11: Lehrerinnen und Lehrer beteiligen sich an der Planung und Umsetzung schulischer Projekte und Vorhaben.	41

## 5. Diskussion der Ergebnisse

Die Ergebnisse der vorliegenden Studie sollen in den nächsten Abschnitten hinsichtlich einer Eignung der generierten Items sowie der Haupt- und Nebengütekriterien diskutiert werden, bevor die Arbeit mit einem Ausblick auf mögliche zukünftige Forschungsaufgaben schließt.

### 5.1 Itemanalyse

Die Analyse der für den Situational Judgement Test generierten 102 Items hinsichtlich ihrer Verteilung, Trennschärfe und Schwierigkeit ermöglicht erste Aussagen über die Eignung der Items zur Erfassung professioneller pädagogischer Kompetenzen entlang der Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004):

#### Die Verteilungsform der 102 Items

Die Ergebnisse des Tests auf Normalverteilung nach Kolmogorov-Smirnov sprechen aufgrund signifikanter Resultate für alle 102 Items gegen eine Normalverteilung. Allerdings wird dieser Test für Stichproben mit  $n > 200$  nicht empfohlen, da bei einer solchen Stichprobengröße auch kleinste Verteilungsunterschiede signifikant werden.

Zur Analyse der Verteilungsform der Items wurde folglich zusätzlich auf die graphische Überprüfung anhand der Quantile-Quantile-Plots sowie der Histogramme zurückgegriffen. Beide graphische Tests sprechen für eine Normalverteilung der meisten Items, wobei 20 Variablen eine Linksschiefe aufweisen und zwei Variablen eine zweigipflige Verteilung zeigen.

Eine linksschiefe Verteilung weist auf Deckeneffekte hin – somit scheinen 20 Items zu leicht zu sein. Ein Ausschluss der genannten Variablen muss also in Erwägung gezogen werden. Vor einer endgültigen Entscheidung sollten jedoch Informationen über die Trennschärfe sowie den Schwierigkeitsindex der Items betrachtet werden:

#### Die Trennschärfe der 102 Items

Die Trennschärfen der Items des überprüften Instruments wurden für die 11 Subskalen des Tests berechnet, wobei die skalenbezogenen Trennschärfen im Bereich zwischen

-.402 und .629 liegen und sich eine mittlere Trennschärfe von .305 ergibt. 51 Items des Situational Judgement Tests weisen eine Trennschärfe unter der allgemein angewandten Grenze von .300 (z. B. Fisseni, 2004) auf.

Hierbei muss bedacht werden, dass Situational Judgement Tests meist eine hohe Heterogenität aufweisen und somit geringe Trennschärfen von Einzelitems im Sinne von Item-Skala-Korrelationen keine Seltenheit sind (siehe z. B. die Metaanalyse von McDaniel, Hartman, Whetzel & Grubb, 2007; Möller, 2010). Eine Entscheidung für die Beibehaltung der fraglichen Items könnte folglich unter dem Gesichtspunkt der Inhaltsvalidität fallen: Die Inhaltsvalidität rekurriert auf die Repräsentativität von Testitems für ein zu messendes Merkmal (Moosbrugger & Kelava, 2007). Ausgangspunkt ist hierbei der Repräsentationsschluss – mithin die Annahme, dass die Items eines Tests eine repräsentative Stichprobe aus einem Itemuniversum darstellen. Nach Moosbrugger und Kelava (2007) ist die Frage nach der Inhaltsvalidität dann am einfachsten zu klären, „wenn die einzelnen Items einen unmittelbaren Ausschnitt aus dem Verhaltensbereich darstellen, über den eine Aussage getroffen werden soll.“ (S. 15)

Da die Auswahl und Eignungseinschätzung der Einzelitems sowohl des Basis-Instrumentes KOSTA (ebd.) als auch des resultierenden Situational Judgement Tests durch ExpertInnen vorgenommen wurde, kann von einer inhaltlichen Validität der 102 Testitems ausgegangen werden. Dabei repräsentiert jedes Item eine distinkte Kompetenz und kann somit als notwendig für die annähernd umfassende Wiedergabe des Itemuniversums betrachtet werden.

Außerdem wird bei der Entscheidung über eine Beibehaltung wenig trennscharfer Items das Problem des Reliabilitäts-Validitäts-Dilemmas der klassischen Testtheorie tangiert, welches besagt, dass die Validität eines Tests sinkt, wenn die Trennschärfen der Items steigen. Rost (2004) plädiert dafür, eine größere Heterogenität von Testitems zuzulassen, wenn ein komplexes Kriterium (wie etwa professionelle pädagogische Kompetenz) durch einen Test vorhergesagt werden soll, da in diesem Falle „ein Testergebnis, das *sehr viele* Bedingungen des derart komplexen Kriteriums abdeckt, *höher* mit der Kriteriumsvariable“ korreliert (S. 394). Der kriterialen Validität wird somit im Rahmen der Erfassung komplexer Konstrukte höhere Bedeutung beigemessen als der Reliabilität. In diesem Sinne könnte zunächst die Entscheidung für die *Beibehaltung* der 102 Items fallen.

Von den oben erwähnten Items mit einer linksschiefen Verteilung weist lediglich das Item „Auswahl von Methoden und Medien“ zusätzlich auch eine Trennschärfe unter .300 auf und könnte somit aufgrund des Vorliegens von *zwei* Ausschlusskriterien aus dem Test eliminiert werden.

### Die Schwierigkeitsindizes der 102 Items

Mittlere Itemschwierigkeitsindizes um  $p=.50$  eignen sich in besonderem Maße zur Differenzierung zwischen fähigen und nicht fähigen ProbandInnen. Trotzdem sollte ein Test auch Items höherer und geringerer Schwierigkeit beinhalten, um das Itemuniversum adäquat abbilden und UntersuchungsteilnehmerInnen mit unterschiedlichen Merkmalsausprägungen annähernd gleich gut differenzieren zu können (Bortz & Döring, 2003).

Die Items des vorliegenden Situational Judgement Tests weisen im Mittel einen Schwierigkeitsindex von  $p=.673$  auf. Das schwierigste Item zeigt dabei einen Index von  $p=.298$ , das leichteste weist einen Schwierigkeitsindex von  $p=.870$  auf. 15 Items sollten aufgrund von Indizes über  $p=.800$  ausgeschlossen werden, da sie durch die hohe Lösungswahrscheinlichkeit zu wenig Informationsgehalt bieten. Items mit einem Schwierigkeitsindex unter .200 finden sich nicht.

## 5.2 Haupt- und Nebengütekriterien

Es folgt die Zusammenfassung und Diskussion der Ergebnisse von Analysen zu den Haupt- und Nebengütekriterien Objektivität, Reliabilität, Validität, Verfälschbarkeit, Nützlichkeit, Ökonomie sowie Normierung.

### Objektivität

„Die Objektivität eines Verfahrens gibt an, in welchem Ausmaß die Testergebnisse vom Testanwender unabhängig sind.“ (Bortz & Döring, 2003, S. 194). Dabei lassen sich die Durchführungs-, die Auswertungs- sowie die Interpretationsobjektivität unterscheiden.

Das im Rahmen der vorliegenden Arbeit entwickelte Instrument wird mit einer schriftlichen Bearbeitungsinstruktion und einer einheitlichen Beantwortungsskala in standardisierter Form ausgegeben. Die *Durchführung* des Tests ist somit vom Testleiter unabhängig.

Die *Auswertungsobjektivität* bezeichnet die Unabhängigkeit der Vergabe von Testpunkten für bestimmte Testantworten vom individuellen Testauswerter. Durch eine eindeutige Vorgabe der Form der Itembeantwortung sowie der Bewertung der Antworten (beispielsweise mithilfe eines Scoring-Schlüssels, also der Anweisung, für bestimmte Antworten eine bestimmte Anzahl an Punkten zu vergeben) kann eine objektive Auswertung erreicht werden. Bei dem entwickelten Situational Judgement Test ist die Auswertungsobjektivität aufgrund der Vorgabe von zahlenbasierten Likert-Skalen gewährt: Zur Auswertung müssen lediglich die mit Zahlen von 1 bis 6 versehenen Markierungen der Skalenstufen abgelesen werden.

Unter der *Interpretationsobjektivität* wird eine standardisierte Deutung von Testergebnissen verstanden. Testergebnisse werden hierbei kriteriums- und/oder normorientiert interpretiert. „Die Frage nach der Auswahl eines geeigneten Vergleichsmaßstabs zur Kennzeichnung eines Probanden auf einer psychischen Merkmalsdimension“ (Rettler, 1999, S. 221) steht hier im Mittelpunkt.

Für das überprüfte Instrument ist die Objektivität der Ergebnisinterpretation aufgrund der numerischen Datensätze des Instrumentes grundsätzlich gegeben. Ob daran anschließend ein norm- oder kriteriumsorientiertes Vorgehen angezeigt wäre, ist noch zu diskutieren. Terhart (für das Sekretariat der ständigen Konferenz der Kultusminister der Länder, 2004) empfiehlt dabei eine kriterienorientierte Skalierung der Standards mit einem zugrundeliegenden Modell verschiedener Kompetenzniveaus. Die Vorgabe von Standards impliziert eine kriterienorientierte Interpretation, denn ein Standard ist ein Kriterium, das es zu erreichen gilt. Kriterienorientiert würde das entwickelte Instrument dann interpretiert, wenn bestimmte Schwellenwerte definiert würden, welche ein Vorliegen oder Nicht-Vorliegen professioneller pädagogischer Kompetenzen markierten. Eine solche Definition von Schwellen sollte unter Berücksichtigung der Testwerte-Verteilung in der entsprechenden Population durch ExpertInnen realisiert werden.

Im Falle einer normorientierten Interpretation müssten Normen für verschiedene Subpopulationen von Lehramtsstudierenden (etwa gruppiert nach der Semesteranzahl) zur Verfügung gestellt werden, mit welchen dann die individuelle Testleistung verglichen werden würde.

Die Bereitstellung von Kriterien oder Normen ist allerdings erst dann zulässig, wenn die (inkrementelle) Validität des Instrumentes sichergestellt ist.

## Reliabilität

In der vorliegenden Studie wurde die Messgenauigkeit des Situational Judgement Tests zunächst durch die Methode der internen Konsistenz bestimmt. Hierbei wird ein Test in so viele Teile zerlegt, wie er Items beinhaltet (Rost, 2004) – jedes Item wird folglich als „Testteil“ betrachtet. Die entsprechende Maßzahl zur Bestimmung der Reliabilität ist Cronbachs Alpha (Cronbach, 1951).

Die interne Konsistenz der elf Skalen des Tests gemäß Cronbachs Alpha liegt zwischen .273 und .805. Obwohl für Situational Judgement Tests meist keine hohen internen Konsistenzen erwartet werden (z. B. McDaniel, Hartman, Whetzel & Grubb, 2007). Trotzdem überschreiten in der vorliegenden Studie die Skalen 5, 9, 10 sowie 11 den Richtwert von .700 (Zinnbauer & Eberl, 2004) und können somit auch aus konservativer Perspektive als akzeptabel gelten. Die Skalen 2, 6 und 7 weisen ein Cronbachs Alpha zwischen .500 und .700 auf, für den Gesamttest ergibt sich ein Wert von .948. Da sich Cronbachs Alpha durch die Anzahl an einbezogenen Items auch bei niedrigen Interitemkorrelationen grundsätzlich erhöht (Cortina, 1993), darf dieser letzte Wert jedoch nicht überbewertet werden.

Nach dem „.700-Kriterium“ sollten die Skalen 1,2,3,4,6,7 und 8 überarbeitet werden, um Skalen zu erhalten, welche homogener den betreffenden Inhaltsbereich erfassen. Für eine Beibehaltung der beinhalteten Items trotz einer teilweise geringen internen Konsistenz könnte allerdings sprechen, dass die elf Skalen zu den von der KMK (ebd.) definierten Kompetenzbereichen heterogene Subkompetenzen beinhalten (Näheres dazu wurde in Abschnitt 1.4 dargelegt). Deutlich wird dies beispielsweise an Kompetenzbereich 7 „Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern“, der augenscheinlich verschiedenste Kompetenzen aus Diagnostik, Förderung und Beratung zusammenfasst. Dies beeinflusst Cronbachs Alpha, da dieses Maß abhängig ist von der Anzahl der Skala zugrundeliegender latenter Indikatoren (Cortina, 1993).

Eine Unkorreliertheit von erhobenen Kompetenzen, Verhaltensweisen oder Fähigkeiten zeichnet die Methode der Situational Judgement Tests in besonderem Maße aus (McDaniel, Hartman, Whetzel & Grubb, 2007, Schmitt & Chan, 2006), und somit werden hier eher geringe interne Konsistenzen erreicht. Die vorliegenden Ergebnisse können also als „typisch“ für einen Situational Judgement Test gelten.

Dies führt zu einem zentralen Argument für die Beibehaltung von Skalen mit niedrigen internen Konsistenzen: In jüngster Zeit hat die Unterscheidung zwischen reflektiven und formativen Messungen Aufmerksamkeit gefunden (Kuß & Eisend, 2010). Bei einer *reflektiven* Messung wird hierbei davon ausgegangen, dass manifeste Items ein latentes Konstrukt widerspiegeln. Die *formative* Messung basiert auf einer umgekehrten Annahme: „Bei diesem Ansatz wird angenommen, dass Veränderungen bei den Indikatoren ... Änderungen der Konstrukte verursachen. Mit anderen Worten formen oder bestimmen die Indikatoren das Konstrukt und dieses wird als (typischerweise lineare) Kombination seiner Indikatoren dargestellt...“ (Diamantopoulos, 2008, S. 1201).

Jarvis, MacKenzie und Podsakoff (2003) unterscheiden reflektive und formative Messungen anhand folgender Kriterien (Tabelle 9):

Tabelle 9. *Die Kriterien reflektiver und formativer Messungen nach Jarvis, MacKenzie und Podsakoff (2003)*

reflektive Messung	formative Messung
Kausalität: Konstrukt → Indikator	Kausalität: Indikator → Konstrukt
Indikatoren sind Manifestationen des Konstrukts	Indikatoren definieren Merkmale des Konstrukts
Veränderungen der Indikatoren verändern das Konstrukt nicht	Veränderungen der Indikatoren bedingen Veränderungen des Konstrukts
Indikatoren sind prinzipiell austauschbar	Indikatoren müssen nicht austauschbar sein
Verzicht auf einen Indikator verändert das Konstrukt nicht	Verzicht auf einen Indikator kann das Konstrukt maßgeblich verändern
Indikatoren müssen homogen sein	Indikatoren müssen nicht inhaltlich homogen sein

Nach diesen Kriterien kann die Formulierung von Standards für die Lehrerbildung: Bildungswissenschaften durch die KMK (ebd.) als Generierung eines formativen Konstrukts gelten, da eine Aufstellung von Einzelkompetenzen (z. B. die Strukturierung von Unterricht, eine individuelle Leistungsdiagnostik oder die Beratung von Eltern und SchülerInnen) erfolgte, über die sich professionelle pädagogische Kompetenz definieren soll. Unter dieser Perspektive kann dafür plädiert werden, dass eine inhaltliche Heterogenität ein konstituierendes Merkmal des Konstrukts – und daraus folgend des entsprechenden Erhebungsinstruments –

sein und die Entfernung von Items das Konstrukt ändern könnte. Eine Zusammenfassung von Items zu Skalen und die Berechnung interner Konsistenzen wäre bei einer Entscheidung für die Annahme eines formativen Konstrukts „professionelle pädagogische Kompetenz“ wenig sinnvoll – vielmehr müsste dann die Diagnostik studentischer Kompetenz itembezogen realisiert werden.

Cronbachs Alpha darf über diese Überlegung hinaus nicht als alleiniges Qualitätskriterium gelten, da ein hohes Cronbachs Alpha zum Einen für eine hohe Redundanz der Items spricht und zum Anderen nichts darüber aussagt, ob das zu messende Konstrukt tatsächlich durch das Instrument erfasst wird (Cortina, 1993).

Trotzdem muss beachtet werden, dass sich ein niedriges Cronbachs Alpha auch dann ergeben kann, wenn die Items mit hohen Messfehlern behaftet sind (Bollen & Lennox, 1991). Hier sind weitere Analysen des Instruments vonnöten.

### Validität

Im Folgenden werden die Ergebnisse der explorativen Datenanalyse zusammengefasst und insbesondere im Hinblick auf die Validität des entwickelten Instrumentes diskutiert. Die Validität ist das wichtigste Testgütekriterium (Bortz & Döring, 2003), da sie angibt, ob und wie gut ein Test in der Lage ist, genau das zu messen, was er zu messen vorgibt. Somit können die Ergebnisse der Validitätsanalysen Hinweise darüber liefern, ob sich das Instrument „Simulationsorientierte Erfassung professioneller pädagogischer Kompetenz Lehramtsstudierender“ überhaupt zur Erfassung professioneller pädagogischer Kompetenzen eignet.

#### *Inhaltsvalidität*

Die inhaltliche Validität eines Instruments bezieht sich auf das Ausmaß, in dem seine Items den interessierenden Merkmals- oder Verhaltensbereich erfassen. Eine statistische Maßzahl kann hier nicht ermittelt werden, stattdessen wird die Inhaltsvalidität durch ExpertInnen der Domäne eingeschätzt (Bortz & Döring, 2003).

In der vorliegenden Studie wurde die Inhaltsvalidität zum Einen sichergestellt, indem die Basis der generierten Items ein aufgrund theoretischer Erwägungen durch ExpertInnen entwickeltes Instrument (KOSTA, Weresch-Deperrois et al., 2009) darstellte. Zum Anderen wurde die Eignung der Items, das Vorliegen oder Nicht-



Vorliegen einer Kompetenz zu reflektieren, durch eine interdisziplinäre ExpertInnengruppe geratet.

Die Eignungsentscheidungen erfolgten durch sechs ExpertInnen, wobei sich hohe Übereinstimmungen bezüglich der Eignungseinschätzungen ergaben: Bei 15 Items zeigte sich dabei eine prozentuale Übereinstimmung der Urteile von 66,7%. 41 Items wurden mit einer Übereinstimmung von 83,3% geratet, bei 46 Items ergab sich eine Interrater-Übereinstimmung von 100%. Die Inhaltsvalidität des entwickelten Instruments wird somit als ausreichend eingeschätzt.

Eine weitere Verbesserung könnte durch die Durchführung der „Critical Incident Technique“ (Flanagan, 1954) erreicht werden. Diese Methode der Arbeits- und Anforderungsanalyse sammelt Verhaltensbeschreibungen zur Analyse von deren Nützlichkeit bei der Lösung praktischer Probleme bzw. mit dem Ziel, „psychologische Prinzipien zu entwickeln“ (Flanagan, 1954, S. 327; eigene Übersetzung). Hierbei werden ExpertInnen in Einzelinterviews zu besonders (in)effektiven Verhaltensweisen sowie deren Antecedentien und Konsequenzen befragt. Danach werden die genannten Ereignisse geclustert und beispielsweise bestimmten Kompetenzklassen zugeordnet. Ein solches Vorgehen könnte die inhaltliche Validität des Instrumentes weiter erhöhen.

#### *Kriteriale Validität*

Die Unterscheidung zwischen konvergenter und diskriminanter Validität geht auf die Multitrait-Multimethod-Methode von Campbell und Fiske (1959) zurück. Hierbei werden mehrere Konstrukte (Multitrait) durch mehrere Methoden (Multimethod) erhoben und die Validitäten anhand der Analyse von wechselseitigen Zusammenhängen zwischen Konstrukten und Methoden abgeschätzt. Konvergente Validität liegt dabei dann vor, wenn mehrere Methoden (z. B. verschiedene Tests oder Außenkriterien) *ein* Konstrukt übereinstimmend messen, diskriminante Validität zeigt sich, wenn *verschiedene* Konstrukte durch eine Methode zur Erfassung eines bestimmten Konstrukts differenziert werden (Bortz & Döring, 2003).

Das entwickelte Instrument sollte anhand konvergenter und diskriminanter Kriterien im Rahmen einer konkurrenten Validierung auf seine kriteriale Validität überprüft werden. Dabei wurden im Rahmen einer explorativen Datenanalyse die Abiturnote (diskriminant), die Anzahl im Rahmen des Studiums besuchter Praktika (konvergent), die Semesteranzahl (konvergent), Prüfungsnoten in

bildungswissenschaftlichen Fächern (konvergent) sowie das Berufsziel (Schulart) erhoben und eine multiple Regression gerechnet, um aufdecken zu können, ob die genannten Variablen auf die Testleistung in den 11 Skalenscores sowie im Gesamtscore Einfluss besitzen.

Die Ausprägungen der Variable „Berufsziel“ klärten dabei zwischen 51,3% (Score 5: „Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern.“) und 32,3% (Gesamtscore) der Varianz auf. Die Wahl des Berufsziels „Realschule“ beeinflusste hierbei die meisten Testscores (Score 2, 3, 5 – gemeinsam mit dem Berufsziel „Gymnasium“, 10, 11 sowie Gesamtscore), gefolgt vom Berufsziel „Gymnasium“ (Score 1 und 5 – gemeinsam mit dem Berufsziel „Realschule“) sowie „Grund- und Hauptschule“ (Score 7). Für die Scores 4, 6, 8 und 9 wurden keine Prädiktoren signifikant.

Ist das entwickelte Instrument geeignet, um zwischen Gruppen mit hoher bzw. niedriger Kompetenzausprägung zu diskriminieren? Dies sollte in einem nächsten Schritt überprüft werden: Da eine Analyse von Extremgruppen aufgrund weitgehend sehr gering besetzter Gruppen „Nicht-Kompetenter“ nicht möglich war, wurden die Skalenscores mithilfe eines Mediansplits dichotomisiert. Für die Variablen Abiturnote, Praxiserfahrung, Studiendauer, bildungswissenschaftliche Prüfungsnote sowie Berufsziel und die Testscores wurden folglich schrittweise binäre logistische Regressionen gerechnet, um aufdecken zu können, ob die Wahrscheinlichkeit des Eintreffens der Ereignisse „kompetent“ oder „nicht kompetent“ durch die genannten Variablen beeinflusst wird.

Modelle für die Scores 1, 5, 7 und 9 wurden signifikant. Insbesondere die Ausprägungen der Variablen „Schulart“ beeinflussen dabei die Zugehörigkeit zur Gruppe „Kompetent“ oder „Nicht-kompetent“. Die Modelle erklären zwischen 37,2% (Realschule/Gymnasium, Score 9) und 18,2% (Gymnasium, Score 5) der jeweiligen Varianz.

Hinweise darauf, dass sich systematisch eine bestimmte Schulart relevant für die Gruppenzugehörigkeit erweist, können allerdings nicht abgeleitet werden, da sowohl die Wahl des Berufsziels Förderschule, – zusammen mit der Abiturnote – Gymnasium, Grund-/Hauptschule als auch Realschule Einfluss auf die Wahrscheinlichkeit, ein Testergebnis über dem Median zu erreichen, besitzen.

Ein möglicher Erklärungsversuch für den Einfluss der Variable „Berufsziel (Schulart)“ könnte darin bestehen, dass die bildungswissenschaftlichen Anteile der Lehrerausbildung in Rheinland-Pfalz zwischen den Schularten nicht gleich groß sind (Gesetz- und Verordnungsblatt für das Land Rheinland-Pfalz: Landesverordnung über die Erste Staatsprüfung für Lehrämter nach Abschluss der Hochschulprüfungen lehramtsbezogener Bachelor- und Masterstudiengänge, 2007) und sich somit das bildungswissenschaftliche Wissen – als Basis professioneller pädagogischer Kompetenz – zwischen Studierenden verschiedener Lehrämter unterscheiden könnten. *Gegen* diese Annahme spricht, dass das zweite Kriterium bezüglich der bildungswissenschaftlichen Wissensbasis – die Prüfungsnote in bildungswissenschaftlichen Fächern – *keinen* prädiktiven Wert bezüglich der Testscores besitzt.

Die Anzahl im Rahmen des Studiums besuchter Praktika, die Semesteranzahl sowie die Prüfungsnoten in bildungswissenschaftlichen Fächern stellen der Datenanalyse zufolge keine Prädiktoren für die Ausprägungen der Testscores dar. Erwartet wurde jedoch eine positive Beziehung zwischen den genannten Variablen und den Testergebnissen, da sowohl die während der Praktika gesammelte Praxiserfahrung als auch das während des Studiums gewonnene bildungswissenschaftliche Wissen die Ausprägung der professionellen pädagogischen Kompetenz beeinflussen sollte.

Gelingen Validitätsnachweise nicht ausreichend, so ist zunächst zu fragen, ob die erhobenen *Kriterien* geeignet sind, d. h., ob sie im Falle der konvergenten Validierung dasselbe bzw. im Falle der diskriminanten Validierung etwas anderes „messen“ als der angewandte Test. Messen also die Anzahl im Rahmen des Studiums besuchter Praktika, die Semesteranzahl, Prüfungsnoten in bildungswissenschaftlichen Fächern sowie das Berufsziel (Schulart) *dasselbe* die professionelle pädagogische Kompetenz von Lehramtsstudierenden bzw. *etwas anderes* als pädagogische Kompetenz (Abiturnote)?

Bezüglich der über die Anzahl absolvierter Praktika erfassten *Praxiserfahrung* könnte behauptet werden, dass Kompetenzen zunehmen, je öfter sie ausgeübt werden. Unter diesem Gesichtspunkt könnte die genannte Frage mit „ja“ beantwortet werden. Allerdings ist die *Quantität* der Praktika kaum ein Indikator für die *Qualität* der geleisteten Arbeit. Somit könnte auf die Frage geantwortet werden: „Nein – das Kriterium misst etwas anderes als der Test.“

Auch bei der Entwicklung eines Situational Judgement Tests von Möller (2010) ergaben sich beispielsweise keine Korrelationen zwischen Testergebnis und Dauer der Berufserfahrung. Weekley und Ployhart (2005) deckten lediglich kleine Korrelationen zwischen Situational Judgement Test-Scores, allgemeiner Berufserfahrung ( $r=.21$ ) sowie der Anzahl der Jahre in der momentanen Position ( $r=.13$ ) auf. Auch Clevenger, Pereira, Wiechmann, Schmitt und Schmidt Harvey (2001) fanden keine signifikanten Korrelationen zwischen Situational Judgement Test-Scores und Berufserfahrung. Eine Erklärung hierfür könnten Hunter und Hunter (1984) liefern: Die Autoren fanden in ihrer Meta-Analyse lediglich eine kleine Korrelation von .18 zwischen Arbeitsleistung und Berufserfahrung. Aus diesen Ergebnissen könnte erneut gefolgert werden, dass Qualität (Arbeitsleistung) und Quantität (Berufserfahrung) der beruflichen Kompetenzausübung nicht hinlänglich deckungsgleich sind und somit eine Validierung von Situational Judgement Tests an Variablen der Berufserfahrung problematisch ist.

Geeigneter zur Erfassung der professionellen pädagogischen Kompetenz von Lehramtsstudierenden anhand eines Außenkriteriums könnten sich Fremdbewertungen der Kompetenzen durch praktikumsbetreuende LehrerInnen darstellen, jedoch werden die Praktikumsleistungen der Studierenden in Rheinland-Pfalz nicht bewertet oder benotet (Ministerium für Bildung, Wissenschaft, Jugend und Kultur Rheinland-Pfalz, 2010), weshalb eine entsprechende Erfassung der Kompetenzen von Lehramtsstudierenden nicht realisiert werden konnte.

Eine zweite Möglichkeit wäre es, die Selbsteinschätzung der Studierenden hinsichtlich ihrer professionellen pädagogischen Kompetenz zu erfassen. Dieses Vorgehen wird jedoch als wenig vielversprechend angesehen, da die Validität von self-assessment-Methoden im Bereich der Kompetenzeinschätzung aufgrund von möglichen Verfälschungen durch soziale Erwünschtheit oder mangelnde selbstreflexive Fähigkeiten als begrenzt eingeschätzt wird. Selbsteinschätzungsverfahren erfassen außerdem eher das Selbstkonzept eigener Fähigkeiten denn die tatsächliche Kompetenz (z. B. Hartig & Jude, 2007, McClelland, 1987). Frese, Fay, Hilburger und Leng (1997) argumentieren zudem, dass traditionelle self-assessment-Verfahren zwar die Wichtigkeit erfassen, die Personen einem Konstrukt zuschreiben, nicht jedoch das Ausmaß tatsächlicher Kompetenz.

Als zweites Außenkriterium zum Nachweis konvergenter Validität wurde die *Prüfungsnote in bildungswissenschaftlichen Fächern* erfasst.

Zunächst ist erneut der Frage nachzugehen: „Messen Prüfungsnoten in bildungswissenschaftlichen oder didaktischen Fächern die professionelle pädagogische Kompetenz von Lehramtsstudierenden?“ Prüfungsnoten erfassen das Wissen von Studierenden. Für eine positive Antwort spricht dabei, dass Wissen im Allgemeinen als Voraussetzung kompetenten Handelns gilt (z. B. Klieme & Hartig, 2007). Trotzdem kann die Frage auch mit „nein“ beantwortet werden, da eine Schere zwischen Theorie und Praxis in der Lehrerbildung häufig beklagt wird: Blomberg (2011) benennt die „Kluft zwischen Wissen und Handeln“ (S.5) als größtes Problem der Lehrerbildung, und auch Helmke, Helmke und Schrader (2007) sowie Schocker-von Ditfurth (2001) erwähnen die Diskrepanz zwischen dem Wissenschaftswissen, welches Lehramtsstudierenden an Universitäten und Hochschulen vermittelt wird, und dem praktischen Handlungswissen. Kolbe (2004) postuliert schließlich, dass das wissenschaftliche Wissen nicht durch Transfer in der Praxis angewandt werde.

In der Metaanalyse von McDaniel und Nguyen (2001) wurden 17 Studien bezüglich der Korrelation zwischen beruflichem Wissen und Situational Judgement Test-Scores untersucht. Die Autoren errechneten dabei eine mittlere Korrelation von  $r=.07$ . Auch in der Studie von Lievens, Buyse und Sackett (2005) besaßen Ergebnisse des entwickelten Situational Judgement Tests keinen prädiktiven Wert bezüglich von Studiennoten.

Die Eignung von bildungswissenschaftlichen Prüfungsnoten als Kriterium für durch einen Situational Judgement Test erfasste pädagogische Kompetenz kann somit kritisch gesehen werden. Um diesbezüglich eine gesicherte Aussage treffen zu können, sind weitere Untersuchungen nötig.

Als weiteres konkurrentes Außenkriterium wurde zur Überprüfung der konvergenten Validität des Instruments die *Anzahl der besuchten Semester im Lehramtsstudium* erfasst.

Wieder ist der Frage nachzugehen: „Misst die Anzahl besuchter Studiensemester die professionelle pädagogische Kompetenz von Lehramtsstudierenden?“ Studierende höheren Semesters sollten sowohl über eine breitere praktische

Erfahrung als auch über eine größere Wissensbasis verfügen als Kommilitonen, die kürzer studieren. Wird also Wissen als Voraussetzung kompetenten Handelns verstanden oder gar mit Kompetenz gleichgesetzt (z. B. Baumert & Kunter, 2006; Klieme & Hartig, 2007), und folgt man der Annahme, dass eine häufige Verhaltensaübung zu kompetenterem Handeln führt, so kann die Frage mit „ja“ beantwortet werden. Wird allerdings davon ausgegangen, dass universitäres Wissenschaftswissen nicht adäquat in pädagogisches Handeln transferiert wird (wie es beispielsweise Kolbe, 2004, tut; auch: Helmke, Helmke & Schrader, 2007, Schocker-von Ditfurth, 2001), so ist die Anzahl besuchter Semester kein geeignetes Außenkriterium. Beide Perspektiven sind plausibel, und so sind auch an dieser Stelle weitere Analysen abzuwarten.

Ein Außenkriterium zum Nachweis *diskriminanter* Validität war die Abiturnote. Die Eignung des Kriteriums scheint gegeben, da in den durch die Abiturnote erfassten Schulfächern keine pädagogischen Inhalte thematisiert werden. Zudem wurden die Situationsbeschreibungen wenig komplex sowie kurz gestaltet, um den Einfluss des kognitiven „Load“ auf die Testergebnisse möglichst gering zu halten. Erwartungsgemäß besaß die Abiturnote keinen signifikanten Einfluss auf die Testscores.

Zum Nachweis der *diskriminanten* Validität wurde weiterhin überprüft, ob Studierende des Lehramts hinsichtlich ihrer Testscores bessere Werte erreichen als Studierende technischer oder naturwissenschaftlicher Studiengänge.

Für alle Scores konnten hochsignifikante und praktisch bedeutsame Mittelwertsunterschiede nachgewiesen werden. Lehramtsstudierende schneiden also in einem Test zur simulationsorientierten Erfassung professioneller pädagogischer Kompetenzen besser ab als Studierende technischer oder naturwissenschaftlicher Studiengänge.

Dieses Ergebnis spricht für die Validität des überprüften Instruments, da erwartungsgemäß Personen mit einer pädagogischen Ausbildung höhere Testscores erreichen als Personen ohne eine solche Ausbildung und es somit naheliegt anzunehmen, dass der Test tatsächlich etwas misst, das mit pädagogischer Kompetenz in Beziehung steht.

Für das vorliegende Instrument konnte der Nachweis der Validität mit Ausnahme des Kriteriums „Studienfach“ noch nicht zufriedenstellend erbracht werden.

Die Ursachen mangelnder kriterialer Validität können a) in einer mangelnden Eignung des Instruments oder b) in einer mangelnden Eignung der Kriterien begründet sein (Bühner, 2006), wobei es sich häufig schwierig gestaltet nachzuweisen, ob *Test* oder *Kriterium* ungeeignet sind.

Die gewählten Kriterien wurden bereits an entsprechender Stelle jeweils diskutiert. Ein weiterer *allgemeiner* Grund mangelnder Validität besteht in der Streuungsrestriktion sowohl a) der Items als auch b) der Stichprobe (Bühner, 2006; Scheurer, 1999).

Zu a): Die Itementwicklung wurde dergestalt realisiert, dass fachspezifische Inhalte ausgeklammert wurden, da am Ende der Entwicklung eines Instrumentes der Erfassung und Bewertung professioneller Kompetenz im Lehrberuf eine „Beurteilung des Entwicklungsstandes der beruflichen Kompetenzen insgesamt ..., nicht eine Beurteilung spezieller Teilkompetenzen für spezifische Gruppen von Lehrern“ stehen soll (Terhart, 2007, S. 48; siehe auch Oser & Oelkers, 2001). Zudem stellen fachübergreifende Lerngelegenheiten einen wichtigen Anteil der Lehrerbildung dar, wobei die Erziehungswissenschaft den verbindlichen Kern darstellt (Tippelt, 2008; Schmidt et al., 2007). Die Erfassung professioneller pädagogischer Kompetenz entlang von durch die KMK (ebd.) formulierten Standards sollte außerdem für die Lehramtsstudierenden sämtlicher Fächer und Schulformen gleichermaßen ermöglicht werden und durch eine kurze, wenig komplexe Situationsschilderung sowohl der Einfluss der Lesefähigkeit als auch der kognitiven Fähigkeit so gering wie möglich gehalten werden. Zusätzlich musste die Eindimensionalität der Items sichergestellt werden. Aus diesen Überlegungen resultierten wenig komplexe, relativ kontextarme Items.

Der Kontextgehalt dieser Items könnte jedoch trotzdem zu gering sein. Terhart (2009) schreibt beispielsweise: „National wie international wird im Rahmen der Forschung zur Lehrerbildung, speziell: zur Wirksamkeit von Lehrerbildung die besondere Rolle der fachdidaktischen Ausbildungselemente, des ‚pedagogical content knowledge‘ deutlich.“ (ebd., S. 434) Hieraus könnte geschlossen werden, dass eine valide Erfassung professioneller pädagogischer Kompetenzen

nur auf der Folie fachdidaktischer – und somit kontexthaltigerer – Inhalte gelingen könnte.

Da außerdem die mittlere Itemschwierigkeit des Situational Judgement Tests bei .673 liegt, ist der Test eher zu leicht, wodurch die Unterscheidung zwischen ProbandInnen mit niedriger Merkmalsausprägung und ProbandInnen mit hoher Merkmalsausprägung nur unzureichend gelingen kann.

Zu b): Eine nicht realisierbare Zufallsauswahl ist für die Lehrerbildungsforschung typisch (Cramer, 2010, in Druck), und auch in der vorliegenden Untersuchung liegt eine Streuungsrestriktion bezüglich der Variablen „Hochschule“, „Studiensemester“, „Anzahl abgeleiteter Praktika“ sowie „Berufsziel“ vor. Zudem könnte sich aufgrund der Freiwilligkeit der Teilnahme an der Studie eine Auswahl derjenigen Studierenden ergeben haben, die insbesondere engagiert und interessiert sind – dies könnte eine Ursache für die eher geringe mittlere Schwierigkeit des entwickelten Situational Judgement Tests darstellen.

Die Ergebnisse der Datenanalyse sind folglich aufgrund der geringen Streuung von Testergebnissen und Kriterien vorsichtig zu interpretieren.

### *Konstruktvalidität*

Beim Nachweis der Konstruktvalidität von Situational Judgement Tests ergeben sich Probleme, welche aus der besonderen Methodik dieser situationsbezogenen Verfahren resultieren:

Die gewöhnlich zur Überprüfung der Konstruktvalidität angewandten Faktorenanalysen sind für Situational Judgement Tests ungeeignet, da zum Einen meist jedes Item eine distinkte Kompetenz reflektiert und sich somit eine hohe Heterogenität der Tests ergibt, und zum Anderen, weil der Einfluss der betreffenden Situation auf jedes Item differentiell wirkt (z. B. McDaniel & Whetzel, 2005, Schmitt & Chan, 2006). Hierbei könnte *ein Item mehrere Konstrukte* repräsentieren oder aber *verschiedene Items* könnten *verschiedene Konstrukte* spiegeln, wodurch sich ein Einfluss sowohl auf die Konstrukt- als auch auf die kriteriale Validität ergeben könnte (McDaniel, Whetzel, Hartman, Nguyen und Grubb, 2006). Bledow und Frese (2009) postulieren sogar, dass die Items eines Situational Judgement Tests keine parallelen Messungen eines Konstruktes seien, sondern dass vielmehr jedes situationsbasierte Item ein unabhängiges Phänomen messe. Die Autoren



schlußfolgern: „This implies that situational aspects of each scenario influence responses, and therefore, ... items are not parallel measures of a unidimensional construct. Each scenario item ... measures a unique phenomenon.“ (S. 238) Situational Judgement Tests erfassen demnach Kompetenzen meist nicht konstruktrein (Christian, Edwards & Bradley, 2010).

In ihrer Metaanalyse gelangen McDaniel und Whetzel (2005) ebenso wie Schmitt und Chan (2006) zu dem Ergebnis, dass keine veröffentlichte Faktorenanalyse eines Situational Judgement Tests mit interpretierbaren Faktoren vorliege und somit Situational Judgement Tests zu keinen eindeutigen Faktorenlösungen gelangen. Christian, Edwards und Bradley (2010) behaupten dementsprechend, dass Situational Judgement Tests Eigenschaften oder Kompetenzen meist nicht konstruktrein abbilden. Auch Weekley und Jones (1999) sowie O'Connell, Hartman, McDaniel, Grubb und Lawrence (2007) gelangen zu dem Schluss, dass die Items von Situational Judgement Tests stets mehrere Konstrukte zugleich erfassen. Dies liegt Kanning, Grewe, Hollenberg und Hadouch (2006) zufolge in der Zielsetzung begründet, Alltagssituationen abbilden zu wollen, welche sich heterogen und komplex gestalten. Lievens und Sackett (2007) fassen dementsprechend zusammen, dass Situational Judgement Tests nicht entwickelt würden, um klar definierte Konstrukte zu messen, sondern zur Erfassung domänenspezifischer Schlüsselkompetenzen. Situational Judgement Tests können als komplexe Problemlöseszenarien interpretiert werden, deren ungenügende Konstruktvalidität beispielsweise von Kersting (1999) bemängelt wird. Nach Achtenhagen und Baethge (2007) ist es zusätzlich – unabhängig von der Erhebungsmethodik – „...so gut wie unmöglich, ein Schema festzulegen, mit dessen Hilfe sich berufliche Kompetenzen beschreiben und messen lassen, ohne dass es zu schwerwiegenden Überlappungen der Kategorien käme.“ (S. 57) Nach Burmeister (2009) ist jedoch bei simulationsorientierten Verfahren nicht die Konstrukt-, sondern die Kontextvalidität das Ziel.

Eine Faktorenanalyse wurde aufgrund der mangelnden Eignung des Verfahrens für die Analyse der Konstruktvalidität von Situational Judgement Tests folglich nicht durchgeführt. Die Methode der Diskriminanzanalyse wird als besser geeignet empfohlen (Stemler & Sternberg, 2006), und so wurden die angefallenen Daten dahingehend analysiert, ob vermutlich mit dem Testergebnis in Beziehung stehende (konvergente oder diskriminante) Kriterien geeignet sind, vorab definierte Gruppen

von TestteilnehmerInnen mit hohen bzw. niedrigen Testscores bestmöglich zu trennen.

Im Ergebnis gelang die Trennung in eine Gruppe, deren Gesamtscore *über* dem Median der Gesamtstichprobe liegt, bzw. eine zweite Gruppe, deren Gesamtscore *unter* dem Median der Gesamtstichprobe lag aufgrund der Variablen „Semester“, „Prüfungsnote“, „Abiturnote“ sowie „Anzahl abgeleiteter Praktika“ nicht. Da bereits im Rahmen der Regressionsanalyse keine Beziehungen zwischen den Variablen und den Testscores nachgewiesen wurden, war dieses Ergebnis allerdings vorhersehbar. Von der Variable „Abiturnote“ zudem angenommen, dass sie *keine* Einflussgröße bezüglich der Testergebnisse darstellen sollte. Somit ist eine nicht nachgewiesene Trennungsleistung dieser Variable zwischen den Gruppen „Gesamtscore < Median“ und „Gesamtscore > Median“ erwartungskonform.

Zur Überprüfung der Konstruktvalidität wurde abschließend eine Zuordnung der entwickelten Items zu den KMK-Standards durch ExpertInnen vorgenommen, um zu untersuchen, ob sich die generierten Items den durch die KMK (ebd.) postulierten elf Kompetenzen inhaltlich zuordnen lassen.

84 von 102 Items wurden dabei mehrheitlich übereinstimmend *einer* KMK-Kompetenz zugeordnet – hiervon wurden 48 Items mehrheitlich der bei der Entwicklung *intendierten* Kompetenz zugeordnet. 18 Items wurden von *keinem/* Experten/in der *intendierten* Kompetenz zugeordnet. Nach Seidel, Blomberg und Stürmer (2010) ist eine Heterogenität von ExpertInnenurteilen bei der Zuordnung von Unterrichtssituationen zu Kriterien erfolgreichen Unterrichts der Schwierigkeit geschuldet, komplexe Unterrichtssituationen durch ein standardisiertes Instrument darzustellen (Seidel, Blomberg & Stürmer, 2010).

Der am häufigsten genannte Kompetenzbereich 2 bezeichnet die Gestaltung von Lernsituationen, die Motivation der SchülerInnen bzw. Transfer und Verknüpfung von Gelerntem und bezieht sich somit auf das Unterrichten als Kernaufgabe von Lehrenden. Zudem lässt sich hierunter eine große Anzahl von Lehrmethoden und Verhaltensweisen subsumieren. Eine häufige Nennung dieses Kompetenzclusters überrascht somit nicht.

Am seltensten wurden Kompetenzbereich 4 („Lehrerinnen und Lehrer kennen die sozialen und kulturellen Lebensbedingungen von Schülerinnen und Schülern und

nehmen im Rahmen der Schule Einfluss auf deren individuelle Entwicklung.“) sowie 11 („Lehrerinnen und Lehrer beteiligen sich an der Planung und Umsetzung schulischer Projekte und Vorhaben.“) gewählt. Die vergleichsweise wenig häufige Nennung des letztgenannten Bereiches könnte damit erklärt werden, dass dieser Kompetenzcluster sehr eng gefasst ist und sich in wenigen, klar umschriebenen Handlungen äußert. Warum Kompetenzbereich 4 zu den am wenigsten genannten Bereichen gehört, ist unklarer zu deuten. Möglicherweise liegt die Ursache darin, dass der Kompetenzbereich zwei nicht vollständig korrelierte Unterfacetten beinhaltet und das gleichzeitige Vorliegen sowohl des Wissens um schülerseitige Lebensbedingungen als auch der individuellen Entwicklungsförderung innerhalb der entwickelten Items von den ExpertInnen nicht aufgefunden werden konnte.

### Verfälschbarkeit und Testübung

Zu den Aspekten der Verfälschbarkeit und Testübung können zum jetzigen Zeitpunkt lediglich allgemeine Aussagen getroffen werden:

Die Antwortinstruktion des Situational Judgement Tests wurden wissensbasiert gestaltet, da solcherlei Antwortinstruktionen als weniger leicht verfälschbar gelten als verhaltensbasierte Instruktionen (Hooper, Cullen & Sackett, 2006). Die Verfälschbarkeit von Situational Judgement Tests wird außerdem allgemein geringer eingeschätzt als die Verfälschbarkeit bei Messungen von Persönlichkeitsdimensionen (ebd.).

Wie anfällig das im Rahmen der vorliegenden Arbeit entwickelte Instrument für Übungeffekte ist, kann noch nicht beantwortet werden. Generell wird der Einfluss von Testübung auf die Ergebnisse von Situational Judgement Tests nicht höher bewertet als beispielsweise bei Intelligenztests (Lievens, Buyse & Sackett, 2005). Zur Klärung dieser Frage müssten weitere Erhebungen der befragten Stichprobe dienen.

### Nützlichkeit

Aufgrund des beklagten Mangels an adäquaten Instrumenten zur realitätsnahen Erfassung der professionellen pädagogischen Kompetenz angehender Lehrpersonen jenseits von Selbsteinschätzungs- und wissensbasierten Verfahren kann eine simulationsorientierte Erfassung professioneller pädagogischer Kompetenz als *grundsätzlich* nützlich eingestuft werden.

Ein weiteres Argument für die *grundsätzliche Nützlichkeit* eines Situational Judgement Tests ist dessen gemäßigte Augenscheinvalidität: Solcherlei Instrumente werden häufig von der relevanten Stichprobe gut akzeptiert (z. B. Lievens, Peeters & Schollaert, 2008). Die Drop-Out-Quote könnte somit durch einen solchen Test verringert werden.

Angesichts uneinheitlicher Validitätsnachweise durch einerseits Studiennoten sowie die Anzahl abgeleiteter Praktika bzw. andererseits das Studienfach können zum jetzigen Zeitpunkt kaum Aussagen zur diagnostischen oder prognostischen inkrementellen Validität des Verfahrens gemacht werden. Ob das vorliegende Verfahren inkrementelle Validität aufweist, müssten somit zukünftige Studien zeigen.

### Ökonomie

Generell zeigen sich paper-pencil-Instrumente ökonomischer als Beobachtungsverfahren: Sowohl die Durchführung als auch die Datenauswertung ist weitaus weniger kosten- und zeitaufwendig als etwa eine Videographie gehaltener Unterrichtsstunden. Ein simulationsorientiertes Verfahren in Form eines Situational Judgement Tests kann somit *grundsätzlich* als ökonomisch beurteilt werden.

Nichtsdestotrotz stellt die Ökonomie eines Verfahrens lediglich ein Nebengütekriterium dar. Ist ein Test nicht objektiv, valide und reliabel, so ist auch seine Ökonomie nicht von Relevanz.

### Normierung

Der letzte Schritt der Entwicklung eines Testverfahrens besteht in der Testnormierung (Scheurer, 1999). Ziel ist es hierbei letztendlich, individuelle Testergebnisse anhand statistischer Bezugswerte zu interpretieren.

Zunächst muss hierfür die Rohwertverteilung des betreffenden Instruments innerhalb einer repräsentativen Referenzpopulation ermittelt werden. Anschließend werden Skalenursprung und –einheit in Abhängigkeit von der Verteilung der Rohwerte festgelegt, wobei sich als Bezugswert für das individuelle Testergebnis der Mittelwert anbietet. Individuelle Testergebnisse lassen sich durch die Differenzbildung des individuellen Messwerts mit diesem Bezugswert als unter- oder

überdurchschnittlich deuten. Die Differenz erfolgt hierbei abhängig von der Rohwerteverteilung meist anhand der Standardabweichung (Scheurer, 1999). Um eine Vergleichbarkeit der Ergebnisse verschiedener Tests zu ermöglichen, wird zuletzt eine Standardisierung der Messwerte realisiert.

Die Voraussetzung einer Testnormierung ist selbstverständlich der Nachweis der Validität eines Verfahrens. Zudem ist eine Normierung nur dann angezeigt, wenn eine normorientierte Diagnostik überhaupt erwünscht ist. Im Falle einer Kriterienorientierung gerät die Verteilung der Testwerte innerhalb der Referenzpopulation nur am Rande (bei der Wahl angemessener, realitätsnaher Kriterien) in den Blick.

Für das vorliegende Instrument kann zum jetzigen Zeitpunkt kein abschließender Nachweis der Validität erbracht werden. Auch eine Entscheidung für eine norm- oder kriterienorientierte Interpretation der Testwerte kann noch nicht erfolgen – allerdings legt die Bereitstellung von zu erreichenden Standards durch die KMK (2004) ein kriterienorientiertes Vorgehen nahe.

## **6. Zusammenfassung und Ausblick**

Die Neustrukturierung (nicht nur) des Lehramtsstudiums im Rahmen der Bologna-Reform ging einher mit der Einführung der Abschlüsse „Bachelor“ und „Master“, der Einführung eines vergleichbaren Leistungspunktesystems (European Credit Transfer System - ECTS), einer europäisch verankerten Qualitätssicherung sowie einer stärkeren Praxisorientierung und Orientierung am Konzept des „Lebenslangen Lernens“ (Konferenz der für die Hochschulen zuständigen europäischen Ministerinnen und Minister, 2005). Die Ziele der Lehrerbildungsreform bestehen dabei insbesondere in einer stärkeren Praxisorientierung des Lehramtstudiums und einer optimierten Verknüpfung des Studiums mit den obligatorischen Praxisphasen.

Entscheidend für diese Neustrukturierung der Lehrerbildung ist ein Wandel von einer Input-Steuerung durch Gesetze, Lehrpläne, Erlasse, Stundentafeln, Ordnungen hin zur Output-Orientierung an mit Standards zu vergleichenden Effekten und Wirkungen (Terhart, 2002). Normative Standards wurden für Fach- und Bildungswissenschaften durch die KMK (2004) formuliert, wobei eine Output-Orientierung letztendlich die Frage klären muss: „Ist die Lehrerbildung wirksam im Sinne einer Standarderreicherung der Studierenden?“ Hierfür muss der Erfüllungsgrad

der Standarderreichung kontinuierlich evaluiert werden, um sowohl für die Studierenden als auch für die Studiengangsentwicklung Optimierungsbedarfe aufdecken zu können.

Zur Evaluation der Lehrerbildung müssen adäquate Instrumente vorliegen, um das (Nicht-)Erreichen von Kompetenzstandards reliabel und valide erfassen zu können. Zwar lassen sich in diesem Bereich positive Entwicklungen verzeichnen, doch wird der Mangel an empirischen Studien zur Wirksamkeit der Lehrerbildung häufig beklagt (König, 2010). Liegen entsprechende Instrumente zwar vor, so erheben sie jedoch Kompetenzen meist durch Selbsteinschätzungsverfahren – ein Vorgehen, das bezüglich seiner Validität kritisch zu betrachten ist (s.o.; z. B. auch Tousignant & DesMarchais, 2002).

Ziel der vorliegenden Arbeit war es folglich, entlang der durch die KMK formulierten Standards für die Lehrerbildung: Bildungswissenschaften (2004) ein valides Instrument zur Erfassung der professionellen pädagogischen Kompetenz Lehramtsstudierender entwickeln. Hierbei sollte der Forderung von Schaper, Hilligus und Reinhold (2009) einer anforderungsbezogenen Messung mithilfe eines simulationsorientierten eignungsdiagnostischen Ansatzes nachgekommen werden.

Als geeigneter Kompromiss zwischen Ökonomie und Validität erschien die Methode der Situational Judgement Tests. Diese simulationsorientierten Tests ermöglichen eine indirekte, situationsspezifische Erfassung multipler Konstrukte (Bledow & Frese, 2009). Hierbei werden den TestteilnehmerInnen meist in schriftlicher und standardisierter Form fiktive Situationen vorgegeben, die im entsprechenden beruflichen Kontext gewöhnlich auftreten können. Critical Incidents, Arbeits- und Anforderungsanalysen oder vorgegebene Kriterien (beispielsweise die Standards für die Lehrerbildung: Bildungswissenschaften der KMK, 2004) liefern bei der Entwicklung eines solchen Situational Judgement Tests die Inhalte der Items. Die vorgegebenen Situationsvignetten (Items) werden durch die ProbandInnen bewertet, indem diese beispielsweise die in einer Situation ihrer Meinung nach erfolg- oder mißerfolgversprechendste Handlungsalternative wählen oder hinsichtlich ihrer Effektivität einschätzen. Durch ein solches Instrument würde eine standardorientierte Evaluation zweiter Stufe nach Terhart (2002) realisiert (siehe Abschnitt 1.6).

Vor der Entwicklung eines solchen Situational Judgement Tests wurde der Stand der Forschung zur professionellen pädagogischen Kompetenz rekapituliert.

Zunächst wurde der Begriff der „Kompetenz“ geklärt und der Definition von Klieme und Hartig (2007) gefolgt, die Kompetenzen als erlernbare kontextspezifische Leistungsdispositionen verstehen, welche sich auf Situationen und Anforderungen in bestimmten Domänen beziehen. Grundlegend sind hierbei der Kontextbezug und die Erlernbarkeit von Kompetenzen.

Anschließend wurde dieser Kompetenzbegriff auf das Feld des Lehrberufs übertragen, wobei konstituierende Merkmale professionellen pädagogischen Handelns identifiziert und auf Passung zu den Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004) überprüft wurden. Hierbei zeigte sich eine weitgehende Übereinstimmung zwischen den Ergebnissen der empirischen Forschung, normativen Modellen sowie den Standards für die Lehrerbildung: Bildungswissenschaften.

Die Entwicklung des Situational Judgement Test sollte auf der Folie der Standards der KMK (ebd.) geschehen, um das Instrument in Falle eines positiven Validitätsnachweises für die standardorientierte Evaluation von „Lehrerbildung“ nutzen zu können.

Der *erste Schritt* der Operationalisierung professioneller pädagogischer Kompetenz von Lehramtsstudierenden entlang der KMK-Standards bestand in der Auswahl geeigneter trennscharfer Items aus dem Evaluationsinstrument der Lehrerbildung an der Universität Koblenz-Landau, Campus Landau (KOSTA; Weresch-Deperrois et al., 2009).

In dem zu entwickelnden Instrument sollte durch die Situationsvignetten ein breites Spektrum an Kompetenzausprägungen reflektiert werden. Um zu realistischen Handlungsalternativen am *unteren* Ende des Kompetenzspektrums zu gelangen, wurden somit in einem *zweiten Schritt* Lehramtsstudierende der ersten Semester als „NovizInnen professioneller pädagogischer Kompetenz“ qualitativ zur Gestaltung von Situationen des Lehramts befragt sowie die Antworten hinsichtlich des Grades ihrer Standardausrichtung von unabhängigen RaterInnen beurteilt. Die „nicht standardkonformen“ Antworten der Studierenden wurden zur Generierung von Items negativer Polung genutzt.

Der *dritte Schritt* der Operationalisierung bestand in der Entwicklung von Items mit Situationsbeschreibungen vom oberen Ende (positive Polung) als auch vom unteren

Ende (negative Polung) des entsprechenden Kompetenzspektrums – mithin der Entwicklung von Situationsvignetten sowie des Antwortformats. Die Entscheidung fiel hierbei für eine relativ kontextarme, eindimensionale und somit fachübergreifende Darstellung von Situationen, um so zum Einen die Varianz aufgrund von Störvariablen wie etwa dem kognitiven „Load“ und der Lesefähigkeit zu minimieren und zum Anderen eine Nutzung des Verfahrens durch Lehramtsstudierende sämtlicher Fach- und Schulrichtungen realisieren zu können. Für jede Situation wurden zwei Items entwickelt, wobei eines ein Ideal vorgibt und das zweite ein nicht-standardkonformes Verhalten schildert. Hierdurch sollten Effekte der sozialen Erwünschtheit minimiert sowie die Schwierigkeit und Validität des Tests erhöht werden (z. B. Ployhart & Ehrhart, 2003; Hooper, Cullen & Sackett, 2006). Beim Antwortformat wurde zugunsten einer sechsstufigen Likert-Skala von einem Multiple-Choice-Format Abstand genommen, um die Varianz der Ergebnisse zu erhöhen (Cucina, Vasilopoulos & Leaman, 2003).

In einem *vierten Schritt* wurde die Eignung der Situationensvignetten, das Vorliegen oder Nicht-Vorliegen von Kompetenzen zu repräsentieren, von einer ExpertInnengruppe aus sechs ExpertInnen der Bereiche Lehramt, Erziehungswissenschaft und Pädagogik eingeschätzt. Als Kriterium der übereinstimmenden Einschätzung wurden Prozentwerte berechnet. Die Interrater-Übereinstimmung bezüglich der Einschätzung der Items als *geeignet* lag bei 66,7% (15 Items), 83,3% (41 Items) sowie 100% (46 Items). Somit können sämtliche Items als geeignet bewertet werden, positiv sowie negativ gepolte Kompetenzausprägungen zu markieren.

Es folgte der erste *Einsatz des Verfahrens* zur Itemanalyse und Überprüfung von Reliabilität und Validität an einer Stichprobe (n = 332) aus Lehramtsstudierenden sowie Studierenden technischer und naturwissenschaftlicher Studiengänge.

Bezüglich der Variablen „Studienort“, „Berufsziels (Schulart)“, „Anzahl der Studiensemester“ sowie folglich „Anzahl abgeleiteter Praktika“ zeigte sich eine eingeschränkte Streuung. Eine solche eingeschränkte Zufallsauswahl ist zwar nicht selten in der Lehrerbildungsforschung zu finden (Cramer, 2010, in Druck), trotzdem ergeben sich aufgrund der eingeschränkten Streuung hinsichtlich der zur Validierung genutzten Kriterien sowohl Repräsentativitäts- als auch Interpretationsprobleme. Die folgenden Ergebnisse sind auf der Folie dieser Datenlage zu interpretieren.



Die Itemanalyse ergab überwiegend eine Normalverteilung der Items. Die durchschnittliche Itemschwierigkeit lag bei .673. Diese relativ hohe „Leichtigkeit“ des Instrumentes könnte die Validität des Verfahrens beeinträchtigen, da zwischen ProbandInnen mit hoher und niedriger Kompetenzausprägung nicht hinreichend diskriminiert werden könnte.

Die Skalen 1, 3 und 4 wiesen ein Cronbachs Alpha zwischen .273 und .435 auf. Skala 2, 6 und 7 erreichten interne Konsistenzen zwischen .500 und .700, und Skala 5, 9, 10 sowie 11 überschritten den Richtwert von .700. Die *Reliabilitätsanalyse* anhand von Cronbachs Alpha kann somit in Anbetracht der bei Situational Judgement Tests zu erwartenden niedrigen Werte (z. B. McDaniel, Hartman, Whetzel & Grubb, 2007) und gerade auch unter der Annahme eines formativen Konstrukts als zufriedenstellend gelten.

Das letztere Argument kann auch bezüglich der Trennschärfe der Items gelten: Die skalenbezogenen Trennschärfen der Items lagen im Bereich von -.402 und .629. Im Mittel ergab sich eine skalenbezogene Trennschärfe von .305. 50 der 102 Items zeigten sich nach dem häufig verwandten Kriterium einer „300-Grenze“ (Moosbrugger & Kelava, 2007) als nicht ausreichend trennscharf. Allerdings sind geringe Trennschärfen von Einzelitems im Sinne von Item-Skala-Korrelationen bei Situational Judgement Tests nicht selten (siehe hierzu die Metaanalyse von McDaniel, Hartman, Whetzel & Grubb, 2007). Zudem könnte die standardorientierte professionelle pädagogische Kompetenz Lehramtsstudierender als ein formatives Konstrukt gelten, bei welchem der Verzicht auf einen Indikator das Konstrukt maßgeblich verändern würde (Diamantopoulos, 2008). Eine Entscheidung für die Beibehaltung der fraglichen Items könnte folglich unter dem Gesichtspunkt der Inhaltsvalidität fallen. Die fünf Items mit negativen Trennschärfen sollten allerdings vom Test ausgeschlossen werden, da negative Item-Skala-Korrelationen sich dann zeigen, wenn weniger kompetente TestteilnehmerInnen bei dem betreffenden Item entgegen der Erwartung besser abschneiden als kompetentere ProbandInnen.

Nach Analyse der entwickelten Items wurde die Beziehung zwischen verschiedenen Außenkriterien und den Testscores in den Blick genommen, um erste Informationen bezüglich der Validität des Instrumentes zu erhalten. Multiple Regressionsanalysen im Hinblick auf die Beziehung zwischen den Variablen „Abiturnote“, „Anzahl im Rahmen des Studiums besuchter Praktika“, „Semesteranzahl“, „Prüfungsnoten in

bildungswissenschaftlichen Fächern“ sowie „Berufsziel (Schulart)“ einerseits sowie den Skalenscores andererseits zeigten lediglich einen Einfluss des gewählten Berufsziels (Grund- und Hauptschule, Förderschule, Realschule sowie Gymnasium) auf die Testscores 1, 2, 3, 5, 7, 10, 11 sowie den Gesamtscore. Hierbei war insbesondere die Schulart „Realschule“ ein Prädiktor. Werden die Testscores am Median dichotomisiert, um zwei Gruppen „kompetenter“ und „nicht-kompetenter“ Studierender zu erhalten, so zeigt sich ebenso das Berufsziel (Schulart) als einflussreich bezüglich der Scores 1, 5, 7 und 9, wobei sämtliche Schularten als Prädiktoren fungieren. Auch die Abiturnote erweist sich als Einflussgröße bezüglich Testscore 1. Erwartet wurde jedoch, dass die Variablen „Anzahl im Rahmen des Studiums besuchter Praktika“ sowie „Semesteranzahl“ die Testscores in positiver Richtung beeinflussen, da sowohl während des Studiums gewonnenes Wissen als auch Praxiserfahrung zu einem Kompetenzgewinn führen sollten. Ein Hinweis auf die Validität des Verfahrens anhand dieser Variablen konnte somit nicht erlangt werden.

Jedoch gelang der Nachweis diskriminanter Validität anhand der Variablen „Studienfach“: Studierende des Lehramts erreichten auf sämtlichen Testscores signifikant und praktisch bedeutsam höhere Werte als Studierende technischer oder naturwissenschaftlicher Studiengänge. Erwartungsgemäß schnitten somit Studierende mit pädagogischer Ausbildung sowie Praxiserfahrung besser ab als Studierende ohne diese Voraussetzungen. Dieser Befund weist darauf hin, dass der entwickelte Situational Judgement Test in der Tat etwas erfasst, das mit professioneller pädagogischer Kompetenz in Beziehung steht.

Sämtliche Kriterien im Rahmen der Validitätsanalysen können als reliabel betrachtet werden, doch ergeben sich einige Probleme bezüglich einer Kriteriumsdefizienz folgender Kriterien:

Das Kriterium „Anzahl abgeleiteter Praktika“ erfasst zwar die Quantität der praktischen Erfahrung von Lehramtsstudierenden, ist allerdings allenfalls indirekt in der Lage, die tatsächliche professionelle pädagogische Kompetenz der Studierenden zu repräsentieren, da die Quantität und Qualität der praktischen Erfahrung durchaus nicht deckungsgleich sein müssen. Eine Näherung an die tatsächliche Güte der Praxiserfahrung hätte darin bestehen können, die im Rahmen der Praktika erhaltenen Noten zu erfassen. Allerdings werden die Orientierenden und Vertiefenden Praktika des Lehramtsstudiums in Rheinland-Pfalz nicht benotet

(z. B. Ministerium für Bildung, Wissenschaft, Jugend und Kultur, 2010), und somit musste von der Erhebung des Kriteriums „Praktikumsnote“ Abstand genommen werden.

Als zweites Kriterium zum Nachweis der konvergenten Validität wurde die letzte Prüfungsnote in bildungswissenschaftlichen oder didaktischen Fächern erfasst. Da Wissen im Allgemeinen als Voraussetzung von Kompetenz erachtet wird (z. B. Klieme & Hartig, 2007), kann dieses Kriterium zunächst als geeignet angesehen werden. Wird jedoch das Argument des mangelnden Transfers des an der Universität erworbenen Wissens auf die pädagogische Praxis in den Blick genommen (z. B. Blomberg, 2011), so kann das Kriterium als ungeeignet gelten. Das Kriterium „letzte Prüfungsnote in bildungswissenschaftlichen oder didaktischen Fächern“ muss somit zumindest als problematisch eingeordnet werden. Gleiches gilt für das letzte erhobene Außenkriterium: Die „Anzahl besuchter Semester im Lehramtsstudium“ kann zwar dann als geeignetes Kriterium gelten, wenn davon ausgegangen wird, dass Studierende mit verschiedenen Studiendauern sich hinsichtlich ihres Wissens und ihrer Praxiserfahrung unterscheiden – wird jedoch das universitär vermittelte Wissen als Basis professionellen pädagogischen Handelns nur als eingeschränkt wirksam erachtet und die Quantität der Praxiserfahrung nicht mit der tatsächlichen Kompetenz der Studierenden gleichgesetzt, so ist auch dieses Kriterium als nicht ausreichend geeignet für einen konvergenten Validitätsnachweis zu erachten.

Die Ergebnisse der *diskriminanten* Validitätsanalyse anhand des Kriteriums „Studienfach“ weisen auf die Validität des entwickelten Instruments hin. Dabei kann das Kriterium als geeignet gelten, da die Inhalte eines technischen oder naturwissenschaftlichen Studiengangs nicht in Bezug zu professioneller pädagogischer Kompetenz stehen.

Um die *Konstruktvalidität* des entwickelten Verfahrens zu überprüfen, wurde anstelle der für Situational Judgement Tests als ungeeignet geltenden Faktorenanalyse eine Diskriminanzanalyse durchgeführt. Im Ergebnis gelang es nicht, anhand der Variablen „Semester“, „Prüfungsnote“, „Abiturnote“ sowie „Anzahl abgeleiteter Praktika“ eine Trennung in eine Gruppe, deren Gesamtscore *über* dem Median der Gesamtstichprobe lag, sowie eine zweite Gruppe, deren Gesamtscore *unter* dem Median der Gesamtstichprobe lag, vorzunehmen. Diese Befundlage überrascht allerdings wenig, da bereits im Rahmen der Regressionsanalysen eine

Beziehung zwischen den genannten Variablen einerseits und den Testscores andererseits nicht nachgewiesen werden konnte.

Um zusätzlich qualitativ zu analysieren, ob sich die generierten Items den durch die KMK (ebd.) postulierten elf Kompetenzen inhaltlich zuordnen lassen, wurde eine zweite, unabhängige ExpertInnenbefragung durchgeführt. 84 der 102 Items wurden dabei durch die ExpertInnen mehrheitlich übereinstimmend einer KMK-Kompetenz zugeordnet. Allerdings wurde nur bei 48 Items mehrheitlich eine Zuordnung zu der im Rahmen der Itementwicklung *intendierten* Kompetenz vorgenommen. Eine eindeutige Allokation der beschriebenen Situationen scheint somit nicht uneingeschränkt gegeben.

Dieser Befund führt zu der Überlegung, ob sich die durch die Standards für die Lehrerbildung: Bildungswissenschaften (ebd.) überhaupt (konstrukt-)valide erfassen lassen. Blömeke, Felbrich und Müller (2008) beschreiben beispielsweise diese Standards als schwer testbar. Lässt sich dieser Ansicht zustimmen? Es bestehen bereits valide Instrumente zur Diagnostik professioneller pädagogischer Kompetenzfacetten (siehe die Abschnitte 1.4 bzw. 3.1.4), und zusätzlich fußen die innerhalb der Standards genannten Kompetenzen auf breiter empirischer Basis. Eindimensionale Skalen sind bei der Entwicklung von standardorientierten Instrumenten zwar aufgrund heterogener und nicht überschneidungsfreier Kompetenzbereiche nicht zu erwarten, jedoch muss diese mangelnde Konstruktreinheit die Validität der Instrumente nicht zwangsläufig schmälern: Professionelle pädagogische Kompetenz kann als formatives Konstrukt gelten, bei dessen Erfassung nicht die Eindimensionalität sondern die Vollständigkeit der erfassten Inhaltsgereiche im Focus steht (siehe Abschnitt 1.12). Die Kritik, dass die durch die Standards definierten Kompetenzbereiche kaum valide zu messen seien, scheint somit wenig angemessen.

Die Fragestellung der vorliegenden Studie lautete:

Lässt sich die professionelle pädagogische Kompetenz Lehramtsstudierender entlang der Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004) durch einen Situational Judgement Test valide erfassen?

Diese Fragestellung kann vorsichtig mit „Ja“ beantwortet werden, wobei die Validitätsanalyse zum jetzigen Zeitpunkt nicht als abgeschlossen betrachtet werden kann. Um die Validität des Situational Judgement Tests

gesicherter nachweisen zu können, sind erneute Studien an einer repräsentativen Zufallsstichprobe unerlässlich. Hierbei könnten folgende Modifikationen vorgenommen werden:

Die Items des Instruments wurden unter Ausklammerung fachspezifischer Inhalte gestaltet, um eine standardorientierte, fachübergreifende Erfassung professioneller pädagogischer Kompetenz ermöglichen zu können. Zudem sollten Einflüsse der Lesekompetenz und des kognitiven „Loads“ durch hochkomplexe Items verhindert werden. Trotzdem könnten die Items einen zu geringen Anteil fachspezifischer Inhalte aufweisen, wenn der Perspektive gefolgt wird, dass fachdidaktische Ausbildungselemente eine wesentliche Rolle bei der Wirksamkeit von Lehrerbildung spielen (Terhart, 2009). *Fachspezifisch gestaltete Items* könnten sich somit als valider erweisen, wobei die dann erfolgende eingeschränkte Anwendungsbreite des Instruments in Kauf genommen werden müsste.

Die *Inhaltsvalidität* des Instruments könnte durch die Generierung von Items auf Basis der „Critical Incident Technique“ (Flanagan, 1954) erhöht werden. Durch ExpertInnenbefragungen zu besonders (in)effektiven Verhaltensweisen sowie deren Antezedentien und Konsequenzen und die anschließende Clusterung der genannten Ereignisse könnten weitere relevante Situationen für die Operationalisierung professioneller pädagogischer Kompetenz als geeignete Situationsvignetten identifiziert werden. Dieses Verfahren könnte zusätzlich dazu genutzt werden, die *Schwierigkeit* der Items und/oder ihren Anteil an fachspezifischen Inhalten zu erhöhen.

Weitere Analysen zum Nachweis der *konvergenten* Validität sind angezeigt. Dabei sollten solcherlei Kriterien erhoben werden, welche professionelle pädagogische Kompetenz ähnlich handlungsnah erfassen wie Situational Judgement Tests, um Methodenfaktoren sowie Kriteriumskontamination oder -defizienz einzuschränken: Ein handlungsnahes Kriterium könnte die Beurteilung professioneller pädagogischer Kompetenz im Rahmen von Unterrichtsbeobachtungen durch unabhängige BeurteilerInnen darstellen. Noch besser geeignet wäre ein bereits validierter Situational Judgement Tests zur Diagnostik professioneller pädagogischer Kompetenz – hier ließen sich Methodenfaktoren vollständig ausschließen.

Ausgeklammert wurden in der vorliegenden Untersuchung aufgrund ihrer Zielsetzung der Entwicklung eines standard- und simulationsorientierten Instruments

zur Evaluation der Lehrerbildung diejenigen Variablen, welche jenseits der Lehrerbildung einen Einfluss auf die professionelle pädagogische Kompetenz Lehramtsstudierender haben könnten wie beispielsweise Persönlichkeitsvariablen, die eigene Lerngeschichte sowie subjektive Überzeugungen und Einstellungen. Diese Variablen könnten im Rahmen eines *Multi-Trait-Ansatzes* zur Validierung des entwickelten Instrumentes genutzt werden, sofern gesicherte Erkenntnisse zur Beziehung dieser Variablen zur professionellen pädagogischen Kompetenz bestehen.

Bezüglich der *spezifischen Methodik* von Situational Judgement Tests ergeben sich folgende Forschungsaufgaben: Bislang besteht keine Klarheit darüber, wie Situational Judgement Tests idealerweise konstruiert, skaliert und bewertet werden sollten, um ein bestimmtes Konstrukt zu erfassen und ihre Reliabilität zu erhöhen. Zudem existiert bislang kein theoretischer Rahmen, innerhalb dessen Situational Judgement Tests eingegliedert werden könnten. Theoretische Überlegungen anderer psychologischer Bereiche (wie etwa die Forschung zur Wahrnehmung von Situationen, zum Treffen von Entscheidungen oder Urteilen) könnten hilfreich sein, um Prinzipien zur Konstruktion von Situational Judgement Tests bereitstellen zu können (Ployhart & Weekley, 2006).

Trotz ihrer mangelnden theoretischen Verortung gilt die Methode des Situational Judgement Tests bezüglich der prädiktiven und inkrementellen Validität im Hinblick auf professionelle Leistungen als geeignet. Die Vorteile gegenüber häufig zur Eignungsdiagnostik eingesetzter kognitiver oder Persönlichkeitstests liegen hierbei in der Augenscheinvalidität und somit hohen Akzeptanz von Situational Judgement Tests. Zudem sind Situational Judgement Tests handlungsnah gestaltet und nähern sich somit dem Konzept der Kompetenz aufgrund ihres Kontextbezuges besser als abstraktere Testformen. Im Vergleich von papier- vs. videobasierten Umsetzungen sollten sich paper-pencil-Tests bezüglich der Einsetzbarkeit bei großen Stichproben sowie der relativen Entwicklungsökonomie besser eignen. Hinsichtlich der Fidelität beider Verfahren stehen *vergleichende Studien* bislang noch aus.

Jenseits eines Einsatzes zur Evaluation der Lehrerbildung könnte das Instrument im Kontext einer Beratung zur Studienorientierung eingesetzt werden, um sowohl Einblicke in den beruflichen Alltag von Lehrpersonen als auch Hinweise auf eine berufliche (Nicht-)Eignung bereitzustellen. Ebenso könnte der entwickelte Situational Judgement Test als Auswahlinstrument jenseits „klassischer“ Kriterien

(Abiturnote, Fachnoten,...) zur Studienplatzvergabe im Lehramtsstudium verwandt werden. Hierfür müssten vorab Analysen zur inkrementellen Validität des Instruments erfolgen.

Ein weiteres Einsatzgebiet des entwickelten Instruments könnte im Bereich der Förderung liegen: Obwohl der Situational Judgement Test in seiner jetzigen Form zu leicht scheint, so könnte er doch der Funktion nachkommen, besonders *wenig kompetente* ProbandInnen zu identifizieren und so den Förderungsaspekt in den Blick zu nehmen: Ein – norm- oder kriterienorientiert – schlechtes Abschneiden könnte entsprechende Maßnahmen wie beispielsweise ein verpflichtendes zusätzliches Selbststudium oder die Teilnahme an Förderkursen zur Folge haben.

Um einen Beitrag zum Qualitätsmanagement der Lehrerbildung leisten zu können, sollte der entwickelte Situational Judgement Test jedoch erneuten Validitätsprüfungen unterzogen sowie gegebenenfalls gemäß der genannten Aspekte modifiziert werden. Angesichts erster ermutigender Ergebnisse bezüglich der diskriminanten Validität des Instruments sowie des Bedarfs an validen Verfahren zur Evaluation der Lehrerbildung scheinen weitere Studien lohnend zu sein.

## 7. Literatur

- Abs, H. J., Döbrich, P., Vögele, E. & Klieme, E. (2005). Skalen zur Qualität der Lehrerbildung – Dokumentation der Erhebungsinstrumente. *Pädagogische Entwicklungsbilanzen an Studienseminaren (PEB-Sem)*. Frankfurt am Main: DIPF.
- Achtenhagen, F. & Baethge, M. (2007) Kompetenzdiagnostik als Large-Scale-Assessment im Bereich der beruflichen Aus- und Weiterbildung. *Zeitschrift für Erziehungswissenschaft*, 8, 51-70.
- Arnold, K.-H., Koch-Priewe, B., Lin-Klitzing, S. (2007). Allgemeine Didaktik, Fachdidaktik und Unterrichtsqualität. In K.-H. Arnold (Hrsg.), *Unterrichtsqualität und Fachdidaktik* (S. 51-72). Bad Heilbrunn: Verlag Julius Klinkhardt.
- Arnold, R., Pätzold, H. (2002). *Schulpädagogik kompakt. Prüfungswissen auf den Punkt gebracht*. Berlin: Cornelsen.
- Balli, S.J., Demo, D. & Wedman, J.F. (1998). Family involvement with children's homework: An intervention in the middle grades. *Family Relations: Interdisciplinary Journal of Applied Family Studies*, 47, 49–157.
- Ballou, D., Podgursky, M. (1999). Teacher Training and Licensure: A Layman's Guide. In M. Kanstoroom & C. Finn (Eds.), *Better Teachers, Better Schools* (pp. 31-82). Washington: Thomas B. Fordham Foundation.
- Bandura, A. (1990). Conclusion: Reflections on nonability determinants of competence. In R. Sternberg & J. Kolligian Jr. (Eds.), *Competence considered* (pp. 315-362). New Haven: Yale University Press.
- Bastian, J. & Helsper, W. (2000). Professionalisierung im Lehrberuf – Bilanzierung und Perspektiven. In J. Bastian, W. Helsper, S. Reh & C. Schelle (Hrsg.), *Professionalisierung im Lehrberuf* (S. 167-192). Opladen: Leske & Budrich.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 469-520.
- Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, S. Krauss, S. & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 2-54). Münster: Waxmann.
- Bergman, M. E., Drasgow, F., Donovan, M. A. & Juraska, S. E. (2003). *Scoring situational judgement tests*. Paper presented at the 18th annual conference of the Society for Industrial and Organizational Psychology, Orlando, FL.



- Bledow, R. & Frese, M. (2009). A situational judgment test of personal initiative and its relationship to performance. *Personnel Psychology*, 62, 2, 229-258.
- Blömeke, S., Felbrich, A. & Müller, C. (2008). Theoretischer Rahmen und Untersuchungsdesign. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und –referendare. Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (S. 15-48). Münster: Waxmann.
- Blömeke, S., Kaiser, G. & Lehmann, R. (Hrsg.). (2008). *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und –referendare. Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung*. Münster: Waxmann.
- Blomberg, G. A. (2011). Der Einsatz von Unterrichtsvideos in der universitären Lehrerbildung [Elektronische Version]. Zugriff am 11.08.2012 unter <http://d-nb.info/1014330505/34>
- Blum, W. & Krauss, S. (2008). *Professionelle Kompetenz von Mathematiklehrkräften: Was heißt das und welche Auswirkungen hat das auf Unterricht und Schülerleistungen? Vortrag im Rahmen des COACTIV-Projekts. Lernen individuell gestalten – 31. Mai – Congress-Centrum Nord Koelnmesse*.
- Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 100 (3), 305-314.
- Borich, G. D. (2007). *Effective teaching methods*. Columbus, Ohio: Pearson Prentice Hall.
- Bortz, J. & Döring, N. (2003). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (3., überarb. Aufl.). Berlin: Springer.
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (S. 177-212). Göttingen: Hogrefe.
- Brophy, J.E. (2000). Teaching. *Educational Practices Series*, Vol. 1. Brussels: International Academy of Education & International Bureau of Education. Zugriff am 30.04.2009 [www.ibe.unesco.org/publications/EducationalPracticesSeriesPdf/prac01e.pdf](http://www.ibe.unesco.org/publications/EducationalPracticesSeriesPdf/prac01e.pdf).
- Brophy, J., & Good, T. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–375).
- Bruder, S. (2011). *Lernberatung in der Schule. Ein zentraler Bereich professionellen Lehrerhandelns* [Elektronische Version]. Zugriff am 31.05.2011 unter

[http://tuprints.ulb.tu-darmstadt.de/2432/1/Dissertation\\_Bruder\\_Lernberatung\\_070311.pdf](http://tuprints.ulb.tu-darmstadt.de/2432/1/Dissertation_Bruder_Lernberatung_070311.pdf)

- Bruder, S., Klug, J., Hertel, S. & Schmitz, B. Messung, Modellierung und Förderung der Beratungskompetenz und Diagnostischen Kompetenz von Lehrkräften. *Lehrerbildung auf dem Prüfstand*, (2010), 173-193.
- Bruder, S., Keller, S., Kelava, A. & Schmitz, B. (2011). A Situational Judgment Test for Measuring Teachers' Counseling Competency. In S. Bruder, *Lernberatung in der Schule. Ein zentraler Bereich professionellen Lehrerhandelns* [Elektronische Version]. Zugriff am 14.08.2012 unter [http://tuprints.ulb.tu-darmstadt.de/2432/1/Dissertation\\_Bruder\\_Lernberatung\\_070311.pdf](http://tuprints.ulb.tu-darmstadt.de/2432/1/Dissertation_Bruder_Lernberatung_070311.pdf)
- Brünken, R. (2009). *Diagnostik in der Lehrerbildung*. Vortrag anlässlich des 43. BAK Seminartags in Saarbrücken am 02.10.2009.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson.
- Bühner, M. & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson.
- Burmeister, K. (2009). *Komplexes Problemlösen im Kontext angewandter Eignungsdiagnostik* [Elektronische Version]. Zugriff am 08.01.2012 unter [http://ub-ed.ulb.uni-grafschwald.de/opus/volltexte/2009/614/pdf/Burmeister\\_Kai\\_Komplexes\\_Problemloesen\\_im\\_Kontext\\_angewandter\\_Eignungsdiagnostik.pdf](http://ub-ed.ulb.uni-grafschwald.de/opus/volltexte/2009/614/pdf/Burmeister_Kai_Komplexes_Problemloesen_im_Kontext_angewandter_Eignungsdiagnostik.pdf)
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Ceci, S. J., Barnett, S. M. & Kanaya, T. (2003). Developing childhood proclivities into adult competencies: the overlooked multiplier effect. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The Psychology of Abilities, Competencies, and Expertise* (pp. 70-92). New York: Cambridge University Press.
- Chan, D. & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational Judgement tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82, 143-159.
- Chan, D. & Schmitt, N. (2002). Situational Judgement Tests and job performance. *Human Performance*, 15, 233-254.
- Chomsky, N. (1968). *Language and Mind*. New York: Harcourt, Brace & World.

- Christian, M. S., Edwards, B. D., Bradley, J. C. (2010). Situational Judgment Tests: Constructs assessed and a meta-analysis of criterion-related validities. *Personnel Psychology* [Elektronische Version]. Zugriff am 09.07.2010 unter [http://findarticles.com/p/articles/mi\\_hb3338/is\\_201004/ain53078768/?tag=content;col1](http://findarticles.com/p/articles/mi_hb3338/is_201004/ain53078768/?tag=content;col1).
- Clausen, M. (2002). *Unterrichtsqualität – eine Frage der Perspektive?* Münster: Waxmann.
- Clevenger, J. P., Pereira, G. M., Wiechmann, D., Schmitt, N. & Schmidt Harvey, V. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, 86, 410-417.
- Cochran-Smith, M. & Fries, M. K. (2002). The discourse of reform in teacher education: Extending the dialogue. *Educational Researcher*, 31 (6), 26-29.
- Cortina, J. M. (1993). What is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology*, 78, 98-104.
- Cramer, C. (2010). Sozioökonomische Stellung Lehramtsstudierender. *Lehrerbildung auf dem Prüfstand.*, 3 (1), 4-22.
- Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 16, 297-334.
- Cucina, J. M., Vasilopoulos, N. L. & Leaman, J. A. (2003). *The bandwidth-fidelity dilemma and situational Judgement test validity*. Paper presented at the 18th annual conference of the Society for Industrial and Organizational Psychology, Orlando, FL.
- Darling-Hammond, L. (2000). Teacher Quality and Student Achievement: A Review of State Policy Evidence. *Education Policy Analysis Archives* 8 (1), pp. 1-46.
- Deutsche Akademie der Wissenschaften (2003). *Digitales Wörterbuch der deutschen Gegenwartssprache des 20. Jahrhunderts* [Elektronische Version]. Zugriff am 20.08.2010 unter [http://www.dwds.de/woerterbuch/1\\_1](http://www.dwds.de/woerterbuch/1_1)
- Diehl, T. (2005). Pädagogische Professionalität – Möglichkeiten ihrer empirischen Erfassung. In A. Frey, R. S. Jäger & U. Renold (Hrsg.), *Kompetenzdiagnostik – Theorien und Methoden zur Erfassung und Bewertung von beruflichen Kompetenzen (Berufspädagogik, Band 5)* (S.116-135). Landau: Verlag Empirische Pädagogik.
- Dobey, D. C. & Schafer, L.E. (1984). *The effects of knowledge on elementary science inquiry teaching*. *Science Education*, 68, 39-51.

- Dries, C. & Bovi, C. (2004). *360 Grad Feedback* [Elektronische Version]. Zugriff am 02.06.2011 unter [http://www.ki-bit.com/web/component?option=com\\_docman/task/doc\\_view/gid,30/Itemid,62/](http://www.ki-bit.com/web/component?option=com_docman/task/doc_view/gid,30/Itemid,62/)
- Erpenbeck, J. & Rosenstiel, von L. (2007). *Handbuch Kompetenzmessung*. Stuttgart: Schäffer-Poeschel.
- Fisseni, H.-J. (2004). *Lehrbuch der psychologischen Diagnostik*. Göttingen: Hogrefe.
- Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin*, *41*, 237-358.
- Frese, M., Fay, D., Hilburger, T. & Leng, K. (1997). The concept of personal initiative: Operationalization, reliability and validity of two German samples. *Journal of Occupational & Organizational Psychology*, *70*, 139-161.
- Frey, A. & Jung, C. (2011). Kompetenzmodelle, Standardmodelle und Professionsstandards in der Lehrerbildung: Stand und Perspektiven. *Lehrerbildung auf dem Prüfstand, Sonderheft*, 1-4.
- Galluzzo, G. & Craig, J. (1990). Evaluation of Preservice Teacher Education Programs. In R. W. Houston (Ed.), *Handbook of Research on Teacher Education* (pp. 590 – 616). New York: Macmillan.
- Gesetz- und Verordnungsblatt für das Land Rheinland-Pfalz: Landesverordnung über die Erste Staatsprüfung für Lehrämter nach Abschluss der Hochschulprüfungen lehramtsbezogener Bachelor- und Masterstudiengänge* (2007) [Elektronische Version]. Zugriff am 28.01.2012 unter [http://www.mbwjk.rlp.de/fileadmin/Dateien/Downloads/Bildung/Lehrerbildung/4.1.1\\_-\\_LVOVersionJM070823.pdf](http://www.mbwjk.rlp.de/fileadmin/Dateien/Downloads/Bildung/Lehrerbildung/4.1.1_-_LVOVersionJM070823.pdf)
- Gettinger, M. & Kohler, K.M. (2006). Process-Outcome Approaches to Classroom Management and Effective Training. In C. M. Evertson & C.S. Weinstein (eds.), *Handbook of Classroom Management. Research, Practice, and Contemporary Issues* (pp. 73-95). Mahwah, N.J.: Lawrence Erlbaum.
- Gräsel, C., Krolak-Schwerdt, S., Nölle, I. & Hörstermann, T. (2010). Diagnostische Kompetenz von Grundschullehrkräften bei der Erstellung der Übergangsempfehlung. Eine Analyse aus der Perspektive der sozialen Urteilsbildung. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes*(S. 286-295). Weinheim: Beltz.
- Griese, C. (2009). Lehrer beraten Eltern. Ratgeberliteratur für Lehrer zur Elternarbeit. *PÄD-Forum: unterrichten erziehen* 37/28, 5, 213-216.
- Grouws, D. A. & Cebulla, K. J. (2000). Improving student achievement in mathematics. In *Educational Practices Series, Vol. 4*. Brussels: International

- Academy of Education & International Bureau of Education Zugriff am 30.04.2009 von [www.ibe.unesco.org/publications/EducationalPracticesSeriesPdf/prac04e.pdf](http://www.ibe.unesco.org/publications/EducationalPracticesSeriesPdf/prac04e.pdf)
- Gudjons, H. (1998). *Didaktik zum Anfassen*. Bad Heilbrunn: Klinkhardt.
- Gudjons, H. (2000). Belastungen und neue Anforderungen. Aspekte der Diskussion um Lehrer und Lehrerinnen in den 80er und 90er Jahren. In J. Bastian, W. Helsper, S. Reh & C. Schelle (Hrsg.), *Professionalisierung im Lehrberuf* (S. 33-54). Opladen: Leske + Budrich.
- Häder, M. (2009). *Delphi-Befragungen: Ein Arbeitsbuch*. Wiesbaden: Verlag für Sozialwissenschaften.
- Hartig, J. & Jude, N. (2007). Empirische Erfassung von Kompetenzen und psychometrische Kompetenzmodelle. In J. Hartig & E. Klieme (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik* (S. 17-36). Berlin: BMBF.
- Hartig, J., Klieme, E., Jude, N., Jurecka, A., Kröhne, U., Maag-Merki, K., Reef, J.-P. & Wirt, J. (2007). *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik*. Bonn, Berlin: BMBF.
- Heidenreich, M. (1999). Berufskonstruktion und Professionalisierung. Erträge der soziologischen Forschung. In H.-J. Apel, K.-P. Horn, P. Lundgreen & U. Sandfuchs (Hrsg.), *Professionalisierung pädagogischer Berufe im historischen Prozess* (S. 35-58). Bad Heilbrunn: Julius Klinkhardt.
- Helmke, A. (2004). Was wissen wir über guten Unterricht?. *Pädagogik (Große Serie 2006: Forschung-Schule-Unterricht. Befunde und Konsequenzen)*, 2, 42-45.
- Helmke, A. (2007). *Unterrichtsqualität erfassen – bewerten – verbessern*. Seelze: Klett Kallmeyer.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze-Velber: Klett Kallmeyer.
- Helmke, A., Helmke, T., Kleinbub, I., Nordheider, I., Schrader F.-W. & Wagner, W. (2007). *Die DESI-Videostudie. Der fremdsprachliche Unterricht – Englisch*, 41 (90), 37-44.
- Helmke, A., Helmke, T. & Schrader, F.-W. (2007). Unterrichtsqualität: Brennpunkte und Perspektiven der Forschung. In K.-H. Arnold (Hrsg.), *Unterrichtsqualität und Fachdidaktik* (S. 51-72). Bad Heilbrunn: Verlag Julius Klinkhardt.
- Helmke, A. & Schrader, F.-W. (2006). Lehrerprofessionalität und Unterrichtsqualität: Den eigenen Unterricht beurteilen und reflektieren. *Schulmagazin 5-10*, 1-13.
- Herzog, W. (2005). *Müssen wir Standards wollen? Skepsis gegenüber einem theoretisch (zu) schwachen Konzept*. *Zeitschrift für Pädagogik*, 51, 252-258.

- Hill, H. C., Rowan, B. & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42, 371-406.
- Hooper, A.C., Cullen, M.J. & Sackett, P.R. (2006). Operational threats to the use of SJTs: faking, coaching, and retesting issues. In J. Weekley & R.E. Ployhart (Eds), *Situational Judgment Tests* (pp. 205-32). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hunter, D. R. (2003). Measuring general aviation pilot judgement using a situational judgement technique. *International Journal of Aviation Psychology*, 13, 373-386.
- Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Imhof, M. (1985). Erziehung zur Konfliktfähigkeit mit Hilfe von Selbsterfahrungsarbeit in der Schule. *Praxis der Kinderpsychologie und Kinderpsychiatrie* 34, 6, 231-239.
- Jäger, R. S. & Petermann, F. (Hrsg.). (1999). *Psychologische Diagnostik. Ein Lehrbuch*. Weinheim: Beltz Psychologie Verlags Union.
- Jäger, R. S. (1999). Der diagnostische Prozess. In R. S. Jäger & F. Petermann (Hrsg.), *Psychologische Diagnostik* (4. Aufl., S. 450-455). Weinheim: Beltz Psychologie Verlags Union.
- Jäger, R.S., Bodensohn, R. & Frey, A. (2008). Bewerten und Prüfen – Beurteilen und Fördern. Bedingungen für das Studium im Lehramt. *Journal für Lehrerinnen- und Lehrerbildung*, 1, 7-15.
- Jarvis, C. B., MacKenzie, S. B. & Podsakoff, P. M. (2003), A critical review of construct indicators and measurement model misspecification in marketing and consumer research, *Journal of Consumer Research*, 30 (2), pp. 199–218.
- Jurt, U. (2007). *Professionsstandards. Beobachtungsinstrument für die berufspraktische Ausbildung*. PHZ Schwyz.
- Kanning, U. P., Grewe, K., Hollenberg, S. & Hadouch, M. (2006). From the Subject's Point of View: Reaction to Different Types of Situational Judgment Items. *European Journal of Psychological Assessment*, 22, 168-176.
- Kelchtermans, G. (2006). Teacher collaboration and collegiality as workplace conditions. A review. *Zeitschrift für Pädagogik*, 52, 220-237
- Keller, G. (2003). *Selbstmanagement im Lehrerberuf*. Donauwörth: Auer Verlag.
- Kersting, M. (1999). *Diagnostik und Personalauswahl mit computergestützten Problemlöseszenarios? Zur Kriteriumsvalidität von Problemlöseszenarios und Intelligenztests*. Göttingen: Hogrefe.

- Kirkpatrick, D.L. (1998). *Evaluating Training Programs. The Four Levels*. San Francisco: Berrett-Koehler.
- Klieme, E. & Hartig, J. (2007). Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. *Kompetenzdiagnostik. Zeitschrift für Erziehungswissenschaft*, 8, 11-32.
- Klieme, E. & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts, *Zeitschrift für Pädagogik* 54 (2), S. 222-237.
- Klieme, E., Schümer, G. und Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: „Aufgabenkultur“ und Unterrichtsgestaltung. In Bundesministerium für Bildung und Forschung (Hrsg.), *TIMSS – Impulse für Schule und Unterricht. Forschungs-befunde, Reforminitiativen, Praxisberichte und Video-Dokument* (S. 43-57). Bonn: BMBF.
- Klusmann, U. (2011). Allgemeine berufliche Motivation und Selbstregulation. In M. Kunter, J. Baumert, W. Blum, S. Krauss, S. & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 277-294). Münster: Waxmann.
- Klusmann, U., Kunter, M., Trautwein, U. & Baumert, J. (2006). Lehrerbelastung und Unterrichtsqualität aus der Perspektive von Lehrenden und Lernenden. *Zeitschrift für Pädagogische Psychologie*, 20, 161-173.
- Köller, O., Baumert, J. & Neubrand, J. (2000). Epistemologische Überzeugungen und Fachverständnis im Mathematik- und Physikunterricht. In J. Baumert, W. Bos & R. H. Lehmann (Hrsg.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn*. (Mathematische und naturwissenschaftliche Kompetenzen am Ende der gymnasialen Oberstufe, Vol.2) (S. 229-270). Oplande: Leske+Budrich.
- Kolbe, F.-U. (2004). Verhältnis von Wissen und Handeln. In S. Blömeke, P. Reinhold, G. Tulodziecki & J. Wildt (Hrsg.), *Handbuch Lehrerbildung* (S. 206-232). Bad Heilbrunn: Julius Klinkhardt.
- König, J. (2010). Längsschnittliche Erhebung pädagogischer Kompetenzen von Lehramtsstudierenden (LEK): Theoretischer Rahmen, Fragestellungen, Untersuchungsanlage und erste Ergebnisse zu Lernvoraussetzungen von angehenden Lehrkräften. *Lehrerbildung auf dem Prüfstand*, 3 (1), 56-83.

- König, J. & Blömeke, S. (2009). Disziplin- oder Berufsorientierung? Zur Struktur des pädagogischen Wissens angehender Lehrkräfte. *Lehrerbildung auf dem Prüfstand*, 2 (1), S. 126-147.
- König, J. & Lebens, M. (2012). Classroom Management Expertise (CME) von Lehrkräften messen: Überlegungen zur Testung mithilfe von Videovignetten und erste empirische Befunde. *Lehrerbildung auf dem Prüfstand*, 5 (1), 3-29.
- Krauss, S., Kunter, M., Brunner, M., Baumert, J., Blum, W., Neubrand, M., Jordan, A., & Löwen, K. (2004). COACTIV: Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule: Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* (S. 31–53). Münster: Waxmann.
- Krauss, S., Blum, W., Brunner, M., Neubrand, M., Baumert, J., Kunter, M., Besser, M. & Elsner, J. (2011). *Konzeptualisierung und Testkonstruktion zum fachbezogenen Professionswissen von Mathematiklehrkräften*. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (S. 135-161). Münster: Waxmann.
- Kounin, J.S. (1976). *Techniken der Klassenführung*. Bern: Huber.
- Krumm, V. (1996): Über die Vernachlässigung der Eltern durch Lehrer und Erziehungswissenschaft. Plädoyer für eine veränderte Rolle der Lehrer bei der Erziehung der Kinder. In A. Leschinsky (Hrsg.), *Die Institutionalisierung von Lehren und Lernen. Beiträge zu einer Theorie der Schule*. 34. Beiheft der *Zeitschrift für Pädagogik*. Weinheim: Beltz, (S. 119–137).
- Kunter, M. (2011). Theorie meets Praxis in der Lehrerbildung – Kommentar. *Erziehungswissenschaft* 22 (2011) 43, S. 107-112.
- Kunter, M., Kleickmann, T., Klusmann, U. & Richter, D. (2011). Die Entwicklung professioneller Kompetenz von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, S. Krauss & M. Neubrand, (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 55-68). Münster: Waxmann.
- Kunter, M. (2010). Modellierung von Lehrerkompetenzen. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (S. 307-313). Weinheim: Beltz.



- Kunter, M., Baumert, J., Blum, W., Krauss, S. & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Kultusministerkonferenz (2004). Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. *Standards für die Lehrerbildung: Bildungswissenschaften*. Beschluss der Kultusministerkonferenz vom 16.12.2004 [Elektronische Version]. Zugriff am 20.05.2009 unter [http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2004/2004\\_12\\_16-Standards-Lehrerbildung.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf).
- Kultusministerkonferenz (2010). *Ländergemeinsame Strukturvorgaben für die Akkreditierung von Bachelor- und Masterstudiengängen* [Elektronische Version]. Zugriff am 10.09.2010 unter [http://www.hrk-bologna.de/bologna/de/download/dateien/2003\\_10\\_10-Laendergemeinsame-Strukturvorgaben.pdf](http://www.hrk-bologna.de/bologna/de/download/dateien/2003_10_10-Laendergemeinsame-Strukturvorgaben.pdf)
- Kuß, A. & Eisend, M. (2010). *Marktforschung. Grundlagen der Datenerhebung und Datenanalyse*. Wiesbaden: Gabler.
- Küstings, H., Billich, M. und Lipowsky, F. (2009). Der Einfluss von Lehrerkompetenz und Lehrerhandeln auf den Schulerfolg von Lernenden. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (S. 655-667). Weinheim und Basel: Beltz.
- Lehmann, G. & Nieke, W. (2008). *Zum Kompetenzmodell* [Elektronische Version]. Zugriff am 21.04.2010 unter <http://www.bildungsserver-mv.de/download/material/text-lehmann-nieke.pdf>
- Leuchtenberger, G. (1909). *Vademecum für junge Lehrer. Pädagogisch-didaktische Erfahrungen und Ratschläge*. Berlin: Weidmannsche Buchhandlung.
- Leutner, D. (1992). *Adaptive Lehrsysteme. Instruktionspsychologische Grundlagen und experimentelle Analysen* (Handwörterbuch Pädagogische Psychologie). Weinheim: Beltz.
- Leutwyler, B. & Ettl, E. (2008). Standards in der Weiterbildung – Perspektiven eines integralen Steuerungskonzeptes. *Beiträge zur Lehrerbildung*, 26, 11-22.
- Lienert, G. A. (1967). *Drahtbiegeprobe (DBP)* (2. Aufl.). Göttingen: Hogrefe.
- Lievens, F., Buyse, T. & Sackett, P.R. (2005a). Retest effects in operational selection settings: development and test of a framework. *Personnel Psychology*, Vol. 58, 981-1007.

- Lievens, F., Buyse, T. & Sackett, P.R. (2005b). The operational validity of a video-based situational Judgement test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology, 90*, 442-452.
- Lievens, F., Peeters, H. & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review, Vol. 37, 4*, 426-441.
- Lievens, F., Sackett, P.R. & Buyse, T. (2009). The Effects of Response Instructions on Situational Judgment Test Performance and Validity in a High-Stakes Context. *Journal of Applied Psychology, Vol. 94*, 1095–1101.
- Lievens, F. & Sackett, P. R. (2007). Situational judgment tests in highstakes settings: Issues and strategies with generating alternate forms. *Journal of Applied Psychology, 92*, 1043-1055.
- Mayer, R. E. (2003). What Causes Individual Differences in Cognitive Performance? In R. J. Sternberg & E. L. Grigorenko (Eds.), *The Psychology of Abilities, Competencies, and Expertise* (pp. 263-274). New York: Cambridge University Press.
- McClelland (1987). *Human motivation*. Cambridge, UK: Cambridge University Press.
- McClelland, D. C. (1973). Testing for competence rather than for intelligence. *American Psychologist, 28*, 1-14.
- McClelland, D. C. (1998). Identifying competencies with behavioral-event interviews. *Psychological Science, 9 (5)*, 331-339.
- McDaniel, M. A., Finnegan, E. B., Morgeson, F. P., Campion, M. A. & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*, 730-740.
- McDaniel, M. A., Hartman, N. S. & Grubb, W. L. (2003a). *Response instructions as moderators of the validity of situational Judgement tests*. Paper presented at the 18<sup>th</sup> annual conference of the Society for Industrial and Organizational Psychology, Orlando, FL.
- McDaniel, M. A., Hartman, N. S., Grubb, W. L. (2003b). *Situational Judgement Tests, knowledge, behavioral tendency, and validity: A meta-analysis*. Paper presented at the 18<sup>th</sup> annual conference of the Society for Industrial and Organizational Psychology, Orlando, FL.
- McDaniel, M.A., Hartman, N.S., Whetzel, D.L & Grubb, W.B. (2007). Situational judgment tests, response instructions and validity: a meta-analysis. In Burke, M.J. (Ed.). *Personnel psychology* (S.63-84). Blackwell Publishing.

- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A. & Braverman, E. P. (2001). Use of situational Judgement tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 80*, 730-740.
- McDaniel, M. A. & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9*, 103-113.
- McDaniel, M.A. & Whetzel, D.L. (2005). Situational judgment test research: Informing the debate on practical intelligence theory. *Intelligence, 33*, 515–525.
- McDaniel, M. A., Whetzel, D. L., Hartman, N. S., Nguyen, N. T. & Grubb, W. L. (2006). Situational Judgement Tests: Validity and an Integrative Model. In J. A. Weekley & R. E. Ployhart (2006). *Situational Judgement tests. Theory, measurement and application* (pp. 183-204). Mahwah, N.J.: Lawrence Erlbaum.
- Messner, R. (2004). Was Bildung von Produktion unterscheidet – oder die Spannung von Freiheit und Objektivierung und das Projekt der Bildungsstandards. In W. Böttcher & R. Messner (Hrsg.), *Mit Standards Menschen bilden? Hofgeismarer Vorträge*. Bd. 23/2004, S. 19-45.
- Meyer, H. (2004). *Was ist guter Unterricht?* Berlin: Cornelsen.
- Ministerium für Bildung, Wissenschaft, Jugend und Kultur Rheinland-Pfalz (2010). *Die neuen lehramtsbezogenen Bachelor- und Masterstudiengänge* [Elektronische Version]. Zugriff am 01.08.2010 unter [http://www.mbwjk.rlp.de/fileadmin/mbwjk/Bildung/lehrerberuf/Reform\\_der\\_Lehrerinnen\\_und\\_Lehrerbildung/Flyer\\_Studiengaenge.pdf](http://www.mbwjk.rlp.de/fileadmin/mbwjk/Bildung/lehrerberuf/Reform_der_Lehrerinnen_und_Lehrerbildung/Flyer_Studiengaenge.pdf)
- Möller, J.H. (2010). *Kundenorientierung im Hotelfach. Die Entwicklung und Validierung eines Situational Judgment Tests*. Münster: Wissenschaftliche Schriften der WWU Münster.
- Moosbrugger, H. & Kelava A. (Hrsg.). (2007). *Testtheorie und Fragebogenkonstruktion*. Heidelberg: Springer.
- Motowidlo, S. J., Dunnette, M. D. & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*, 640-647.
- Motowidlo, S. J., Hanson, M. A. & Crafts, J. L. (1997). Low-fidelity simulations. In D. L. Whetzel & G. R. Wheaton (Eds.), *Applied measurement methods in industrial psychology* (pp. 241-260). Palo Alto, CA: Davies-Black.
- Motowidlo, S. J., Hooper, A. C. & Jackson, H. L. (2006). A Theoretical Basis for Situational Judgement Tests. In J. A. Weekley & R. E. Ployhart (2006).

- Situational Judgement tests. Theory, measurement and application* (pp. 57-79). Mahwah, N.J.: Lawrence Erlbaum.
- Mulder, R. H., Messmann, G. & Gruber, H. (2009). Professionelle Entwicklung von Lehrenden als Verbindung von Professionalität und professionellem Handeln. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (S. 401-409). Weinheim und Basel: Beltz.
- Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78 (3), 691-692.
- National Board for Professional Teaching Standards (2009). *The five core propositions* [Elektronische Version]. Zugriff am 31.05.2010 unter [http://www.nbpts.org/the\\_standards/the\\_five\\_core\\_propositions](http://www.nbpts.org/the_standards/the_five_core_propositions)
- Nerdinger, F., Blickle, G. & Schaper, N. (2008). *Lehrbuch Arbeits- und Organisationspsychologie*. Heidelberg: Springer.
- Neuweg, G. H. (2004). Figuren der Relationierung von Lehrerwissen und Lehrerkönnen. In B. Hackl & G. H. Neuweg (Hrsg.), *Zur Professionalisierung pädagogischen Handelns* (S. 1-26). Münster: LIT.
- Nguyen, N. T., Biderman, M. & McDaniel, M. A. (2005). Effects of response instructions on faking in a situational Judgement test. *International Journal of Selection and Assessment*, 13 (4), 250-260.
- O'Connell, M.S., Hartman, N.S., McDaniel, M.A., Lee Grubb, W. & Lawrence, A. (2007). Incremental Validity of Situational Judgement Tests for Task and Contextual Job Performance. *International Journal of Selection and Assessment*, 15, 19-29.
- Oevermann, U. (1997). Theoretische Skizze einer revidierten Theorie professionalisierten Handelns. In A. Combe & W. Helsper (Hrsg.), *Pädagogische Professionalität* (S. 70-182). Frankfurt a. M.: Suhrkamp Verlag.
- Olson-Buchanan, J. B., Drasgow, F., Moberg, P. J., Mead, A. D., Keenan, P. A. & Donovan, M. A. (1998). Interactive video assessment of conflict resolution skills. *Personnel Psychology*, 51, 1-24.
- Olson-Buchanan, J. B. & Drasgow, F. (2006). Multimedia Situational Judgment Tests: The Medium Creates the Message. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 253-278). Mahwah: Lawrence Erlbaum Associates Publishers.
- Olweus, D. (2002). *Gewalt in der Schule: Was Lehrer und Eltern wissen sollten – und tun können*. Bern: Huber.

- Oser, F. & Oelkers, J. (Hrsg.). (2001). *Die Wirksamkeit der Lehrerbildungssysteme. Von der Allrounderbildung zur Ausbildung professioneller Standards*. Chur, Zürich: Rüegger.
- Oser, F. (2003). Professionalisierung der Lehrerbildung durch Standards. In D. Lemmermöhle & D. Jahreis (Hrsg.), *Professionalisierung der Lehrerbildung* (S. 71-82). Weinheim: Juventa.
- Oser, F. (2004). Standardorientierte Evaluation der Lehrerbildung. In S. Blömeke, P. Reinhold, G. Tulodziecki & J. Wildt (Hrsg.), *Handbuch Lehrerbildung* (S. 184-206). Münster: Westermann.
- Oser, F. (2005). Schrilles Theoriegezerre, oder warum Standards gewollt sein sollen. Eine Replik auf Walter Herzog. *Zeitschrift für Pädagogik*, 51, 266-274.
- Oser, F. (2006). Moralentwicklung und Moralförderung. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 502-509). Weinheim: Beltz.
- Palhaus, D.L. (1984). Two component models of social desirable responding. *Journal of Personality and Social Psychology*, 46, 598-609.
- Perrez, M., Huber, G.L., Geißler, K. A. (2001). Psychologie der pädagogischen Interaktion. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (S. 357-414). Weinheim: Beltz.
- Phillips, J. F. (1992). Predicting sales skills. *Journal of Business and Psychology*, 7, 151-160.
- Pixner, J. & Schüpbach, H. (2008). Chancen und Grenzen von Situational Judgment Inventories für die Studierendenauswahl. In H. Schuler (Hrsg.), *Studierendenauswahl und Studienentscheidung* (S. 139-146). Göttingen: Hogrefe.
- Ployhart, R. E. & Ehrhart, M. G. (2003). Be careful what you ask for: Effects of response instructions on the construct validity and reliability of situational judgement tests. *International Journal of Selection and Assessment*, 11, 1-16.
- Ployhart, R. E. & Weekley, J. A. (2006). Situational Judgment: Some Suggestions for Future Science and Practice. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational Judgement tests. Theory, measurement and application* (pp. 345-350). Mahwah, N.J.: Lawrence Erlbaum.
- Prenzel, M., Seidel, T., Lehrke, M., Rimmel, R., Duit, R., Euler, M., Geiser, H., Hoffmann, L., Müller, C. & Widodo, A. (2002). Lehr-Lernprozesse im Physikunterricht - eine Videostudie. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (S. 139-156). Weinheim: Beltz.

- Reinmann-Rothmeier, G. & Mandl, H.. (2001). Unterrichten und Lernumgebungen gestalten. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (S. 601-646). Weinheim: Beltz.
- Rekus, J. (2009). Werterziehung in Schule und Familie. *PÄD-Forum: unterrichten erziehen* 37/28, 2, 71-74.
- Rettler, H. (1999). Normorientierte Diagnostik. In R. S. Jäger & F. Petermann (Hrsg.), *Psychologische Diagnostik* (4. Aufl., S. 221-226) Weinheim: BeltzPVU.
- Reynolds, D. H., Winter, J. L. & Scott, D. R. (1999). *Development, validation, and translation of a professional-level situational Judgement inventory*. Paper presented at the 14<sup>th</sup> annual conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.
- Reynolds, D. H., Sydell, E. J., Scott, D. R. & Winter, J. L. (2000). *Factors Affecting Situational Judgement Test Characteristics*. Paper presented at the 15<sup>th</sup> annual conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.
- Rheinberg, F. (2006). Motivationstraining und Motivierung. In D. H. Rost (Hrsg.), *Hand-wörterbuch Pädagogische Psychologie* (S. 510-515). Weinheim: Beltz.
- Rheinberg, F., Bromme, R., Minsel, B., Winteler, A., Weidenmann, B. (2001). Die Erziehenden und Lehrenden. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (S. 513-564). Weinheim: Beltz.
- Riese, J. (2010). Empirische Erkenntnisse zur Wirksamkeit der universitären Lehrerbildung. Indizien für notwendige Veränderungen der fachlichen Ausbildung von Physik Lehrkräften. *Physik und Didaktik in Schule und Hochschule*, 1/9, 25-33.
- Riese, J. & Reinhold, P. (2008). Entwicklung und Validierung eines Instruments zur Messung professioneller Handlungskompetenz bei (angehenden) Physik Lehrkräften. *Lehrerbildung auf dem Prüfstand*, 2, S. 625-641.
- Rolff, H.-G. (Hrsg.). (2011). *Qualität mit System. Ein Arbeitsbuch Unterrichtsbezogenen Qualitätsmanagements (UQM)*. Köln: Link.
- Rost, J. (2004). *Testtheorie – Testkonstruktion*. Bern: Verlag Hans Huber.
- Sacco, J. M., Scheu, C. R., Ryan, A. M., Schmitt, N., Schmidt, D. B., & Rogg, K. L. (2000). *Reading level and verbal test scores as predictors of subgroup differences and validities of Situational Judgment Tests*. Paper presented at the 15<sup>th</sup> Annual Meeting of the Society for Industrial and Organizational Psychology. New Orleans, LA.
- Sarges, W. (Hrsg.). (2001). *Weiterentwicklungen der Assessment Center-Methode* (2., überab. u. erw. Aufl., S. VII-XXXII). Göttingen: Hogrefe.

- Saterdag, H. (2004). Professionalität als Ziel der Lehrerbildung. *Journal für Lehrerinnen- und Lehrerbildung*, 1, 60-68.
- Schaarschmidt, U. (2005). *Halbtagsjobber? Psychische Gesundheit im Lehrerberuf – Analyse eines veränderungswürdigen Zustandes*. Weinheim: Beltz.
- Schaper, N. (2009). Aufgabenfelder und Perspektiven bei der Kompetenzmodellierung und –messung in der Lehrerbildung. *Lehrerbildung auf dem Prüfstand*, 2 (1), 166-199.
- Schaper, N., Hilligus, A. H. & Reinhold, P. (2009). Kompetenzmodellierung und –messung in der Lehrerbildung. *Lehrerbildung auf dem Prüfstand*, 2 (1), 1-9.
- Scheerens, J. & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.
- Scheithauer, H. & Bull, H.D. (2009). Aggression und Gewalt an Schulen. Erfolgreiche Handlungsansätze. *PÄD-Forum: unterrichten erziehen* 37/28, 1, 13-16.
- Scheurer, H. (1999). Diagnostik als Testung. In R. S. Jäger & F. Petermann (Hrsg.), *Psychologische Diagnostik* (4. Aufl., S. 257- 263). Weinheim: BeltzPVU.
- Schirp, H. (2003). Werteerziehung und Schulentwicklung. Konzeptuelle und organisatorische Ansätze zur Entwicklung einer demokratischen und sozialen Lernkultur. In C. Palentien, & K. Hurrelmann (Hrsg.), *Schülerdemokratie: Mitbestimmung in der Schule* (S. 47–67). München: Luchterhand.
- Schmelzing, S. , Fuchs, C., Wüsten, S., Sandmann, A. & Neuhaus, B. (2009). Entwicklung und Evaluation eines Instruments zur Erfassung des fachdidaktischen Reflexionswissens von Biologielehrkräften. *Lehrerbildung auf dem Prüfstand*, 2 (1), 57-81.
- Schmidt, W. H., Tatto, M. T., Bankov, K., Blömeke, S., Cedillo, T., Cogan, L., Han, S.-I., Houang, R., Hsieh, F.-J., Paine, L., Santillan, M. N. & Schwille, J. (2007). *The Preparation Gap: Teacher Education for Middle School Mathematics in Six Countries – Mathematics Teaching in the 21<sup>st</sup> Century (MT 21)* [Elektronische Version]. Zugriff am 11.05.2010 unter <http://usteds.msu.edu/MT21Report.pdf>.
- Schmitt, N. and Chan, D. (2006). Situational judgment tests: method or construct?. In J. Weekley & R.E. Ployhart (Eds.), *Situational Judgment Tests* (pp. 135-56). Mahwah, NJ: Lawrence Erlbaum Associates
- Schmotz, C. & Blömeke, S. (2009). Zum Verhältnis von fachbezogenem Wissen und epistemologischen Überzeugungen bei angehenden Lehrkräften. *Lehrerbildung auf dem Prüfstand*, 2 (1), 148-165.

- Schnabel, K. (2001). Psychologie der Lernumwelt. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (S. 467-512). Weinheim: Beltz.
- Schnebel, S. (2007). *Professionell beraten. Beratungskompetenz in der Schule*. Weinheim: Beltz.
- Schneewind, K. A. & Pekrun, R. (1994). Theorien der Erziehungs- und Sozialisationspsychologie. In K. A. Schneewind (Hrsg.), *Psychologie der Erziehung und Sozialisation* (S. 3-40), Göttingen: Hogrefe.
- Schocker-von Dittfurth, M. (2001). *Forschendes Lernen in der fremdsprachlichen Lehrerbildung*. Tübingen: Narr.
- Schönig, W. (2002). Organisationskultur der Schule als Schlüsselkonzept der Schulentwicklung. *Zeitschrift für Pädagogik* 48, 6, S. 815-834.
- Schröder, K. (2007). Kompetenz, Bildungsstandards und Lehrerbildung aus fachdidaktischer Sicht. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung* (S.290-298). Weinheim: Beltz.
- Schuler, H. (1998). *Psychologische Personalauswahl. Einführung in die Berufeignungs-diagnostik* (2. unveränderte Aufl.). Göttingen: Hogrefe.
- Schuler, H., Hell, B. & Trapmann, S. (2007). *Eignungsdiagnostische Auswahl von Studierenden* [Elektronische Version]. Zugriff am 21.04.2010 unter [http://www.studieneignung.de/Abschlussbericht\\_Internetversion.pdf](http://www.studieneignung.de/Abschlussbericht_Internetversion.pdf)
- Schuler, H. & Höft, S. (2004). Diagnose beruflicher Eignung und Leistung. In H. Schuler, *Lehrbuch Organisationspsychologie* (S. 289-345). Bern: Huber.
- Seidel, T., Blomberg, G. & Stürmer, K. (2010). „Observer“ – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. Projekt OBSERVE. In E. Klieme, D. Leutner & M. Kenk, Martina (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (S. 296-306). Weinheim: Beltz.
- Seidel, T. & Prenzel, M. (2007) Erfassung pädagogisch-psychologischer Kompetenzen mit Videosequenzen. *Zeitschrift für Erziehungswissenschaft*, 8, 201-216.
- Seifert, A. Hilligus, A. H. & Schaper, N. (2009). Entwicklung und psychometrische Überprüfung eines Messinstruments zur Erfassung pädagogischer Kompetenzen in der universitären Lehrerbildung. *Lehrerbildung auf dem Prüfstand*, 3, 82-103. Landau: Verlag Empirische Pädagogik.
- Sekretariat der ständigen Konferenz der Kultusminister der Länder (Hrsg.). (2004). *Standards für die Lehrerbildung. Bericht der Arbeitsgruppe* [Elektronische Version]. Zugriff am 14.08.2012 unter



[http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2004/2004\\_12\\_16-Standards\\_Lehrerbildung-Bericht\\_der\\_AG.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards_Lehrerbildung-Bericht_der_AG.pdf)

- Shohamy, E. (1996). Competence and performance in language testing. In G. Bronn, K. Memkjer, J. Williams (Eds.), *Performance and Competence in Second Language Acquisition*, 136-151.
- Shulman, L. S. (1986). Paradigms and research programs in the study of teaching. A contemporary perspective. In M.C. Wittrock (Ed.) (3<sup>d</sup> Ed.), *Handbook of Research on Teaching* (S. 3-36). New York: Macmillan.
- Simonton, D. K. (2003). Expertise, Competence, and Creative Ability: The Perplexing Complexities. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The Psychology of Abilities, Competencies, and Expertise* (pp. 213-239). New York: Cambridge University Press.
- Slavin, R. E. (1989) (Ed.). *School and classroom organization*. Hillsdale, NJ: Erlbaum.
- Slavin, R. E. (1997). *Educational Psychology* (5<sup>th</sup> ed.). Needham Heights, MA: Allyn and Bacon.
- Spinath, B. (2004). Diagnostische Kompetenzen von Lehrerinnen und Lehrern. *Lernende Schule*, 26, 16-17.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer/innen und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19, 85-95
- Stadler, H. (2005). Videoanalysen – ein Weg zur Verbesserung von Unterrichtsqualität. *IMST-Newsletter*, 4 (14), 6-7.
- Steinert, B., Klieme, E., Maag Merki, K., Döbrich, P., Halbheer, U. & Kunz, A. (2006). Lehrerkooperation in der Schule: Konzeption, Erfassung, Ergebnisse. *Zeitschrift für Pädagogik* 52, 2, S. 185-204.
- Stemler, S. E. & Sternberg, R. (2006). Using Situational Judgment Tests to measure practical intelligence. In J. A. Weekley & R. E. Ployhart (2006) (Eds.), *Situational Judgement tests. Theory, measurement and application* (pp. 107-131). Mahwah, N.J.: Lawrence Erlbaum.
- Stevens, M. A. & Campion, M. J. (1999). Staffing work teams: Development and validation of a selection test for teamworking settings. *Journal of Management*, 25, 207-208.
- Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S. & Serrano, A. (1999). *The TIMSS videotape classroom study : methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan,*

- and the United States ; a research and development report.* Washington: U.S. Government Printing Office.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257-285.
- Tent, L. (2006). Zensuren. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 873-879). Weinheim: Beltz.
- Terhart, E. (2000). Lehrerbildung und Professionalität. Strukturen, Probleme und aktuelle Reformtendenzen. In J. Bastian, W. Helsper, S. Reh & C. Schelle (Hrsg.), *Professionalisierung im Lehrerberuf* (S. 73-86). Opladen: Leske & Budrich.
- Terhart, E. (2002). *Standards für die Lehrerbildung. Eine Expertise für die Kultusministerkonferenz*. Münster: Institut für Schulpädagogik und Allgemeine Didaktik, Westfälische Wilhelms-Universität.
- Terhart, E. (2005). Standards für die Lehrerbildung - ein Kommentar. *Zeitschrift für Pädagogik*, 51 (2), S. 275-279.
- Terhart, E. (2007). Erfassung und Beurteilung der beruflichen Kompetenz von Lehrkräften. In M. Lüders & J. Wissinger (Hrsg.), *Forschung zur Lehrerbildung. Kompetenzentwicklung und Programmevaluation* (S. 37-62). Münster: Waxmann.
- Terhart, E. (2009). Erste Phase: Lehrerbildung an der Universität. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (S. 425-437). Weinheim und Basel: Beltz.
- Thorndike, R. L. & Stein, S. (1937). An evaluation of the attempts to measure social intelligence. *Psychological Bulletin*, 34, 275-285.
- Tippelt, R. (2008). Kerncurricula in der Erziehungswissenschaft: Verbindlichkeit, Forschungsbezug und Transparenz – Zur Einführung in die Empfehlungen. In DGfE, *Kerncurriculum Erziehungswissenschaft. Empfehlungen der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE)* (S. 6-12). Opladen: Barbara Budrich.
- Tousignant, M. & DesMarchais, J. E. (2002). Accuracy of Student Self Assessment Ability Compared to Their Own Performance in a Problem-Based Learning Medical Program: A Correlation Study. *Advances in Health Sciences Education*, 7, 19-27.
- Tulodziecki, G. & Grafe, S. (2006). Stellenwert und Kritik von Standards für die Lehrerbildung aus internationaler Sicht. Vergleiche und Einschätzungen zur Situation. *Journal für LehrerInnenbildung* 6, 1, S. 34-44

- Voss, T. & Kunter, M. (2011). Pädagogisch-psychologisches Wissen von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 193-214). Münster: Waxmann.
- Wagner, R. K. & Sternberg, R. J. (1985). Practical intelligence in real-world pursuits: The role of tacit knowledge. *Journal of Personality and Social Psychology*, *48*, 436-458.
- Wang, M.C., Haertel, L.G.D. & Walberg, H.J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, *63*, 249-294.
- Weekley, J. A. & Jones, C. (1999). Further studies of situational tests. *Personell Psychology*, *50*, 25-49.
- Weekley, J. A. & Ployhart, R. E. & Holtz, B. C. (2006). On the Development of Situational Judgement Tests: Issues in Item Development, Scaling, and Scoring. In J. A. Weekley & R. E. Ployhart, (Eds.). *Situational Judgement tests. Theory, measurement and application* (pp. 157-182). Mahwah, N.J.: Lawrence Erlbaum.
- Weekley, J. A. & Ployhart, R. E. (2005), Situational judgment: Antecedents and relationships with performance, *Human Performance*, *18*, 81-104.
- Weekley, J. A. & Ployhart, R. E. (2006). An introduction to situational judgment testing. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational Judgement tests. Theory, measurement and application* (pp. 1-12). Mahwah, N.J.: Lawrence Erlbaum.
- Weinert, F.E. (2000). *Lehren und Lernen für die Zukunft – Ansprüche an das Lernen in der Schule*. Pädagogische Nachrichten Rheinland-Pfalz, *2*, 1-16.
- Weinert, F. E. (2001). Concepts of Competence: A Conceptual Clarification. In D. S. Rychen & L. Hersh Salganik (Eds.), *Defining and Selecting Competence* (pp. 45-65). Seattle, Göttingen: Hogrefe & Huber.
- Weinert, F. E. & Helmke, A. (1984). *Zwischenbericht für das DFG-Projekt „Unterrichtsqualität und Leistungszuwachs bei Formen direkter Instruktion im Mathematikunterricht fünfter Hauptschulklassen.“* München: Max-Planck-Institut für psychologische Forschung.
- Weinert, F. E. & Schrader, F.-W. (1986). Diagnose des Lehrers als Diagnostiker. In H. Petillon, J. W. L. Wagner & B. Wolf (Hrsg.), *SchülerInnengerechte Diagnose* (S. 11-29). Weinheim: Beltz.

- Weinstein, C. E. & Mayer, R. F. (1986). The Teaching of Learning Strategies. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (S. 315-327). New York: Macmillan.
- Weresch-Deperrois, I., Bodensohn, R. & Jäger, R. S. (2009). Curriculare Standards in der Praxis: Einschätzung ihres Stellenwerts, ihrer Anwendungshäufigkeit, Schwierigkeit und Bedeutung in der Lehrerbildung und universitären Vorbereitung im Bachelor-Studium der Lehrerbildung - eine Erkundungsstudie. *Lehrerbildung auf dem Prüfstand*, 2 (2), 324-345.
- Wild, E. (2003). Einbeziehung des Elternhauses durch Lehrer: Art, Ausmaß und Bedingungen der Elternpartizipation aus der Sicht von Gymnasiallehrern. *Zeitschrift für Pädagogik* 49 (2003) 4, 513-533.
- Wild, E., Hofer, M. & Pekrun, R. (2001). Psychologie des Lerners. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (S. 207-269). Weinheim: Beltz.
- Wild, K.-P. & Krapp, A. (2001). Pädagogisch-psychologische Diagnostik. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (S. 513-564). Weinheim: Beltz.
- Wilson, S. M. & Youngs, P. (2005). Research on Accountability Processes in Teacher Education. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying Teacher Education. The Report of the AERA Panel on Research and Teacher Education* (pp. 591-643). Mahwah: Erlbaum.
- Wuttke, E. (2009). Zum Einfluss der professionellen Lehrkompetenz auf die Konstruktion und Steuerung von Lehr-Lern-Umgebungen und -Prozessen. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (S. 669-678). Weinheim und Basel: Beltz.
- Zinnbauer, M. & Eberl, M. (2004). *Die Überprüfung von Spezifikation und Güte von Strukturgleichungsmodellen: Verfahren und Anwendung* [Elektronische Version]. Zugriff am 01.08.2010 unter [http://www.imm.bwl.uni-muenchen.de/forschung/schriftenefo/ap\\_efoplan\\_21.pdf](http://www.imm.bwl.uni-muenchen.de/forschung/schriftenefo/ap_efoplan_21.pdf)

## **Anhang**

**Anhang A: Verzeichnis der Abbildungen**

**Anhang B: Verzeichnis der Tabellen**

**Anhang C: Die Standards für die Lehrerbildung: Bildungswissenschaften  
(KMK, 2004)**

**Anhang D: Als Basis fungierende Items des Instrumentes KOSTA sowie deren  
Trennschärfen**

**Anhang E: Das Instrument zur Gewinnung der Basis negativ gepolter Items**

**Anhang F: Interrater-Reliabilität und deskriptive Ergebnisse der qualitativen  
Datenerhebung zur Gewinnung der Basis negativ gepolter Items**

**Anhang G: Eignungseinschätzung der 102 Items des Situational Judgement  
Tests durch 6 ExpertInnen**

**Anhang H: Der Situational Judgement Test**

**Anhang I: Mittelwerte, Standardabweichungen, Trennschärfen sowie  
Schwierigkeitsindizes der Items des Situational Judgement Tests**

**Anhang J: Zuordnung der Items zu Kompetenzen durch 8 ExpertInnen**

## **Anhang A: Verzeichnis der Abbildungen**

<i>Abbildung 1:</i> Wissen und Können innerhalb der KMK-Kompetenz "Unterrichtsplanung und -durchführung" (Beispiel).....	50
<i>Abbildung 2:</i> Das Wirkmodell der Lehrerbildung nach Galluzzo und Craig (1990), eigene Übersetzung.....	74
<i>Abbildung 3:</i> Beispielitem eines Situational Judgement Tests – Forced-Choice- Format (Weekley & Ployhart, 2006, S. 3; Übersetzung durch den Verfasser) .....	77
<i>Abbildung 4:</i> Beispielitem eines Situational Judgement Tests – Likert-Skala (Weekley & Ployhart, 2006, S. 3; Übersetzung durch die Verfasserin) .....	78
<i>Abbildung 5:</i> Aufgabe EW20 (Seifert, Hilligus & Schaper, 2009, S. 93).....	93
<i>Abbildung 6:</i> Beispiel einer Unterrichtsvignette (Riese & Reinhold, 2008, S. 634).....	95
<i>Abbildung 7:</i> Itembeispiele aus dem COACTIV-Test zur Erfassung fachdidaktischen Wissens (Krauss et al., 2011, S. 140) .....	96
<i>Abbildung 8:</i> Beispielitem KOSTA (Weresch-Deperrois, Bodensohn & Jäger, 2009) .....	101
<i>Abbildung 9:</i> Schritt 2 der Itemgenerierung: Umwandlung gewählter Items in ein qualitatives Frageformat .....	103
<i>Abbildung 10:</i> Beispielitem und –antwort aus dem Fragebogen „Qualitative Erfassung professioneller pädagogischer Kompetenzen Lehramtsstudierender“ der Vorstudie.....	104
<i>Abbildung 11:</i> Beispielitems „Problemorientiertes Lernen und Lehren“.....	125
<i>Abbildung 12:</i> Schritt 3 der Itemgenerierung: Entwicklung positiv und negativ gepolter Situationen.....	126
<i>Abbildung 13:</i> Beispielitem der ExpertInnen-Befragung „Simulationsorientierte Erfassung professioneller Kompetenzen im Lehrberuf“/positiver Pol .....	127
<i>Abbildung 14:</i> Beispielitem der ExpertInnen-Befragung „Simulationsorientierte Erfassung professioneller Kompetenzen im Lehrberuf“/negativer Pol.....	128
<i>Abbildung 15:</i> Vollständiger Prozess der Itemgenerierung.....	129

*Abbildung 16:* Beispielhafte Darstellung eines Q-Q-Plots anhand der Variablen

„Auswahl von Methoden und Medien“ ..... 137

*Abbildung 17:* Beispielhafte Darstellung eines Histogramms anhand der Variablen

„Beeinflussung von Lernprozessen durch Beurteilung +“ ..... 138

## Anhang B: Verzeichnis der Tabellen

Tabelle 1. Vergleich der Anforderungen gemäß MT21 (Blömeke et al., 2008) mit den Standards für die Lehrerbildung: Bildungswissenschaften (KMK, 2004).....	56
Tabelle 2. Überblick über Entscheidungsalternativen bei der Entwicklung von Situational Judgement Tests (Weekley, Ployhart & Holtz, 2006, S. 158, eigene Übersetzung) .....	119
Tabelle 3. Ranges der Itemkennwerte.....	140
Tabelle 4. Cronbachs Alpha der elf Subskalen des Instrumentes .....	145
Tabelle 5. Ergebnisse der multiplen Regression zur Vorhersage der Testleistungen	149
Tabelle 6. Ergebnisse der binären logistischen Regression zur Vorhersage dichotomisierter Testleistungen .....	153
Tabelle 7. Ergebnisse der t-Tests für unabhängige Stichproben gemeinsam mit den entsprechenden Effektstärken (** $p = .000$ , * $p < .05$ ) .....	158
Tabelle 8. Häufigkeit der Nennung von Kompetenzbereichen durch ExpertInnen bei der Itemzuordnung.....	161
Tabelle 9. Die Kriterien reflektiver und formativer Messungen nach Jarvis, MacKenzie und Podsakoff (2003) .....	167



**Anhang C: Die Standards für die Lehrerbildung: Bildungswissenschaften  
(KMK, 2004)**

**Kompetenzbereich: Unterrichten**

**Lehrerinnen und Lehrer sind Fachleute für das Lehren und Lernen.**

**Kompetenz 1:**

Lehrerinnen und Lehrer planen Unterricht fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.

Standards für die theoretischen Ausbildungsabschnitte	Standards für die praktischen Ausbildungsabschnitte
<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- kennen die einschlägigen Bildungstheorien, verstehen bildungs- und erziehungstheoretische Ziele sowie die daraus abzuleitenden Standards und reflektieren diese kritisch.</li> <li>- kennen allgemeine und fachbezogene Didaktiken und wissen, was bei der Planung von Unterrichtseinheiten beachtet werden muss.</li> <li>- kennen unterschiedliche Unterrichtsmethoden und Aufgabenformen und wissen, wie man sie anforderungs- und situationsgerecht einsetzt.</li> <li>- kennen Konzepte der Medienpädagogik und –psychologie und Möglichkeiten und Grenzen eines anforderungs- und situationsgerechten Einsatzes von Medien im Unterricht.</li> <li>- kennen Verfahren für die Beurteilung von Lehrleistung und Unterrichtsqualität.</li> </ul>	<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- verknüpfen fachwissenschaftliche und fachdidaktische Argumente und planen und gestalten Unterricht.</li> <li>- wählen Inhalte und Methoden, Arbeits- und Kommunikationsformen aus.</li> <li>- integrieren moderne Informations- und Kommunikationstechnologien didaktisch sinnvoll und reflektieren den eigenen Medieneinsatz.</li> <li>- überprüfen die Qualität des eigenen Lernens.</li> </ul>

**Kompetenzbereich: Unterrichten**  
**Lehrerinnen und Lehrer sind Fachleute für das Lehren und Lernen.**

<b>Kompetenz 2:</b> Lehrerinnen und Lehrer unterstützen durch die Gestaltung von Lernsituationen das Lernen von Schülerinnen und Schülern. Sie motivieren Schüler und befähigen sie, Zusammenhänge herzustellen und Gelerntes zu verknüpfen.	
<b>Standards für die theoretischen Ausbildungsabschnitte</b>	<b>Standards für die praktischen Ausbildungsabschnitte</b>
Die Absolventinnen und Absolventen... <ul style="list-style-type: none"> <li>- kennen Lerntheorien und Formen des Lernens.</li> <li>- wissen, wie man Lernende aktiv in den Unterricht einbezieht und Verstehen und Transfer unterstützt.</li> <li>- kennen Theorien der Lern- und Leistungsmotivation und Möglichkeiten, wie sie im Unterricht angewendet werden.</li> </ul>	Die Absolventinnen und Absolventen... <ul style="list-style-type: none"> <li>- regen unterschiedliche Formen des Lernens an und unterstützen diese.</li> <li>- gestalten Lehr-Lernprozesse unter Berücksichtigung der Erkenntnisse über den Erwerb von Wissen und Fähigkeiten.</li> <li>- wecken und stärken bei Schülerinnen und Schülern Lern- und Leistungsbereitschaft.</li> <li>- führen und begleiten Lerngruppen.</li> </ul>

**Kompetenzbereich: Unterrichten**  
**Lehrerinnen und Lehrer sind Fachleute für das Lehren und Lernen.**

<b>Kompetenz 3:</b> Lehrerinnen und Lehrer fördern die Fähigkeiten von Schülerinnen und Schülern zum selbstbestimmten Lernen und Arbeiten.	
<b>Standards für die theoretischen Ausbildungsabschnitte</b>	<b>Standards für die praktischen Ausbildungsabschnitte</b>
Die Absolventinnen und Absolventen... <ul style="list-style-type: none"> <li>- kennen Lern- und Selbstmotivationsstrategien, die sich positiv auf Lernerfolg und Arbeitsergebnisse auswirken.</li> <li>- kennen Methoden der Förderung selbstbestimmten, eigenverantwortlichen und kooperativen Lernens und Arbeitens.</li> <li>- wissen, wie sie weiterführendes Interesse und Grundlagen des lebenslangen Lernens im Unterricht entwickeln.</li> </ul>	Die Absolventinnen und Absolventen... <ul style="list-style-type: none"> <li>- vermitteln und fördern Lern- und Arbeitsstrategien.</li> <li>- vermitteln den Schülerinnen und Schülern Methoden des selbstbestimmten, eigenverantwortlichen und kooperativen Lernens und Arbeitens.</li> </ul>

**Kompetenzbereich: Erziehen****Lehrerinnen und Lehrer üben ihre Erziehungsaufgabe aus.****Kompetenz 4:**

Lehrerinnen und Lehrer kennen die sozialen und kulturellen Lebensbedingungen von Schülerinnen und Schülern und nehmen im Rahmen der Schule Einfluss auf deren individuelle Entwicklung.

Standards für die theoretischen Ausbildungsabschnitte	Standards für die praktischen Ausbildungsabschnitte
<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- kennen pädagogische, soziologische und psychologische Theorien der Entwicklung und der Sozialisation von Kindern und Jugendlichen.</li> <li>- kennen etwaige Benachteiligungen von Schülerinnen und Schülern beim Lernprozess und Möglichkeiten der pädagogischen Hilfen und Präventivmaßnahmen.</li> <li>- kennen interkulturelle Dimensionen bei der Gestaltung von Bildungs- und Erziehungsprozessen.</li> <li>- kennen die Bedeutung geschlechtsspezifischer Einflüsse auf Bildungs- und Erziehungsprozesse.</li> </ul>	<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- erkennen Benachteiligungen und realisieren pädagogische Hilfen und Präventionsmaßnahmen.</li> <li>- unterstützen individuell</li> <li>- beachten die kulturelle und soziale Vielfalt der Lerngruppe.</li> </ul>

**Kompetenzbereich: Erziehen****Lehrerinnen und Lehrer üben ihre Erziehungsaufgabe aus.****Kompetenz 5:**

Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern.

Standards für die theoretischen Ausbildungsabschnitte	Standards für die praktischen Ausbildungsabschnitte
<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- kennen und reflektieren demokratische Werte und Normen sowie ihre Vermittlung.</li> <li>- wissen, wie man wertbewusste Haltungen und selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern fördert.</li> <li>- wissen, wie Schülerinnen und Schüler im Umgang mit persönlichen Krisen- und Entscheidungssituationen unterstützt werden.</li> </ul>	<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- reflektieren Werte und Werthaltungen und handeln entsprechend.</li> <li>- üben mit den Schülerinnen und Schülern eigenverantwortliches Urteilen und Handeln schrittweise ein.</li> <li>- setzen Formen des konstruktiven Umgangs mit Normkonflikten ein.</li> </ul>

**Kompetenzbereich: Erziehen****Lehrerinnen und Lehrer üben ihre Erziehungsaufgabe aus.****Kompetenz 6:**

Lehrerinnen und Lehrer finden Lösungsansätze für Schwierigkeiten und Konflikte in Schule und Unterricht.

Standards für die theoretischen Ausbildungsabschnitte	Standards für die praktischen Ausbildungsabschnitte
<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- verfügen Kenntnisse zu Kommunikation und Interaktion (unter besonderer Berücksichtigung der Lehrer-Schüler-Interaktion).</li> <li>- kennen Regeln der Gesprächsführung sowie Grundsätze des Umgangs miteinander, die in Unterricht, Schule und Elternarbeit bedeutsam sind.</li> <li>- kennen Risiken und Gefährdungen des Kindes- und Jugendalters sowie Präventions- und Interventionsmöglichkeiten.</li> <li>- analysieren Konflikte und kennen Methoden der konstruktiven Konfliktbearbeitung und des Umgangs mit Gewalt.</li> </ul>	<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- gestalten soziale Beziehungen und soziale Lernprozesse in Unterricht und Schule.</li> <li>- erarbeiten mit den Schülerinnen und Schülern Regeln des Umgangs miteinander und setzen sie um.</li> <li>- wenden im konkreten Fall Strategien und Handlungsformen der Konfliktprävention und -lösung an.</li> </ul>

**Kompetenzbereich: Beurteilen****Lehrerinnen und Lehrer üben ihre Beurteilungsaufgabe gerecht und verantwortungsbewusst aus.**

## Kompetenz 7:

Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern.

Standards für die theoretischen Ausbildungsabschnitte	Standards für die praktischen Ausbildungsabschnitte
<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- wissen, wie unterschiedliche Lernvoraussetzungen Lehren und Lernen beeinflussen und wie sie im Unterricht berücksichtigt werden.</li> <li>- kennen Formen von Hoch- und Sonderbegabung, Lern- und Arbeitsstörungen.</li> <li>- kennen die Grundlagen der Lernprozessdiagnostik.</li> <li>- kennen Prinzipien und Ansätze der Beratung von Schülerinnen/Schülern und Eltern.</li> </ul>	<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- erkennen Entwicklungsstände, Lernpotentiale, Lernhindernisse und Lernfortschritte.</li> <li>- erkennen Lernausgangslagen und setzen spezielle Fördermöglichkeiten ein.</li> <li>- erkennen Begabungen und kennen Möglichkeiten der Begabungsförderung.</li> <li>- stimmen Lernmöglichkeiten und Lernanforderungen aufeinander ab.</li> <li>- setzen unterschiedliche Beratungsformen situationsgerecht ein und unterscheiden Beratungsfunktion und Beurteilungsfunktion.</li> <li>- kooperieren mit Kolleginnen und Kollegen bei der Erarbeitung von Beratung/Empfehlung.</li> <li>- kooperieren mit anderen Institutionen bei der Entwicklung von Beratungsangeboten.</li> </ul>

**Kompetenzbereich: Beurteilen**  
**Lehrerinnen und Lehrer üben ihre Beurteilungsaufgabe gerecht und verantwortungsbewusst aus.**

Kompetenz 8:

Lehrerinnen und Lehrer erfassen Leistungen von Schülerinnen und Schülern auf der Grundlage transparenter Beurteilungsmaßstäbe.

Standards für die theoretischen Ausbildungsabschnitte	Standards für die praktischen Ausbildungsabschnitte
<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- kennen unterschiedliche Formen der Leistungsbeurteilung, ihre Funktionen und ihre Vor- und Nachteile.</li> <li>- kennen verschiedene Bezugssysteme der Leistungsbeurteilung und wägen sie gegeneinander ab.</li> <li>- kennen Prinzipien der Rückmeldung von Leistungsbeurteilung.</li> </ul>	<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- konzipieren Aufgabenstellungen kriteriengerecht und formulieren sie adressatengerecht.</li> <li>- wenden Bewertungsmodelle und Bewertungsmaßstäbe fach- und situationsgerecht an.</li> <li>- verständigen sich auf Beurteilungsgrundsätze mit Kolleginnen und Kollegen begründen Bewertungen und Beurteilungen adressatengerecht und zeigen Perspektiven für das weitere Lernen auf.</li> <li>- nutzen Leistungsüberprüfungen als konstruktive Rückmeldung über die eigene Unterrichtstätigkeit.</li> </ul>

**Kompetenzbereich: Innovieren**  
**Lehrerinnen und Lehrer entwickeln ihre Kompetenzen ständig weiter.**

Kompetenz 9:

Lehrerinnen und Lehrer sind sich der besonderen Anforderungen des Lehrerberufs bewusst. Sie verstehen ihren Beruf als ein öffentliches Amt mit besonderer Verantwortung und Verpflichtung.

Standards für die theoretischen Ausbildungsabschnitte	Standards für die praktischen Ausbildungsabschnitte
<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- kennen die Grundlagen und Strukturen des Bildungssystems und von Schule als Organisation.</li> <li>- kennen die rechtlichen Rahmenbedingungen ihrer Tätigkeit (z. B. Grundgesetz, Schulgesetze).</li> <li>- reflektieren ihre persönlichen berufsbezogenen Wertvorstellungen und Einstellungen.</li> <li>- kennen wesentliche Ergebnisse der Belastungs- und Stressforschung.</li> </ul>	<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- lernen, mit Belastungen umzugehen.</li> <li>- setzen Arbeitszeit und Arbeitsmittel zweckdienlich und ökonomisch ein.</li> <li>- praktizieren kollegiale Beratung als Hilfe zur Unterrichtsentwicklung und Arbeitsentlastung.</li> </ul>

**Kompetenzbereich: Innovieren**  
**Lehrerinnen und Lehrer entwickeln ihre Kompetenzen ständig weiter.**

**Kompetenz 10:**

Lehrerinnen und Lehrer verstehen ihren Beruf als ständige Lernaufgabe.

Standards für die theoretischen Ausbildungsabschnitte	Standards für die praktischen Ausbildungsabschnitte
<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- kennen Methoden der Selbst- und Fremdevaluation.</li> <li>- rezipieren und bewerten Ergebnisse der Bildungsforschung.</li> <li>- kennen organisatorische Bedingungen und Kooperationsstrukturen an Schulen.</li> </ul>	<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- reflektieren die eigenen beruflichen Erfahrungen und Kompetenzen und deren Entwicklung und können hieraus Konsequenzen ziehen.</li> <li>- nutzen Ergebnisse der Bildungsforschung für die eigene Tätigkeit.</li> <li>- dokumentieren für sich und andere die eigene Arbeit und ihre Ergebnisse.</li> <li>- geben Rückmeldungen und nutzen die Rückmeldungen anderer dazu, ihre pädagogische Arbeit zu optimieren.</li> <li>- nehmen Mitwirkungsmöglichkeiten wahr.</li> <li>- kennen und nutzen Unterstützungsmöglichkeiten für Lehrkräfte.</li> <li>- nutzen formelle und informelle, individuelle und kooperative Weiterbildungsangebote.</li> </ul>

**Kompetenzbereich: Innovieren****Lehrerinnen und Lehrer entwickeln ihre Kompetenzen ständig weiter.**

## Kompetenz 11:

Lehrerinnen und Lehrer beteiligen sich an der Planung und Umsetzung schulischer Projekte und Vorhaben.

Standards für die theoretischen Ausbildungsabschnitte	Standards für die praktischen Ausbildungsabschnitte
<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- kennen und reflektieren den spezifischen Bildungsauftrag einzelner Schularten, Schulformen und Bildungsgänge.</li> <li>- kennen Ziele und Methoden der Schulentwicklung</li> <li>- kennen die Bedingungen für erfolgreiche Kooperation.</li> </ul>	<p>Die Absolventinnen und Absolventen...</p> <ul style="list-style-type: none"> <li>- wenden Ergebnisse der Unterrichts- und Bildungsforschung auf die Schulentwicklung an.</li> <li>- nutzen Verfahren und Instrumente der internen Evaluation von Unterricht und Schule.</li> <li>- planen schulische Projekte und Vorhaben kooperativ und setzen sie um.</li> <li>- unterstützen eine Gruppe darin, gute Arbeitsergebnisse zu erreichen.</li> </ul>



## **Anhang D: Als Basis fungierende Items des Instrumentes KOSTA sowie deren Trennschärfen**

---

### **Kompetenzbereich 1: Lehrerinnen und Lehrer planen Unterricht fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.**

---

1. Ich strukturiere meinen Unterricht klar. (struk01as-es)
  2. Ich setze verschiedene Methoden anforderungs- und situationsgerecht ein. (met02as-es)/Ich setze verschiedene Medien anforderungs- und situationsgerecht ein. (med03as-es)
  3. Ich plane meinen Unterricht auf der Basis fachwissenschaftlichen und fachdidaktischen Wissens. (faw04as-es)
  4. Ich gehe auf Lern- und Verständnisprobleme der Schüler flexibel ein. (ler05as-es)
- 

### **Kompetenzbereich 2: Lehrerinnen und Lehrer unterstützen durch die Gestaltung von Lernsituationen das Lernen von Schülerinnen und Schülern. Sie motivieren Schüler und befähigen sie, Zusammenhänge herzustellen und Gelerntes zu verknüpfen.**

---

5. Ich unterrichte so, das die Schüler Transferleistungen erbringen müssen. (tra07as-es)
  6. Ich gestalte Lernumgebungen so, dass problemorientiertes Lernen möglich ist. (lern08as-es)
  7. Ich berücksichtige Erkenntnisse über den Erwerb von Wissen und Fertigkeiten für meine Unterrichtsgestaltung. (erw09as-es)
  8. Ich nutze Fehler der Schüler, um neue Lernprozesse zu initiieren oder diese fortzuführen. (fehl10as-es)
  9. Ich motiviere die Schüler, indem ich den Stoff auf deren individuelle Lebenswelten beziehe. (ind13as-es)
- 

### **Kompetenzbereich 3: Lehrerinnen und Lehrer fördern die Fähigkeiten von Schülerinnen und Schülern zum selbstbestimmten Lernen und Arbeiten.**

---

10. Ich vermittele und fördere Lern-, Arbeits- und Selbstregulationsstrategien. (stra15as-es)
  11. Ich veranlasse die Schüler, auch selbstständig Neues zu entdecken. (selb16as-es)
  12. Ich führe die Schüler an verschiedene Arbeitstechniken heran. (ate17as-es)
  13. Ich fördere die Erfolgszuversicht der Schüler. (erfo19as-es)
  14. Ich verdeutliche den Zusammenhang zwischen der eigenen Anstrengung der Schüler und deren individuellem Arbeitsergebnis. (zus20as-es)
-

---

**Kompetenzbereich 4: Lehrerinnen und Lehrer kennen die sozialen und kulturellen Lebensbedingungen von Schülerinnen und Schülern und nehmen im Rahmen der Schule Einfluss auf deren individuelle Entwicklung.**

---

- 15. Ich passe den Unterricht individuellen Lernvoraussetzungen an. (le23as-es)
- 16. Ich fördere individuelle Interessen der Schüler. (int25as-es)
- 17. Ich vermindere Vorurteile und Berührungängste, indem ich die Schüler über kulturelle Unterschiede informiere. (vor26as-es)

---

**Kompetenzbereich 5: Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern.**

---

- 18. Ich übe mit den Schülern eigenverantwortliches Urteilen und Handeln ein. (urt27as-es)
- 19. Ich diskutiere mit den Schülern Dilemmata, die sich aus deren Lebenswelten ergeben. (dil29as-es)
- 20. Ich beziehe die Schüler in Entscheidungsprozesse mit ein. (ent30as-es)
- 21. Ich Sorge für ein vertrauensvolles Klima. (kli31as-es)
- 22. Ich fördere hilfsbereites Verhalten. (hil32as-es)

---

**Kompetenzbereich 6: Lehrerinnen und Lehrer finden Lösungsansätze für Schwierigkeiten und Konflikte in Schule und Unterricht.**

---

- 23. Ich thematisiere im Unterricht Aggression und Gewalt sowie deren Folgen. (agg34as-es)
- 24. Ich gestalte soziale Lernprozesse. (sle35as-es)
- 25. Ich löse Konflikte transparent und gemeinsam mit den Schülern. (kon36as-es)
- 26. Ich übe mit den Schülern, in Konfliktsituationen verschiedene Perspektiven einzunehmen. (per37as-es)

---

**Kompetenzbereich 7: Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern.**

---

- 27. Ich erkenne individuelle Lernvoraussetzungen und -fortschritte. (erk38as-es)
  - 28. Ich nutze die durch meine diagnostische Kompetenz gewonnenen Erkenntnisse zur Gestaltung meines Unterrichts. (dia39as-es)
  - 29. Ich kooperiere bei Verhaltens- oder Lernproblemen eines Schülers mit dessen Eltern. (ko40as-es)
  - 30. Ich achte darauf, in Beratungen keine Beurteilung einfließen zu lassen. (be41as-es)
  - 31. Ich nutze zur Leistungsfeststellung verschiedene Methoden. (lei42as-es)
-

---

**Kompetenzbereich 8: Lehrerinnen und Lehrer erfassen Leistungen von Schülerinnen und Schülern auf der Grundlage transparenter Beurteilungsmaßstäbe.**

---

32. Ich beurteile Schüler anhand einer angemessenen Kombination aus sozialer und individueller Bezugsnorm. (bez44as-es)
33. Ich berücksichtige, dass Lehrerbeurteilungen subjektiv sein können. (sub45as-es)
34. Ich beachte, dass Beurteilungen und Erwartungen Lernvorgänge beeinflussen. (lerv46as-es)
35. Ich berücksichtige und vereinbare transparente Beurteilungskriterien. (tran47as-es)
36. Ich beziehe meine Beurteilungen auf formulierte Lernziele. (lerz48as-es)

---

**Kompetenzbereich 9: Lehrerinnen und Lehrer sind sich der besonderen Anforderungen des Lehrerberufs bewusst. Sie verstehen ihren Beruf als ein öffentliches Amt mit besonderer Verantwortung und Verpflichtung.**

---

37. Ich überprüfe meine eigenen Erwartungen und Ansprüche an den Lehrberuf. (er49as-es)
38. Ich handle als Lehrkraft meiner Vorbildfunktion entsprechend. (vo50as-es)
39. Ich nehme Mitwirkungsmöglichkeiten in der Schule wahr. (mit51as-es)
40. Ich überdenke immer wieder meine berufsbezogenen Wertvorstellungen und Einstellungen. (wert52as-es)
41. Ich kenne die rechtlichen Rahmenbedingungen des Lehrberufs. (rah53as-es)

---

**Kompetenzbereich 10: Lehrerinnen und Lehrer verstehen ihren Beruf als ständige Lernaufgabe.**

---

42. Ich nutze auch außerhalb der Universität verschiedene Weiterbildungsangebote. (weit55as-es)
43. Ich nutze Ergebnisse der Bildungsforschung für meine pädagogische Arbeit. (erg56as-es)
44. Ich reflektiere meine beruflichen Erfahrungen und ziehe entsprechende Konsequenzen. (kons57as-es)
45. Ich gebe meinen Kommilitonen oder Kollegen Feedback und nutze auch deren Rückmeldungen, um meine Arbeit zu optimieren. (fee58as-es)
46. Ich überprüfe immer wieder mein Wissen und Können auf seine Aktualität. (akt59as-es)

---

**Kompetenzbereich 11: Lehrerinnen und Lehrer beteiligen sich an der Planung und Umsetzung schulischer Projekte und Vorhaben.**

---

47. Ich wende Ergebnisse der Unterrichts- und Bildungsforschung auf die Planung und Durchführung von Projekten an. (pla61as-es)
48. Ich nutze für Unterricht und Schule Verfahren und Instrumente der Evaluation. (eva62as-es)
49. Ich nutze die Ergebnisse der Evaluation zur Gestaltung von Unterricht und Schule. (gest63as-es)
50. Ich beteilige mich an Projektarbeiten und unterstütze diese. (proj64as-es)
51. Ich plane schulische Projekte zusammen mit Kommilitonen oder Kollegen und führe diese mit ihnen zusammen durch. (prop65as-es)
-

**Trennschärfen der ausgewählten Items des Instrumentes KOSTA**

Kompetenzbereich 1		Kompetenzbereich 2		Kompetenzbereich 3	
Item <sup>a)</sup>	Trennschärfe	Item	Trennschärfe	Item	Trennschärfe
struk01ds	.458	tra07ds	.541	stra15ds	.540
met02ds*	.446	lern08ds	.617	selb16ds	.591
med03ds*	.453	erw09ds	.499	ate17ds	.547
faw04ds	.459	fehl10ds	.540	erfo19ds	.563
ler05ds	.453	ind13ds	.551	zus20ds	.570
Kompetenzbereich 4		Kompetenzbereich 5		Kompetenzbereich 6	
Item	Trennschärfe	Item	Trennschärfe	Item	Trennschärfe
le23ds	.453	urt27ds	.600	agg34ds	.679
int25ds	.523	dil29ds	.629	sle35ds	.570
vor26ds	.569	ent30ds	.724	kon36ds	.671
		kli31ds	.687	per37ds	.663
		hil32ds	.732		
Kompetenzbereich 7		Kompetenzbereich 8		Kompetenzbereich 9	
Item	Trennschärfe	Item	Trennschärfe	Item	Trennschärfe
erk38ds	.659	bez44ds	.647	er49as	.547
dia39ds	.588	sub45ds	.587	vo50as	.595
ko40ds	.587	ler46ds	.656	mit51as	.693
be41ds	.596	tran47ds	.634	wert52as	.642
lei42ds	.540	lerz48ds	.541	rah53as	.357
Kompetenzbereich 10		Kompetenzbereich 11			
Item	Trennschärfe	Item	Trennschärfe		
weit55ds	.491	pla61cs	.649		
erg56ds	.426	eva62cs	.674		
kons57ds	.579	gest63cs	.673		
fee58ds	.569	proj64cs	.664		
akt59ds	.643	prop65cs	.660		

\* Die Items met02ds und med03ds wurden zu einem Item kombiniert.

## **Anhang E: Das Instrument der qualitativen Vorstudie „Qualitative Erfassung professioneller pädagogischer Kompetenzen im Lehrberuf“**

### **Qualitative Erfassung professioneller pädagogischer Kompetenzen Lehramtsstudierender**

Liebe Studierende,

um die Qualität schulischer Bildung überprüfen und sichern zu können, wurden durch die Kultusministerkonferenz verbindliche Standards formuliert, welche sich auf verschiedene Bildungsbereiche in Schule und Hochschule beziehen. Beispielsweise für den Bereich der Lehrerbildung wurden durch die „Standards für die Lehrerbildung: Bildungswissenschaften“ (KMK, 2004) elf Kompetenzbereiche definiert, über welche Lehramtsstudierende (in Abhängigkeit von der Studiendauer) verfügen sollen.

Um das Vorhandensein dieser Kompetenzen überprüfen zu können, müssen valide Erhebungsinstrumente zur Verfügung stehen. Ein solches Instrument befindet sich in Form eines situations- und kontextbezogenen Tests in der Entwicklung.

Sie können durch die Beantwortung des vorliegenden an den Kompetenzbereichen der KMK (2004) bzw. dem Erhebungsinstrument KOSTA<sup>®</sup> (Kompetenz- und Standardorientierung in der Lehrerbildung; Zentrum für Lehrerbildung, Zentrum für empirische pädagogische Forschung, Landau; 2009) orientierten Fragebogens einen wichtigen Beitrag zur Entwicklung des situations- und kontextbezogenen Tests leisten. Ihre Antworten werden einen Überblick über die weite Spanne professioneller Kompetenz im Lehrberuf liefern.

Für Ihre Mithilfe bedankt sich im Voraus herzlich

Ines Weresch-Deperrois, Dipl.-Psych, Dipl.-Mus.

## **Bearbeitungshilfe**

Der vorliegende Fragebogen besteht aus **55 offenen Fragen** zu verschiedenen Aspekten professioneller Kompetenz im Lehrberuf. Zu jeder Frage ist eine Beispielantwort angegeben.

- 1. Bitte beantworten Sie jede Frage kurz stichwortartig oder mit höchstens vier Sätzen.**
- 2. Hierbei orientieren Sie sich unbedingt nicht an einem Ideal, sondern an Ihrem tatsächlichen Handeln.**
- 3. Falls Sie bislang noch keine Klasse unterrichtet haben, geben Sie bitte an, wie Sie handeln würden.**

Dieser Fragebogen ist **kein Test Ihrer individuellen Kompetenzen!** Es soll lediglich ein **Überblick über das Spektrum professioneller Kompetenz im Lehrberuf** geschaffen werden.

Falls Sie an einem **Feedback** interessiert sind, geben Sie bitte am Ende des Fragebogens Ihre e-mail-Adresse an.

**Angaben zum Studienverlauf**  
(Zutreffendes bitte ankreuzen oder angeben)

**Ihre Codenummer:**

erster Buchstabe des Vornamens Ihrer Mutter

erster Buchstabe des Vornamens Ihres Vaters

erste Zahl des eigenen Geburtstags

(z. B.: „H“ für „Heike“, „B“ für „Benjamin“ und „27“ für „27.07.“)

1. Im wievielten Semester studieren Sie im Lehramtsstudium?

--	--

2. In welchem Bundesland studieren Sie? \_\_\_\_\_

3. Ihr Berufsziel ist das Lehramt an...

Grundschule     Realschule     Gymnasium     Förderschule   
 Berufsschule

Ich habe mich **noch nicht** für eine Schulart entschieden.

4. Sie sind...                      weiblich                       männlich

5. An welchen Praktika haben Sie bereits teilgenommen?

Orientierendes Praktikum            1     2     3

Vertiefendes Praktikum                1     2

Bislang habe ich **noch nicht** an einem Praktikum teilgenommen.

## Fragenkatalog

1
Wie würden Sie den Ablauf einer typischen durch Sie gehaltenen Unterrichtsstunde beschreiben?
Bsp.: Ich frage die Hausaufgaben ab, wiederhole den Stoff der letzten Stunde kurz und schaue dann, wie schnell wir vorankommen.

2
Nach welchen Überlegungen wählen Sie die Unterrichtsmethoden für Ihre Klassen aus?
Bsp.: Ich wähle die Methoden, mit denen ich die meisten Erfahrungen sammeln konnte.

3
Wie nutzen Sie Evaluationsergebnisse (z. B. einen Test des Sprachgebrauchs von Viertklässlern) zur Gestaltung Ihres Unterrichts?
Bsp.: Falls die Ergebnisse schlecht ausfallen, werde ich mich im Unterricht mehr auf diesen Bereich konzentrieren. Falls sie sehr gut sind, werde ich die Schwierigkeit erhöhen.

4
Wie reagieren Sie, wenn eine im Unterricht von Ihnen angewendete Strategie (z. B. eine bestimmte Methode, Umgang mit SchülerInnen) kaum Erfolg zeigt?
Bsp.: Nicht alle Strategien sind erfolgreich – damit müssen Lehrer leben.

5
In welchen Situationen handeln Sie als Lehrkraft Ihrer Vorbildfunktion entsprechend (Beispiele)?
Bsp.: Ich handle meinen SchülerInnen gegenüber immer respektvoll.



6
Aufgrund welcher Normen beurteilen Sie Ihre SchülerInnen?
Bsp.: Ich halte mich an die soziale Bezugsnorm und vergleiche die Einzelleistung mit der Klassenleistung.

7
Wie gehen Sie bei längerfristigen Verhaltens- oder Lernproblemen einer/s SchülerIn vor?
Bsp.: Ich vereinbare ein Elterngespräch.

8
Wie vermeiden Sie, dass Beurteilungen und Erwartungen Lernvorgänge negativ beeinflussen?
Bsp.: Ich versuche, den SchülerInnen nicht zu zeigen, wenn sie mir nicht so sympathisch sind.

9
Welche Arbeitstechniken lassen Sie Ihre SchülerInnen im Unterricht anwenden?
Bsp.: Ich lasse die SchülerInnen alleine und in Gruppen arbeiten.

10
Welche SchülerInnen fördern Sie?
Bsp.: Ich fördere besonders schwache SchülerInnen.

11
Welche Prinzipien des Wissenserwerbs beachten Sie bei der Unterrichtsgestaltung?
Bsp.: Modellernen - Ich zeige den SchülerInnen mehrmals, wie sie eine Aufgabe lösen sollen.

12
Wie unterstützen Sie Ihre SchülerInnen beim Lernen (z. B. auf eine Klassenarbeit)?
Bsp.: Ich weise sie darauf hin, dass regelmäßiges Lernen wichtig ist.

13
Wie reagieren Sie auf in Ihrem Unterricht aufkommende Lern- und Verständnisprobleme einzelner SchülerInnen?
Bsp.: Ich führe den Unterricht fort und spreche in der Pause mit den SchülerInnen über ihre Lern- und Verständnisprobleme.

14
Welche Methoden der Leistungsfeststellung nutzen Sie?
Bsp.: Klassenarbeiten, mündliche und schriftliche Tests.

15
Welche Werte sind Ihnen im Rahmen des Lehrberufs besonders wichtig?
Bsp.: Die Leistung meiner Klasse.

16
Wie häufig überprüfen Sie Ihr Wissen und Können auf Aktualität?
Bsp.: Der Wissensstand in meinem Fach ändert sich nicht so schnell, einige Jahre ist mein Wissen noch aktuell.

17
Welche schulischen Projekte haben Sie mit Kommilitonen/Kollegen geplant/an welchen Projekten haben Sie teilgenommen?
Bsp.: Ich habe beim Sommerfest die Aufsicht geführt.

18
Nach welchen Überlegungen wählen Sie die <b>Unterrichtsmedien</b> für Ihre Klassen aus?
Bsp.: Ich stimme die Medien auf den Entwicklungsstand der Klasse ab.

19
Sie bemerken das besondere Interesse eines/r Schülers/Schülerin an einem Themenbereich. Wie reagieren Sie hierauf?
Bsp.: Ich unterhalte mich mit ihm/ihr über das Thema.

20
Wie unterstützen Sie eigenverantwortliches Urteilen und Handeln Ihrer SchülerInnen?
Bsp.: Wir machen Rollenspiele und sprechen über das Erlebte.

21
Wie sorgen Sie für einen respektvollen Umgang innerhalb Ihrer Klasse?
Bsp.: Ich entwerfe einen Regelkatalog.

22
Auf Basis welcher Informationen planen Sie schulische Projekte?
Bsp.: Ich wende bei der Planung die Ergebnisse der Bildungsforschung an (z. B. Erkenntnisse über Wissenserwerb).

23
Wie berücksichtigen Sie, dass Lehrerbeurteilungen subjektiv sein können?
Bsp.: Ich vergleiche die Leistungen meiner Klasse mit denen einer Parallelklasse.

24
Wie werden Entscheidungen getroffen, die Ihre Klasse betreffen (z. B. bei der Wahl eines Ausflugszieles)?
Bsp.: Ich befrage die SchülerInnen nach ihrer Meinung und entscheide dann.

25
Wie lösen Sie Konflikte innerhalb Ihrer Klasse (z. B. bei rivalisierenden Cliques)?
Bsp.: Ich spreche Sanktionen aus.

26
Wie erreichen Sie Transferleistungen Ihrer SchülerInnen (Beispiele)?
Bsp.: Ich erzähle den SchülerInnen, in welchem alltäglichen Rahmen der behandelte Stoff nützlich ist.

27
Wie unterstützen Sie SchülerInnen bei Projektarbeiten?
Bsp.: Ich ermutige die SchülerInnen „dranzubleiben“.

28
Nach welchen Kriterien wählen Sie Förderangebote für Ihre SchülerInnen aus?
Bsp.: Ich fördere SchülerInnen, die Lernschwierigkeiten haben.

29
Welche Mitwirkungsmöglichkeiten nehmen Sie an Ihrer Schule wahr?
Bsp.: Ich nehme an Lehrerkonferenzen teil.

30
Wie erreichen Sie, dass Ihre SchülerInnen Ihre Beurteilungen nachvollziehen können?
Bsp.: Ich erkläre z. B. vor einer Klassenarbeit, auf was ich besonders achten werde.

31
Wie motivieren Sie Ihre SchülerInnen im Unterricht?
Bsp.: Ich lasse sie Beispiele für den Stoff aus Ihrem Alltag finden.

32
Wie reagieren Sie auf Vorurteile und Berührungsängste Ihrer Klasse hinsichtlich bestimmter ethnischer Gruppen?
Bsp.: Wenn ich bemerke, dass SchülerInnen ausgeschlossen werden, unterbinde ich das.

33
Wie fördern Sie hilfreiches Verhalten in Ihrer Klasse?
Bsp.: Ich spreche im Unterricht an, dass die SchülerInnen sich gegenseitig helfen sollen.

34
Wie nutzen Sie diagnostische Erkenntnisse zur Unterrichtsgestaltung?
Bsp.: Ich ändere das Tempo meines Unterrichts.

35
Welche Informationen und Quellen nutzen Sie zur Optimierung Ihrer Arbeit?
Bsp.: Bei Problemen befrage ich meine Kollegen.

36
Wie diagnostizieren Sie individuelle Lernvoraussetzungen und –fortschritte?
Bsp.: Ich beobachte SchülerInnen bei der Still- und Gruppenarbeit.

37
Woran orientieren Sie sich bei der Beurteilung von SchülerInnen?
Bsp.: Ich überprüfe, ob formulierte Lernziele erreicht wurden.

38
An welchem Wissen orientieren Sie sich bei der Unterrichtsplanung?
Bsp.: An meinem Wissen aus den fachwissenschaftlichen Universitätsveranstaltungen.

39
Wie fördern Sie die Erfolgsoversicht Ihrer SchülerInnen (Beispiele)?
Bsp.: Nach schlechten Arbeitsergebnissen sage ich, dass das nächste Mal sicher besser wird.

40
Wie passen Sie Ihren Unterricht individuellen Lernvoraussetzungen an?
Bsp.: Ich gebe schnellen SchülerInnen mehr Aufgaben und arbeite mit den schlechteren in Stillarbeitsphasen zusammen.

41
Worüber denken Sie nach, wenn Sie an Ihre pädagogische Arbeit denken?
Bsp.: Ich überprüfe, ob meine Erwartungen realistisch sind.

42
In welcher Form evaluieren Sie Ihren Unterricht?
Bsp.: Ich nutze die Evaluationsmaterialien des Landesbildungsservers.

43
Wie reagieren Sie auf Fehler eines/r Schülers/Schülerin im Unterricht (z. B. beim Lösen einer Aufgabe an der Tafel)?
Bsp.: Ich sage: „Das ist falsch.“ und bitte eine/n andere/n SchülerIn, es erneut zu versuchen.

44
Wie unterstützen Sie ein vertrauensvolles Klassenklima?
Bsp.: Ich spreche SchülerInnen unter vier Augen an, wenn ich denke, dass ein Problem vorliegt.

45
Wie üben Sie mit Ihren SchülerInnen, in Konfliktsituationen verschiedene Perspektiven einzunehmen?
Bsp.: Ich veranstalte Rollenspiele.



46
Welche Weiterbildungsangebote nutzen Sie?
Bsp.: Ich lese Fachartikel.

47
Wie führen Sie im Unterricht meist in ein neues Thema/eine neue Fragestellung ein?
Bsp.: Ich gebe den SchülerInnen ein Problem/eine Situation vor und lasse sie überlegen, wie man es lösen könnte/wie sie zustande kommen könnte.

48
Wie thematisieren Sie Aggression und Gewalt in Ihrer Klasse?
Bsp.: Ich lasse die SchülerInnen über ein aktuelles Beispiel diskutieren.

49
Welche Kompetenzen eines/r SchülerIn mit Migrationshintergrund sollten Ihrer Meinung nach insbesondere gefördert werden?
Bsp.: Lesekompetenzen.

50
Welche Ergebnisse der Bildungsforschung nutzen Sie für Ihren Unterricht?
Bsp.: Ich lese aktuelle Fachzeitschriften.

51
Wie vermitteln Sie Ihren SchülerInnen moralische Grundwerte (Beispiele)?
Bsp.: Ich diskutiere mit den SchülerInnen moralische Dilemmata.

52
Wie gestalten Sie in Ihrem Unterricht soziale Lernprozesse?
Bsp.: Ich lasse die SchülerInnen diskutieren.

53
Welche rechtlichen Rahmenbedingungen des Lehrberufs halten Sie für besonders wichtig?
Bsp.: Das Einzahlen von kleineren Geldbeträgen bei Fehlverhalten ist unzulässig.

54
Wie erreichen Sie problemorientiertes Lernen (Beispiele)?
Bsp.: Ich gebe den SchülerInnen eine realistische Aufgabe vor und erfrage von ihnen Lösungsvorschläge (z. B. im Physikunterricht: „Wir wollen ein Radio bauen. Was könnten wir hierfür benötigen?“).

55
Wie erklären Sie Ihren SchülerInnen, wie deren Arbeitsergebnisse (z. B. in einer Klassenarbeit) zustande kommen?
Bsp.: Ich sage: „Wer nicht gelernt hat, kann auch keine gute Note schreiben.“

Ich möchte eine Rückmeldung erhalten und gebe hier meine e-mail-Adresse an:

---

## Anhang F: Interrater-Reliabilität und deskriptive Ergebnisse der qualitativen Datenerhebung zur Gewinnung der Basis negativ gepolter Items

### Die Homogenität der Randverteilungen der Urteile zweier RaterInnen

Items	Z**	p	Items	Z	p
1. Strukturierung des Unterrichts	-1.00 <sup>a</sup>	.157	29. Schulentwicklung	-.535 <sup>b</sup>	.593
2. Unterrichtsmethoden	-1.00 <sup>a</sup>	.317	30. Beurteilungskriterien	-.557 <sup>b</sup>	.577
3. Nutzung von Evaluationsergebnissen	-.447 <sup>a</sup>	.655	31. Schülermotivierung	-1.29 <sup>a</sup>	.197
4. Flexibler Strategieeinsatz	-.577 <sup>a</sup>	.564	32. Umgang mit Vorurteilen	-1.60 <sup>b</sup>	.108
5. Vorbildfunktion	-1,41 <sup>a</sup>	.157	33. Hilfsbereitschaft	-.707 <sup>b</sup>	.480
6. Beurteilungsnormen	-.632 <sup>b</sup>	.527	34. Nutzung diagnostischer Erkenntnisse	-.816 <sup>b</sup>	.414
7. Elternarbeit	-1.27 <sup>b</sup>	.206	35. Optimierungsquellen	-.577 <sup>a</sup>	.564
8. Beeinflussung von Lernergebnissen	-.707 <sup>a</sup>	.480	36. Individualdiagnostik	-.943 <sup>b</sup>	.346
9. Arbeitstechniken	-.277 <sup>a</sup>	.782	37. Lernziele	-.447 <sup>b</sup>	.655
10. Schülerförderung	-.707 <sup>a</sup>	.480	38. Wissensbasis Unterrichtsplanung	-1.00 <sup>b</sup>	.317
11. Wissenserwerb	-.905 <sup>a</sup>	.366	39. Erfolgszuversicht	-.302 <sup>b</sup>	.763
12. Lernhilfen	-1,41 <sup>a</sup>	.157	40. Lernervoraussetzungen	-1.00 <sup>b</sup>	.317
13. Flexible Unterrichtsgestaltung	-.426 <sup>b</sup>	.670	41. Reflektion	-.233 <sup>b</sup>	.816
14. Leistungsfeststellung	-.218 <sup>a</sup>	.827	42. Unterrichts-evaluation	-1.31 <sup>a</sup>	.190

---

**Die Homogenität der Randverteilungen der Urteile zweier RaterInnen**


---

Items	Z**	p	Items	Z	p
15. Werte	-1.0 <sup>b</sup>	.317	43. Fehlernutzung	-1.60 <sup>a</sup>	.109
16. Aktualisierung von Wissen	-1.00 <sup>b</sup>	.317	44. Klassenklima	-2.11 <sup>b</sup>	.035*
17. Projektplanung	-.025 <sup>a</sup>	.980	45. Perspektivenwechsel	-.733 <sup>a</sup>	.464
18. Medieneinsatz	-.655 <sup>b</sup>	.513	46. Weiterbildung	-1.50 <sup>a</sup>	.134
19. Schülerinteresse	-.447 <sup>a</sup>	.655	47. Themeneinführung	-.632 <sup>b</sup>	.527
20. eigenverantwortliches Urteilen der Schüler	-2.12 <sup>a</sup>	.034*	48. Umgang mit Aggression	-1.63 <sup>a</sup>	.102
21. Respektvoller Umgang	-1.00 <sup>a</sup>	.317	49. Förderung bei Migrationshintergrund	-1,00 <sup>b</sup>	.317
22. Informationsbasis	-1,41 <sup>a</sup>	.157	50. Bildungsforschung	-1.73 <sup>b</sup>	.083
23. Subjektivität	-0.50 <sup>b</sup>	.961	51. Moralentwicklung	-1.89 <sup>a</sup>	.059
24. Entscheidungsfindung	-.471 <sup>a</sup>	.637	52. Soziale Lernprozesse	-1.15 <sup>a</sup>	.248
25. Konfliktlösung	-.790 <sup>b</sup>	.430	53. Rechtlicher Rahmen	-1.00 <sup>b</sup>	.317
26. Transfer	-.333 <sup>a</sup>	.739	54. Problemorientierung	-1,00 <sup>a</sup>	.317
27. Schülerprojekte	-.100 <sup>a</sup>	.317	55. Arbeitsergebnisse	-.447 <sup>a</sup>	.655
28. Förderungskriterien	-1.41 <sup>b</sup>	.157			

---

\* p>.05 a. basierend auf negativen Rängen b. basierend auf positiven Rängen

\*\* standardnormalverteilte Prüfgröße Z

**Rangkorrelationen der Urteile zweier RaterInnen**

Items	r	p	Items	r	p
1. Strukturierung des Unterrichts	.950**	.000	29. Schulentwicklung	.864**	.000
2. Unterrichtsmethoden	.941**	.000	30. Beurteilungskriterien	.548**	.000
3. Nutzung von Evaluationsergebnissen	.923**	.000	31. Schülermotivierung	.770**	.000
4. Flexibler Strategieeinsatz	.935**	.000	32. Umgang mit Vorurteilen	.745**	.000
5. Vorbildfunktion	.961**	.000	33. Hilfsbereitschaft	.872**	.000
6. Beurteilungsnormen	.828**	.000	34. Nutzung diagnostischer Erkenntnisse	.929**	.000
7. Elternarbeit	.806**	.000	35. Optimierungsquellen	.848**	.000
8. Beeinflussung von Lernergebnissen	.863**	.000	36. Individualdiagnostik	.657**	.000
9. Arbeitstechniken	.719**	.000	37. Lernziele	.914**	.000
10. Schülerförderung	.741**	.000	38. Wissensbasis Unterrichtsplanung	.986**	.000
11. Wissenserwerb	.799**	.000	39. Erfolgszuversicht	.836**	.000
12. Lernhilfen	.873**	.000	40. Lerner Voraussetzungen	.649**	.000
13. Flexible Unterrichtsgestaltung	.595**	.000	41. Reflektion	.698**	.000
14. Leistungs feststellung	.577**	.000	42. Unterrichts evaluation	.852**	.000
15. Werte	.980**	.000	43. Fehlernutzung	.787**	.000
16. Aktualisierung von Wissen	.990**	.000	44. Klassenklima	.853**	.000
17. Projektplanung	.648**	.000	45. Perspektivenwechsel	.577**	.000
18. Medieneinsatz	.658**	.000	46. Weiterbildung	.773**	.000
19. Schülerinteresse	.872**	.000	47. Themen einführung	.751**	.000

**Rangkorrelationen der Urteile zweier RaterInnen**

<b>Items</b>	<b>r</b>	<b>p</b>	<b>Items</b>	<b>r</b>	<b>p</b>
20. Urteilen der Schüler	.981**	.000	48. Umgang mit Aggression	.903**	.000
21. Respektvoller Umgang	.986**	.000	49. Förderung Migrationshintergrund	.964**	.000
22. Informationsbasis	.754**	.000	50. Bildungsforschung	.967**	.000
23. Subjektivität	.688**	.000	51. Moralentwicklung	.927**	.000
24. Entscheidungsfindung	.634**	.000	52. Soziale Lernprozesse	.815**	.000
25. Konfliktlösung	.606**	.000	53. Rechtlicher Rahmen	.987**	.000
26. Transfer	.891**	.000	54. Problemorientierung	.986**	.000
27. Schülerprojekte	.924**	.000	55. Arbeitsergebnisse	.929**	.000
28. Förderungskriterien	.970**	.000			

\*\*p = .000

**Prozentuale Anteile sowie absolute Anzahlen „standardkonformer“, „teilweise standardkonformer“ und „nicht standardkonformer“ Antworten pro Item des Instruments "Qualitative Erfassung professioneller pädagogischer Kompetenzen Lehramtsstudierender"**

Item	standardkonform	teilweise	
		standardkonform	nicht standardkonform
1. Strukturierung des Unterrichts	26,2% (28)	57,9% (62)	11,2% (12)
2. Unterrichtsmethoden	38,3% (41)	54,2% (58)	3,7% (4)
3. Nutzung von Evaluationsergebnissen	59,8% (64)	28,0% (30)	3,7% (4)
4. Flexibler Strategieeinsatz	76,6% (82)	18,7% (20)	1,9% (2)
5. Vorbildfunktion	55,1% (59)	40,2% (43)	0,0% (0)
6. Beurteilungsnormen	28,0% (30)	5,3% (57)	4,7% (5)
7. Elternarbeit	78,5% (84)	13,1% (14)	2,8% (3)
8. Beeinflussung von Lernergebnissen	21,5% (23)	61,7% (66)	5,6% (6)
9. Arbeitstechniken	70,1% (75)	21,5% (23)	0,9% (1)
10. Schülerförderung	86,0% (92)	4,7% (5)	1,9% (2)
11. Wissenserwerb	26,2% (28)	54,2% (58)	3,7% (4)
12. Lernhilfen	30,8% (33)	60,7% (65)	4,7% (5)
13. Flexible Unterrichtsgestaltung	52,3% (56)	26,2% (28)	0,9% (1)
14. Leistungsfeststellung	65,4% (70)	13,1% (14)	1,9% (2)
15. Werte	68,2% (73)	29,9% (32)	0,9% (1)
16. Aktualisierung von Wissen	29,9% (32)	50,5% (54)	15,9% (17)
17. Projektplanung	10,3% (11)	52,3% (56)	16,8% (18)
18. Medieneinsatz	40,2% (43)	39,3% (42)	0,9% (1)
19. Schülerinteresse	10,3% (11)	79,4% (85)	5,6% (6)

**Prozentuale Anteile sowie absolute Anzahlen „standardkonformer“, „teilweise standardkonformer“ und „nicht standardkonformer“ Antworten pro Item des Instruments "Qualitative Erfassung professioneller pädagogischer Kompetenzen Lehramtsstudierender"**

Item	standardkonform	teilweise	
		standardkonform	nicht standardkonform
20. Urteilen der Schüler	46,7% (50)	43,9% (47)	2,8% (3)
21. Respektvoller Umgang	56,1% (60)	38,3% (41)	4,7% (5)
22. Informationsbasis	22,4% (24)	44,9% (48)	7,5% (8)
23. Subjektivität	53,3% (57)	20,6% (22)	9,3% (10)
24. Entscheidungsfindung	63,6% (68)	19,6% (21)	1,9% (2)
25. Konfliktlösung	65,4% (70)	15,9% (17)	4,7% (5)
26. Transfer	43,0% (46)	38,3% (41)	6,5% (7)
27. Schülerprojekte	19,6% (21)	71,0% (76)	4,7% (5)
28. Förderungskriterien	53,3% (57)	41,1% (44)	3,7% (4)
29. Schulentwicklung	29,0% (31)	37,4% (40)	15,0% (16)
30. Beurteilungskriterien	53,3% (57)	20,6% (22)	2,8% (3)
31. Schülermotivierung	57,0% (61)	25,2% (27)	2,8% (3)
32. Umgang mit Vorurteilen	52,3% (56)	26,2% (28)	3,7% (4)
33. Hilfsbereitschaft	39,3% (42)	48,6% (52)	2,8% (3)
34. Nutzung diagnostischer Erkenntnisse	37,4% (40)	26,2% (28)	9,3% (10)
35. Optimierungsquellen	90,7% (97)	4,7% (5)	0,9% (1)
36. Individualdiagnostik	60,7% (65)	18,7% (20)	1,9% (2)
37. Lernziele	65,4% (70)	27,1% (29)	1,9% (2)
38. Wissensbasis	72,0% (77)	21,5% (23)	5,6% (6)
Unterrichtsplanung			
39. Erfolgszuversicht	56,1% (60)	29,9% (32)	2,8% (3)
40. Lernvoraussetzungen	72,0% (77)	15,0% (16)	1,9% (2)



**Prozentuale Anteile sowie absolute Anzahlen „standardkonformer“, „teilweise standardkonformer“ und „nicht standardkonformer“ Antworten pro Item des Instruments "Qualitative Erfassung professioneller pädagogischer Kompetenzen Lehramtsstudierender"**

Item	standardkonform	teilweise	
		standardkonform	nicht standardkonform
41. Reflektion	32,7% (35)	38,3% (41)	5,6% (6)
42. Unterrichtsevaluation	23,4% (25)	38,3% (41)	18,7% (20)
43. Fehlernutzung	24,3% (26)	60,7% (65)	5,6% (6)
44. Klassenklima	62,6% (67)	29,0% (31)	0,9% (1)
45. Perspektivenwechsel	75,7% (81)	15,0% (16)	1,9% (2)
46. Weiterbildung	42,1% (45)	40,2% (43)	3,7% (4)
47. Themeneinführung	77,6% (83)	15,0% (16)	0,9% (1)
48. Umgang mit Aggression	62,6% (67)	27,1% (29)	2,8% (3)
49. Förderung	86,9% (93)	10,3% (11)	0,9% (1)
Migrationshintergrund			
50. Bildungsforschung	29,9% (32)	41,1% (44)	13,1% (14)
51. Moralentwicklung	51,4% (55)	28,0% (30)	9,3% (10)
52. Soziale Lernprozesse	42,1% (45)	42,1% (45)	3,7% (4)
53. Rechtlicher Rahmen	20,6% (22)	29,9% (32)	15,9% (17)
54. Problemorientierung	64,5% (69)	21,5% (23)	4,7% (5)
55. Arbeitsergebnisse	20,6% (22)	59,8% (64)	12,1% (13)

Zugunsten der Übersichtlichkeit wird auf die Darstellung der Anteile für Zwischenzuordnungen (0,5 oder 1,5) verzichtet.

In Klammern finden sich die Personenanzahlen entsprechend der Prozente.

Die Items mit den häufigsten Nennungen von „standardkonformen“ und „nicht standardkonformen“ Kompetenzen sind kursiv markiert.

## Anhang G: Eignungseinschätzung der 102 Items des Situational Judgement

### Tests durch 6 ExpertInnen

#### Eignungseinschätzung der 102 Items des Situational Judgement Tests durch 6 ExpertInnen

Item	Experte 1	Experte 2	Experte 3	Experte 4	Experte 5	Experte 6
Unterrichtsstruktur+	1	1	1	1	1	1
Unterrichtsstruktur -	1	1	0	0	1	1
Auswahl Methoden/Medien +	1	1	1	1	1	1
Auswahl Methoden/Medien -	1	1	0	1	0	1
Nutzung von Forschungs- ergebnissen +	0	1	1	1	1	1
Nutzung von Forschungs- ergebnissen -	1	1	1	1	0	1
Flexibilität Unterrichtsgestaltung +	1	1	0	0	1	1
Flexibilität Unterrichtsgestaltung -	1	1	1	1	0	1
Transfer +	1	1	1	1	1	1
Transfer -	1	1	1	1	1	1
Problemorientierung +	1	0	1	1	1	1
Problemorientierung -	0	1	1	1	0	1
Prinzipien Wissenserwerb +	1	1	0	0	1	1
Prinzipien Wissenserwerb -	1	1	1	1	0	1
Fehlernutzung +	1	1	1	1	1	1
Fehlernutzung -	1	1	1	1	1	0
Schülermotivation +	1	1	0	1	1	1
Schülermotivation -	1	1	1	1	1	0
Lernhilfen+	1	1	1	1	1	1
Lernhilfen -	1	1	1	1	1	1
Einführung in neues Thema +	1	1	1	1	1	1
Einführung in neues Thema -	1	1	1	1	1	1
Arbeitstechniken +	1	1	1	0	1	1
Arbeitstechniken -	1	1	1	0	1	1
Erfolgszuversicht +	0	1	0	1	1	1
Erfolgszuversicht -	1	1	1	1	1	1
Zustandekommen von Arbeitsergebnissen +	1	1	1	1	1	1
Zustandekommen von Arbeitsergebnissen -	1	1	0	0	1	1
Schülerzentrierung +	1	1	1	1	1	1
Schülerzentrierung -	1	1	1	1	1	1
Schülerinteresse +	1	1	1	0	1	1
Schülerinteresse -	1	1	1	0	1	0
Umgang mit Vorurteilen +	1	1	1	1	1	1
Umgang mit Vorurteilen -	1	1	1	1	1	0
Eigenverantwortliches Urteilen +	1	1	1	1	1	1

**Eignungseinschätzung der 102 Items des Situational Judgement Tests durch 6 ExpertInnen**

Item	Experte 1	Experte 2	Experte 3	Experte 4	Experte 5	Experte 6
Eigenverantwortliches Urteilen -	1	1	1	1	1	0
Vermittlung Werte und Normen +	1	0	1	1	1	1
Vermittlung Werte und Normen -	1	1	1	1	0	1
Entscheidungsfindung +	1	1	1	1	1	1
Entscheidungsfindung -	1	1	1	1	1	1
Vertrauensvolles Klima +	1	0	1	0	1	1
Vertrauensvolles Klima -	0	1	1	1	1	1
Förderung Hilfsbereitschaft +	1	0	0	1	1	1
Förderung Hilfsbereitschaft -	1	1	1	1	1	0
Thematisieren von Aggression +	1	1	1	1	1	1
Thematisieren von Aggression -	1	1	1	1	1	0
Soziale Lernprozesse +	1	0	1	1	1	1
Soziale Lernprozesse -	1	1	1	1	1	1
Transparenter Umgang mit Konflikten +	1	1	1	1	1	1
Transparenter Umgang mit Konflikten -	1	1	1	1	1	1
Perspektivenwechsel +	0	1	1	1	1	1
Perspektivenwechsel -	1	1	1	0	1	1
Individualdiagnostik +	1	1	1	1	1	1
Individualdiagnostik -	1	1	1	1	1	0
Nutzen von Diagnostik zur Unterrichtsgestaltung +	1	1	1	1	1	1
Nutzen von Diagnostik zur Unterrichtsgestaltung -	1	1	1	1	1	1
Elternarbeit +	1	1	1	1	1	1
Elternarbeit -	1	1	1	1	0	0
Beratung ohne Beurteilung +	1	1	1	0	1	1
Beratung ohne Beurteilung -	1	1	1	1	1	1
Methoden der Leistungsfeststellung +	1	1	1	1	1	1
Methoden der Leistungsfeststellung -	0	1	0	1	1	1
Einfluss von Beurteilung auf Lernprozesse +	1	1	1	1	1	0
Einfluss von Beurteilung auf Lernprozesse -	1	1	0	0	1	1
Transparente Beurteilungskriterien +	1	1	1	1	1	0
Transparente Beurteilungskriterien -	1	1	1	1	1	0
Kombination von Beurteilungsnormen +	1	1	1	1	1	1
Kombination von Beurteilungsnormen -	1	1	1	1	1	1
Subjektivität von Beurteilungen +	1	1	1	1	1	0
Subjektivität von Beurteilungen -	0	1	1	1	1	1

## Eignungseinschätzung der 102 Items des Situational Judgement Tests durch 6 ExpertInnen

Item	Experte 1	Experte 2	Experte 3	Experte 4	Experte 5	Experte 6
Bezug Lernziel- Beurteilung +	1	1	1	1	1	1
Bezug Lernziel- Beurteilung -	1	1	1	1	1	1
Erwartungsreflektion +	1	1	1	1	1	0
Erwartungsreflektion -	1	1	1	0	1	0
Vorbildfunktion +	1	1	1	1	1	1
Vorbildfunktion -	1	1	1	1	1	1
Mitwirkungs- möglichkeiten +	1	1	1	1	1	1
Mitwirkungs- möglichkeiten -	1	1	0	1	1	1
Wertereflektion +	1	1	1	1	1	1
Wertereflektion -	1	1	1	0	1	1
Rechtlicher Rahmen +	1	1	1	0	1	1
Rechtlicher Rahmen -	1	1	1	0	1	1
Weiterbildung +	1	1	1	1	1	1
Weiterbildung -	0	1	1	1	0	1
Nutzung von Bildungsforschung +	1	1	1	1	1	1
Nutzung von Bildungsforschung -	1	1	1	1	1	0
Reflektion beruflicher Erfahrungen +	1	1	1	1	1	1
Reflektion beruflicher Erfahrungen -	1	1	1	1	1	0
Aktualisierung von Wissen +	1	1	1	1	1	1
Aktualisierung von Wissen -	1	1	1	1	1	1
Nutzung von Feedback +	1	1	1	1	1	1
Nutzung von Feedback -	0	1	1	1	1	1
Nutzung von Evaluations- ergebnissen +	1	1	1	1	1	1
Nutzung von Evaluations- ergebnissen -	0	1	1	0	1	1
Forschungsbasierte Projekte +	1	1	1	1	1	0
Forschungsbasierte Projekte -	1	1	1	1	1	1
Unterrichtsevaluation +	1	1	1	1	1	1
Unterrichtsevaluation -	1	1	1	1	1	0
Durchführung schulischer Projekte +	1	1	1	1	1	1
Durchführung schulischer Projekte -	1	1	1	1	1	0
Unterstützung von SchülerInnenprojekten +	1	1	1	1	1	0
Unterstützung von SchülerInnenprojekten -	1	1	1	1	1	0

1= Einschätzung eines Items als *geeignet*0= Einschätzung eines Items als *ungeeignet*

## Anhang H: Der Situational Judgement Test

### Simulationsorientierte Erfassung professioneller pädagogischer Kompetenzen Lehramtsstudierender

Sehr geehrte Studentin, sehr geehrter Student,

um die Qualität schulischer Bildung überprüfen und sichern zu können, wurden durch die Kultusministerkonferenz verbindliche Standards formuliert, welche sich auf verschiedene Bildungsbereiche in Schule und Hochschule beziehen. Beispielsweise für den Bereich der Lehrerbildung wurden durch die „Standards für die Lehrerbildung: Bildungswissenschaften“ (KMK, 2004) **elf Kompetenzbereiche** definiert, über welche Lehramtsstudierende (in Abhängigkeit von der Studiendauer) verfügen sollen.

Valide Erhebungsinstrumente müssen zur Verfügung stehen um das Vorhandensein dieser Kompetenzen überprüfen zu können. Ein neu entwickeltes Instrument halten Sie in Form eines situations- und kontextbezogenen Tests in Händen. Dieser Test soll zunächst auf seine Eignung zur Erfassung professioneller Kompetenzen im Lehrberuf überprüft werden.

Das Instrument „Simulationsorientierte Erfassung professioneller Kompetenzen im Lehrberuf“ soll die **pädagogischen Kompetenzen** angehender Lehrpersonen **kontextbezogen** erfassen. Hierfür werden 102 fiktive Handlungsweisen und Überlegungen aus dem Alltag von LehrerInnen vorgegeben, die von Ihnen bezüglich ihrer Effektivität bewertet werden sollen.

Bitte schätzen Sie **spontan** ein, ob Ihnen das jeweils genannte Verhalten/die genannte Überlegung als sehr ineffektiv (1), ineffektiv (2), eher ineffektiv (3), eher effektiv (4), effektiv (5) oder sehr effektiv (6) erscheint. Kreuzen Sie die von Ihnen **gewählte Ziffer bitte deutlich und eindeutig** an.

Außerdem bitten wir Sie um einige **Angaben zu Ihrem Studium sowie um die Angabe Ihrer letzte Note im Studium bzw. Ihrer Abiturnote** – diese Informationen sind außerordentlich wichtig, um den Nutzen des Instrumentes beurteilen zu können.

**Selbstverständlich werden Ihre Daten absolut anonym und vertraulich behandelt und keinesfalls an Dritte weitergegeben!**

Für Ihre Mitarbeit und wertvolle Unterstützung bedanken wir uns im Voraus herzlich!

Beste Grüße,

Ines Weresch-Deperrois, Dipl.-Psych., Dipl.-Mus.

**Ihre Codenummer:**

erster Buchstabe des Vornamens Ihrer Mutter

erster Buchstabe des Vornamens Ihres Vaters

erste Zahl des eigenen Geburtstags

(z. B.: „H“ für „Heike“, „B“ für „Benjamin“ und „27“ für „27.07.“)

6. Im wievielten Semester studieren Sie im Lehramtsstudium?

7. In welchem Bundesland studieren

Sie? \_\_\_\_\_

8. Ihr Berufsziel ist das Lehramt an...

Grundschule  Realschule  Gymnasium  Förderschule   
 Berufsschule

Ich habe mich **noch nicht** für eine Schulart entschieden. 9. Sie sind... weiblich  männlich 

10. An welchen Praktika haben Sie bereits teilgenommen?

Orientierendes Praktikum 1  2  3 Vertiefendes Praktikum 1  2 Bislang habe ich **noch nicht** an einem Praktikum teilgenommen. 

11. Bitte nennen Sie die Note Ihrer letzten schriftlichen oder mündlichen Prüfung während des Studiums:

12. Bei welcher Prüfung haben Sie diese Note erhalten? \_\_\_\_\_

8. Bitte nennen Sie Ihre Abiturdurchschnittsnote:

<b>Für wie effektiv halten Sie das Verhalten/die Überlegungen der jeweiligen Lehrperson?</b>							
1=sehr ineffektiv 2= ineffektiv 3=eher ineffektiv 4=eher effektiv 5= effektiv 6=sehr effektiv Bitte kreuzen Sie die Ziffer an, die Ihrer Einschätzung entspricht.							

1.	Lehrerin B beginnt den Englisch-Unterricht in ihrer Klasse. Lehrerin B sagt: „Wir behandeln heute das ‚future perfect‘. Wir werden erst den Stoff von gestern wiederholen, dann gemeinsam	1	2	3	4	5	6
2.	Lehrer Es Klasse hat seine hochgesteckten Ziele in diesem Schuljahr nicht erreicht. Lehrer E denkt: „Habe ich zu hohe Erwartungen?“	1	2	3	4	5	6
3.	Einige Schüler von Lehrer Ks zehnter Klasse möchten eine Klassenzeitung gründen und berichten Lehrer K davon. Lehrer K sagt: „Also dann: Ich wünsche euch viel Erfolg.“	1	2	3	4	5	6
4.	Schülerin G stört häufig den Unterricht von Lehrer T und gehört zu den eher schwachen Schülern. Lehrer T überlegt, wie er weiter vorgehen soll. Lehrer T denkt: „Ich werde mit den Eltern	1	2	3	4	5	6
5.	Lehrer Z stellt seiner Klasse eine Frage und erhält von einem Schüler eine falsche Antwort. Lehrer Z sagt: „Nein, das stimmt nicht. Wer kann es richtig sagen?“	1	2	3	4	5	6
6.	Familie M kommt bezüglich ihrer Tochter T zum Beratungsgespräch zu Lehrerin F. Lehrer F sagt: „Ts Leistungen liegen unter dem Klassenschnitt. Sie muss also mehr lernen.“	1	2	3	4	5	6
7.	Schüler B zeigt im Vergleich zur Klasse eher schwache Leistungen. In einer Klassenarbeit schneidet B nicht gut ab. Lehrer I hat das erwartet. Lehrer I sagt zu B: „Ich weiß, dass Du es	1	2	3	4	5	6

<b>Für wie effektiv halten Sie das Verhalten/die Überlegungen der jeweiligen Lehrperson?</b>						
1=sehr ineffektiv 2= ineffektiv 3=eher ineffektiv 4=eher effektiv 5= effektiv 6=sehr effektiv Bitte kreuzen Sie die Ziffer an, die Ihrer Einschätzung entspricht.						

8.	Lehrer K möchte die SchülerInnen seiner Klasse beurteilen. Lehrer K denkt: „Ich bespreche auch mit den Kollegen, die diese Klasse kennen, wie sie die einzelnen SchülerInnen einschätzen.“	1	2	3	4	5	6
9.	Lehrerin M plant ein großes Schulprojekt und fragt Lehrerin L, ob sie mitarbeiten möchte. Lehrerin L sagt: „Ich nehme teil. Ich finde es wichtig, auch außerhalb des Unterrichts mit den SchülerInnen zu arbeiten.“	1	2	3	4	5	6
10.	Lehrerin As Klasse schneidet in einem standardisierten Rechentest eher schlecht ab. Lehrerin A überlegt, was sie tun sollte. Lehrerin A denkt: „Der Test stimmt mit meinem Eindruck von der Klasse nicht überein,	1	2	3	4	5	6
11.	Lehrer Z möchte in der nächsten Stunde in ein neues Thema einführen. Er überlegt, wie er vorgehen soll. Lehrer Z denkt: „Ich beginne mit einem Vortrag, der das Thema darstellt.“	1	2	3	4	5	6
12.	Lehrerin G übt mit den Schülern ihrer siebten Klasse das Verstehen von Sachtexten. Sie gibt der Klasse Arbeitsanweisungen. Lehrerin G sagt: „Lest den Text durch und streicht das Wichtigste an. Dann sammeln wir	1	2	3	4	5	6
13.	Ein Schüler hat bei einer Klassenarbeit schlecht abgeschnitten. Lehrer H spricht mit ihm über das Ergebnis. Lehrer H sagt: „Du hast nur wenig gelernt? Dann konntest du natürlich die Aufgaben nicht	1	2	3	4	5	6
14.	In der Klasse 9d werden immer wieder Bemerkungen laut wie: „Alle Muslime sind Terroristen – vor denen muss man Angst haben.“ Lehrer F möchte die Vorurteile abbauen.	1	2	3	4	5	6



**Für wie effektiv halten Sie das Verhalten/die Überlegungen der jeweiligen Lehrperson?**

1=sehr ineffektiv 2= ineffektiv 3=eher ineffektiv 4=eher effektiv 5= effektiv 6=sehr effektiv  
Bitte kreuzen Sie die Ziffer an, die Ihrer Einschätzung entspricht.

15.	Ein Wandertag ist in Klasse 9d geplant. Lehrer R bespricht mit den SchülerInnen, wo es hingehen soll. Lehrer R sagt: „Wir sammeln einmal Vorschläge, wo wir hinfahren könnten.“	1	2	3	4	5	6
16.	Schüler P scheint im Unterricht niedergeschlagen und abwesend, ohne dass ein offensichtlicher Grund vorliegt. Lehrer M spricht dies in der Pause an. Lehrer M sagt: „Was lenkt dich denn ab?“	1	2	3	4	5	6
17.	Lehrerin S möchte hilfsbereites Verhalten ihrer SchülerInnen fördern. Lehrerin S sagt: „Wer schnell mit seiner Aufgabe fertig ist, fragt bitte den Nachbarn, ob er ihn unterstützen kann.“	1	2	3	4	5	6
18.	An einer benachbarten Schule werden seit einiger Zeit immer wieder SchülerInnen von anderen gewaltsam angegriffen. Lehrer Bs Klasse ist nicht direkt betroffen. Lehrer B denkt: „Ich lasse die Schüler über die“	1	2	3	4	5	6
19.	Lehrerin S plant den Deutsch-Unterricht in ihrer achten Klasse und überlegt, welche Methoden sie vor allem anwenden soll. Lehrerin S denkt: „Ich lasse am Besten die Schüler in Stillarbeit Aufgaben lösen.“	1	2	3	4	5	6
20.	Lehrerin N hat ihre SchülerInnen einen Fragebogen zur Unterrichtsevaluation bearbeiten lassen und betrachtet nun die Ergebnisse. Lehrerin N denkt: „Ich behalte meinen“	1	2	3	4	5	6
21.	Lehrer U möchte seinen Schülern Werte und Normen vermitteln und überlegt, wie er dies umsetzen könnte. Lehrer U denkt: „Ich stelle Regeln auf.“	1	2	3	4	5	6

**Für wie effektiv halten Sie das Verhalten/die Überlegungen der jeweiligen Lehrperson?**

1=sehr ineffektiv 2= ineffektiv 3=eher ineffektiv 4=eher effektiv 5= effektiv 6=sehr effektiv  
Bitte kreuzen Sie die Ziffer an, die Ihrer Einschätzung entspricht.

22.	Lehrerin E möchte ihren Schüler S beurteilen. Lehrerin E denkt: „Die Klassenarbeiten von S waren gut. Ich werde ihn also gut beurteilen.“	1	2	3	4	5	6
23.	Lehrer X hat Schüler Bs Deutsch-Klassenarbeit korrigiert. Lehrer X schreibt einen kurzen Kommentar unter die mit 3,5 benotete Arbeit. Lehrer X schreibt: „Du hast deine	1	2	3	4	5	6
24.	Lehrerin D erwartet von seinen SchülerInnen Pünktlichkeit. Sie selbst kommt häufig zu spät, worauf sie die SchülerInnen ansprechen. Lehrerin D sagt: „Ich werde in Zukunft darauf achten.“	1	2	3	4	5	6
25.	An der Schule von Lehrer Y sollen die LehrerInnen stark an der Schulentwicklung beteiligt werden. Lehrer Y denkt: „Ich möchte nicht so häufig teilnehmen, da ich es wichtiger finde, mich voll	1	2	3	4	5	6
26.	Lehrer B beginnt den Englisch-Unterricht in seiner Klasse. Lehrer B sagt: „Wir werden heute gut voran kommen. Lisa, was war die Hausaufgabe? Martin hat noch eine Frage zum gestrigen	1	2	3	4	5	6
27.	Lehrer G plant den Unterricht in Klasse 11c. Diese Klasse kennt er noch nicht. Lehrer G denkt: „Zur Unterrichtsplanung orientiere ich mich an den Erfahrungen in meiner eigenen Schulzeit, da ich über die	1	2	3	4	5	6
28.	Lehrerin A möchte, dass ihre Schüler das Prinzip des Dividierens gut verstehen. Lehrerin A sagt: „Rechnet bitte: 10 durch 2.“	1	2	3	4	5	6

**Für wie effektiv halten Sie das Verhalten/die Überlegungen der jeweiligen Lehrperson?**

1=sehr ineffektiv 2= ineffektiv 3=eher ineffektiv 4=eher effektiv 5= effektiv 6=sehr effektiv  
Bitte kreuzen Sie die Ziffer an, die Ihrer Einschätzung entspricht.

29.	Lehrer K möchte im Physik-Unterricht die Elektrik-Kenntnisse seiner SchülerInnen vertiefen. Lehrer K sagt: „Wir wollen zusammen ein Radio bauen. Was brauchen wir dafür?“	1	2	3	4	5	6
30.	Schüler P hat bei Klassenarbeiten von Lehrer J wiederholt nicht gut abgeschnitten. Lehrer J sagt: „Ich weiß, dass du es besser kannst.“	1	2	3	4	5	6
31.	Die Schüler B und C hinken im Mathematikunterricht der 9d etwas hinterher und haben insbesondere Schwierigkeiten mit Textaufgaben. Lehrerin K gibt einen Arbeitsauftrag.	1	2	3	4	5	6
32.	Lehrer O möchte Umgangsregeln in seiner Klasse einrichten. Lehrer O sagt: „Wir erarbeiten gemeinsam einen Regelkatalog und vereinbaren, was geschehen soll, wenn die Regeln nicht	1	2	3	4	5	6
33.	In der Klasse 9d werden immer wieder Bemerkungen laut wie: „Alle Muslime sind Terroristen – vor denen muss man Angst haben.“ Lehrer F möchte die Vorurteile abbauen.	1	2	3	4	5	6
34.	In Lehrer Ps Klasse haben sich zwei Schülergruppen gebildet, die sich feindlich gegenüber stehen. Lehrer P thematisiert die Situation. Lehrer P sagt: „In der Klasse haben sich zwei	1	2	3	4	5	6
35.	Lehrerin U kündigt eine Deutsch-Klassenarbeit mit dem Thema „Erörterung“ an. Lehrerin U sagt: „Wer mehr als 30 Punkte erreicht, bekommt eine 1.“	1	2	3	4	5	6

**Für wie effektiv halten Sie das Verhalten/die Überlegungen der jeweiligen Lehrperson?**

1=sehr ineffektiv 2= ineffektiv 3=eher ineffektiv 4=eher effektiv 5= effektiv 6=sehr effektiv  
Bitte kreuzen Sie die Ziffer an, die Ihrer Einschätzung entspricht.

36.	Ein Wandertag ist in Klasse 9d geplant. Lehrer I bespricht mit den SchülerInnen, wo es hingehen soll. Lehrer R sagt: „Wir fahren zur Burg, das gefällt euch sicher am besten.“	1	2	3	4	5	6
37.	Lehrerin A plant den Unterricht für die Klassen 5c und 8a. Sie überlegt dabei, welche Methoden sie anwenden soll. Lehrerin A denkt: „Welche Methoden und Medien eignen sich für welche Klasse?“	1	2	3	4	5	6
38.	Lehrerin E lässt die Schüler in der 10. Klasse ihrer Realschule Aufgaben in Kleingruppen bearbeiten. Als sie durch die Reihen geht, bemerkt sie, dass die Aufgaben kaum gelöst werden.	1	2	3	4	5	6
39.	Lehrerin C hat die achte Klasse ihres Gymnasiums in das Prinzip der linearen Funktionen eingeführt. Dann sagt Lehrerin C: „Ich rechne eine Aufgabe vor, dann rechnen wir gemeinsam mit	1	2	3	4	5	6
40.	Schülerin J spricht in den Pausen häufig begeistert der Kfz-Werkstatt ihres Onkels. Lehrer B möchte dieses Interesse fördern. Lehrer B sagt: „Welches ist denn dein Lieblingsauto?“	1	2	3	4	5	6
41.	Schüler T wird in Lehrer As sechster Klasse ausgegrenzt und lächerlich gemacht. Lehrer A möchte thematisieren. Lehrer A sagt: „Heute sprechen wir über euer unfaires Verhalten T gegenüber.“	1	2	3	4	5	6
42.	Lehrerin U kündigt eine Deutsch-Klassenarbeit mit dem Thema „Erörterung“ an. Lehrerin U sagt: „Ich werde vor allem darauf achten, ob ihr eure Gedanken klar und differenziert darstellt.“	1	2	3	4	5	6

**Für wie effektiv halten Sie das Verhalten/die Überlegungen der jeweiligen Lehrperson?**

1=sehr ineffektiv 2= ineffektiv 3=eher ineffektiv 4=eher effektiv 5= effektiv 6=sehr effektiv  
Bitte kreuzen Sie die Ziffer an, die Ihrer Einschätzung entspricht.

43.	Lehrerin E bespricht zu Schuljahrsbeginn in ihrer Klasse den Englisch-Unterricht der nächsten Monate. Lehrerin E sagt: „Ich bin zufrieden, wenn ihr englische Texte gut ins Deutsche übersetzen“	1	2	3	4	5	6
44.	Lehrer A und Lehrer J diskutieren über ihre Wertvorstellungen bezüglich ihres Berufs. Beide vertreten unterschiedliche Meinungen. Lehrer J denkt: „Ich weiß, dass meine Werte die richtigen sind.“	1	2	3	4	5	6
45.	Lehrerin S hat von einer neuartigen Unterrichtsmethode gehört, die interessant zu sein scheint. Lehrerin S denkt: „Ich habe genügend Methoden zur Auswahl.“	1	2	3	4	5	6
46.	Lehrerin D kommt mit Klasse 7a sehr gut zurecht, mit Klasse 9b hat sie leider immer wieder Probleme. Lehrerin D denkt: „In Klasse 9b werde ich anders vorgehen müssen.“	1	2	3	4	5	6
47.	Lehrer P plant die Themen für seine Physik-AG. Lehrer P denkt: „Ich informiere mich, ob es neue wissenschaftliche Erkenntnisse gibt.“	1	2	3	4	5	6
48.	Lehrerin M plant eine Projektwoche mit ihrer Klasse. Sie überlegt, wie sie das Projekt gestalten soll. Lehrerin M denkt: „Ich möchte ganz kreativ vorgehen und mich nicht durch zu viel Theorie“	1	2	3	4	5	6
49.	Lehrerin E lässt die Schüler in der 10. Klasse ihrer Realschule Aufgaben in Kleingruppen bearbeiten. Als sie durch die Reihen geht, bemerkt sie, dass die Aufgaben kaum gelöst werden.	1	2	3	4	5	6

**Für wie effektiv halten Sie das Verhalten/die Überlegungen der jeweiligen Lehrperson?**

1=sehr ineffektiv 2= ineffektiv 3=eher ineffektiv 4=eher effektiv 5= effektiv 6=sehr effektiv  
Bitte kreuzen Sie die Ziffer an, die Ihrer Einschätzung entspricht.

50.	Lehrer U möchte seinen Schülern Werte und Normen vermitteln und überlegt, wie er dies umsetzen könnte. Lehrer U denkt: „Ich gebe den Schülern ein moralisches Dilemma vor, über das wir dann	1	2	3	4	5	6
51.	Lehrerin L plant den Deutsch-Unterricht in ihrer achten Klasse und überlegt, welche Methoden sie vorwiegend anwenden soll. Lehrerin L denkt: „Ich lasse am Besten die Schüler in Gruppen Aufgaben diskutieren.“	1	2	3	4	5	6
52.	Schüler B gehört zu den schwächeren Mathematik-Schülern in Lehrer Hs Klasse. In einer Stillarbeitsphase geht Lehrer H durch die Reihen und bemerkt Bs Arbeit. Lehrer H sagt zu B: „Das machst du richtig.	1	2	3	4	5	6
53.	Familie M kommt bezüglich ihrer Tochter T zum Beratungsgespräch zu Lehrer F. Lehrer F sagt: „Ich stelle Ihnen einige Lernstrategien vor, mit denen T Erfolg haben kann.“	1	2	3	4	5	6
54.	Einige Schüler von Lehrer Ks zehnter Klasse möchten eine Klassenzeitung gründen und berichten Lehrer K davon. Lehrer K sagt: „Ich berate euch bei eurer Arbeit.“	1	2	3	4	5	6
55.	Lehrerin C hat die achte Klasse ihres Gymnasiums in das Prinzip der linearen Funktionen eingeführt. Dann sagt Lehrerin C: „Jeder rechnet nun bitte bis zum Ende der Stunde die Beispiele 1 bis 5	1	2	3	4	5	6
56.	Die LehrerInnen einer Schule diskutieren über die rechtlichen Rahmenbedingungen des Lehrberufs. Lehrerin W sagt: „Ich kenne die Rahmenbedingungen nicht genau, da es	1	2	3	4	5	6

**Für wie effektiv halten Sie das Verhalten/die Überlegungen der jeweiligen Lehrperson?**

1=sehr ineffektiv 2= ineffektiv 3=eher ineffektiv 4=eher effektiv 5= effektiv 6=sehr effektiv  
Bitte kreuzen Sie die Ziffer an, die Ihrer Einschätzung entspricht.

57.	Die Leistungen von Lehrer Bs Klasse sinken, und die Schüler verhalten sich ihm gegenüber respektlos. Lehrer B überlegt, was er tun könnte. Lehrer B denkt: „Ich frage bei meinen	1	2	3	4	5	6
58.	Eine Kollegin von Lehrer H lässt ihre SchülerInnen ihren Unterricht evaluieren. Lehrer H überlegt, ob er dies auch tun soll. Lehrer H sagt: „Die Leistungen der SchülerInnen reichen aus, im daraus	1	2	3	4	5	6
59.	Lehrer Z stellt seiner Klasse eine Frage und erhält von einem Schüler eine falsche Antwort. Lehrer Z sagt: „Nein, das stimmt nicht. Wer kann erklären, warum?“	1	2	3	4	5	6
60.	Lehrerin B übt mit ihrer Klasse das Übersetzen aus dem Englischen ins Deutsche. Sie möchte die SchülerInnen motivieren und vergibt die Hausaufgaben für den nächsten Tag. Sie sagt: „Übersetzt bis morgen den Text auf	1	2	3	4	5	6
61.	Schüler P scheint im Unterricht niedergeschlagen und abwesend, ohne dass ein offensichtlicher Grund vorliegt. Lehrer R spricht dies in der Pause an. Lehrer R sagt: „Möchtest Du einmal mit mir	1	2	3	4	5	6
62.	Schüler B gehört zu den schwächeren Mathematik-Schülern in Lehrer Hs Klasse. In einer Stillarbeitsphase geht Lehrer H durch die Reihen und bemerkt Bs Arbeit. Lehrer H sagt zu B: „Das machst du richtig.	1	2	3	4	5	6
63.	Lehrerin D kommt mit Klasse 7a sehr gut zurecht, mit Klasse 9b hat sie leider immer wieder Probleme. Lehrerin D denkt: „Schade, dass Klasse 9b nicht wie Klasse 7a ist.“	1	2	3	4	5	6

**Für wie effektiv halten Sie das Verhalten/die Überlegungen der jeweiligen Lehrperson?**

1=sehr ineffektiv 2= ineffektiv 3=eher ineffektiv 4=eher effektiv 5= effektiv 6=sehr effektiv  
Bitte kreuzen Sie die Ziffer an, die Ihrer Einschätzung entspricht.

64.	Lehrerin N hat ihren SchülerInnen einen Fragebogen zur Unterrichtsevaluation bearbeiten lassen und betrachtet nun die Ergebnisse. Lehrerin N denkt: „Ich werde die Bereiche	1	2	3	4	5	6
65.	Schülerin G stört häufig den Unterricht von Lehrer T und gehört zu den eher schwachen Schülern. Lehrer T überlegt, wie er weiter vorgehen soll. Lehrer T denkt: „Ich werde strenger sein	1	2	3	4	5	6
66.	Lehrerin E möchte ihren Schüler S beurteilen. Lehrerin E denkt: „Die schriftlichen Leistungen von S waren gut, allerdings hat er mündlich kaum mitgearbeitet und kein Referat gehalten. Ich werde das in mein Urteil mit einbeziehen.“	1	2	3	4	5	6
67.	An der Schule von Lehrer Y sollen die LehrerInnen stark an der Schulentwicklung beteiligt werden. Lehrer Y denkt: „Ich nehme gern teil, da ich es wichtig finde, die Schule mit gestalten zu	1	2	3	4	5	6
68.	Lehrerin möchte, dass ihre Schüler das Prinzip des Dividierens gut verstehen. Lehrerin A sagt: „Jonas hat einen Freund zum Kuchenessen eingeladen. Vier Kuchenstücke hat er noch. Wie viele Kuchenstücke bekommt	1	2	3	4	5	6
69.	An einer benachbarten Schule werden seit einiger Zeit immer wieder SchülerInnen von anderen gewaltsam angegriffen. Lehrer Bs Klasse ist nicht direkt betroffen. Lehrer B denkt: „Da unsere Schule nicht	1	2	3	4	5	6
70.	Lehrer X hat Schüler Bs Deutsch-Klassenarbeit korrigiert. Lehrer X schreibt einen kurzen Kommentar unter die mit 3,5 benotete Arbeit. Lehrer X schreibt: „Du befindest dich im	1	2	3	4	5	6



**Für wie effektiv halten Sie das Verhalten/die Überlegungen der jeweiligen Lehrperson?**

1=sehr ineffektiv 2= ineffektiv 3=eher ineffektiv 4=eher effektiv 5= effektiv 6=sehr effektiv  
Bitte kreuzen Sie die Ziffer an, die Ihrer Einschätzung entspricht.

71.	Die Leistungen von Lehrer Bs Klasse sinken, und die Schüler verhalten sich ihm gegenüber respektlos. Lehrer B überlegt, was er tun könnte. Lehrer B denkt: „Ich möchte nicht, dass meine	1	2	3	4	5	6
72.	Schüler T wird in Lehrerin As siebter Klasse ausgegrenzt und lächerlich gemacht. Lehrer A möchte dies thematisieren. Lehrerin A sagt: „In Klasse X wird eine Schülerin von ihren Kameraden ausgegrenzt.	1	2	3	4	5	6
73.	Schüler P hat bei Klassenarbeiten von Lehrer J wiederholt nicht gut abgeschnitten. Lehrer J sagt: „Beim nächsten Mal sollte es besser werden.“	1	2	3	4	5	6
74.	Lehrer P kündigt eine Klassenarbeit an. Er möchte die SchülerInnen auf das Lernen vorbereiten. Lehrer P sagt: „Es geht um eine ganze Menge Stoff, also besprechen wir nun, wie ihr am	1	2	3	4	5	6
75.	Eine Kollegin von Lehrer H lässt ihre SchülerInnen ihren Unterricht evaluieren. Lehrer H überlegt, ob er dies auch tun soll. Lehrer H sagt: „Ja, denn mir ist es wichtig zu wissen, was ich aus Schülersicht verbessern	1	2	3	4	5	6
76.	Lehrer P plant die Themen für seine Physik-AG. Lehrer P denkt: „Ich kann die Materialien vom letzten Jahr wieder verwenden.“	1	2	3	4	5	6
77.	Lehrerin B übt mit ihrer Klasse das Übersetzen aus dem Englischen ins Deutsche. Sie möchte die SchülerInnen motivieren und vergibt die Hausaufgaben für den nächsten Tag. Sie sagt: „Übersetzt bis morgen einen Song	1	2	3	4	5	6

**Für wie effektiv halten Sie das Verhalten/die Überlegungen der jeweiligen Lehrperson?**

1=sehr ineffektiv 2= ineffektiv 3=eher ineffektiv 4=eher effektiv 5= effektiv 6=sehr effektiv  
Bitte kreuzen Sie die Ziffer an, die Ihrer Einschätzung entspricht.

78.	Lehrer G plant den Unterricht in Klasse 11c. Diese Klasse kennt er noch nicht. Lehrer G denkt: „Zur Unterrichtsplanung orientiere ich mich an didaktischem Wissen.“	1	2	3	4	5	6
79.	Lehrerin E bespricht zu Schuljahrsbeginn in ihrer Klasse den Englisch-Unterricht der nächsten Monate. Lehrerin E sagt: „Ich bin zufrieden, wenn wir gut vorankommen.“	1	2	3	4	5	6
80.	Lehrer Ws Klasse hat seine hochgesteckten Ziele in diesem Schuljahr nicht erreicht. Lehrer W denkt: „Die Klasse muss sich mehr anstrengen.“	1	2	3	4	5	6
81.	Lehrerin Q erwartet von ihren SchülerInnen Pünktlichkeit. Sie selbst kommt häufig zu spät, worauf sie die SchülerInnen ansprechen. Lehrerin Q sagt: „Das lässt sich leider nicht vermeiden, da ich sehr viel zu tun habe.“	1	2	3	4	5	6
82.	Lehrer C hat einen wissenschaftlichen Artikel über effektive allgemeine Unterrichtsmethoden gelesen. Lehrer C denkt: „Dies werde ich auf jeden Fall umsetzen.“	1	2	3	4	5	6
83.	Lehrerin As Klasse schneidet in einem standardisierten Rechentest eher schlecht ab. Lehrerin A überlegt, was sie tun sollte. Lehrerin A denkt: „Ich werde problematische Aufgabenstellungen noch einmal behandeln.“	1	2	3	4	5	6
84.	In Lehrer Zs Klasse haben sich zwei Schülergruppen gebildet, die sich feindlich gegenüber stehen. Lehrer Z thematisiert die Situation. Lehrer Z sagt: „In der Klasse haben sich zwei	1	2	3	4	5	6

**Für wie effektiv halten Sie das Verhalten/die Überlegungen der jeweiligen Lehrperson?**

1=sehr ineffektiv 2= ineffektiv 3=eher ineffektiv 4=eher effektiv 5= effektiv 6=sehr effektiv  
Bitte kreuzen Sie die Ziffer an, die Ihrer Einschätzung entspricht.

85.	Schülerin J spricht in den Pausen häufig begeistert von der Kfz-Werkstatt ihres Onkels. Lehrer B möchte dieses Interesse fördern. Lehrer B sagt: „Möchtest du an der Technik-AG teilnehmen?“	1	2	3	4	5	6
86.	Lehrerin G übt mit den Schülern ihrer siebten Klasse das Verstehen von Sachtexten. Sie gibt der Klasse Arbeitsanweisungen zu einem neuen Text. Lehrerin G sagt: „Lest bitte den Text einmal	1	2	3	4	5	6
87.	Lehrerin A plant den Unterricht für die Klassen 5c und 8a. Sie überlegt dabei, welche Methoden sie anwenden soll. Lehrerin A denkt: „Welche Methoden und Medien haben mir in meiner Schulzeit gut	1	2	3	4	5	6
88.	Schüler B zeigt im Vergleich zur Klasse eher schwache Leistungen. In einer Klassenarbeit schneidet B nicht gut ab. Lehrer I hat das erwartet. Lehrer I sagt zu B: „Schade, das habe ich	1	2	3	4	5	6
89.	Lehrer A und Lehrer J diskutieren über ihre Wertvorstellungen bezüglich ihres Berufs. Beide vertreten unterschiedliche Meinungen. Lehrer J denkt: „Ich denke darüber nach, ob meine Werte die richtigen sind.“	1	2	3	4	5	6
90.	Lehrerin S hat von einer neuartigen Unterrichtsmethode gehört, die interessant zu sein scheint. Lehrerin S denkt: „Ich werde ein Seminar zu dem Thema besuchen.“	1	2	3	4	5	6
91.	Lehrerin M plant eine Projektwoche mit ihrer Klasse. Sie überlegt, wie sie das Projekt gestalten soll. Lehrerin M denkt: „Ich lese nach, welche Forschungsergebnisse es zur Gestaltung von	1	2	3	4	5	6

**Für wie effektiv halten Sie das Verhalten/die Überlegungen der jeweiligen Lehrperson?**

1=sehr ineffektiv 2= ineffektiv 3=eher ineffektiv 4=eher effektiv 5= effektiv 6=sehr effektiv  
Bitte kreuzen Sie die Ziffer an, die Ihrer Einschätzung entspricht.

92.	Lehrer O möchte Umgangsregeln in seiner Klasse einrichten. Lehrer O sagt: „Ich habe einen Regelkatalog ausgearbeitet, an den ihr euch halten sollt.“	1	2	3	4	5	6
93.	Lehrer K möchte die SchülerInnen seiner Klasse beurteilen. Lehrer K denkt: „Ich kenne diese Klasse schon lange, deshalb kann ich mich ganz auf mein Urteil verlassen.“	1	2	3	4	5	6
94.	Die LehrerInnen einer Schule diskutieren über die rechtlichen Rahmenbedingungen des Lehrberufs. Lehrerin W sagt: „Es ist wichtig, die rechtlichen Rahmenbedingungen zu kennen, um sicher	1	2	3	4	5	6
95.	Lehrerin M plant ein großes Schulprojekt und fragt Lehrerin L, ob sie mitarbeiten möchte. Lehrerin L sagt: „Ich nehme nicht teil. Ich möchte die Zeit lieber in die Gestaltung meines Unterrichts investieren.“	1	2	3	4	5	6
96.	Lehrer K möchte im Physik-Unterricht die Elektrik-Kenntnisse seiner SchülerInnen vertiefen. Lehrer K sagt: „Ich erkläre euch nun, wie ein Radio funktioniert.“	1	2	3	4	5	6
97.	Lehrer P kündigt eine Klassenarbeit an. Er möchte die SchülerInnen auf das Lernen vorbereiten. Lehrer P sagt: „Es geht um eine ganze Menge Stoff, also müsst ihr viel lernen.“	1	2	3	4	5	6
98.	Ein Schüler hat bei einer Klassenarbeit schlecht abgeschnitten. Lehrer H spricht mit ihm über das Ergebnis. Lehrer H sagt: „Schade, dass du nicht gut abgeschnitten hast. Dir liegt das Fach	1	2	3	4	5	6

**Für wie effektiv halten Sie das Verhalten/die Überlegungen der jeweiligen  
Lehrperson?**

1=sehr ineffektiv 2= ineffektiv 3=eher ineffektiv 4=eher effektiv 5= effektiv 6=sehr effektiv  
Bitte kreuzen Sie die Ziffer an, die Ihrer Einschätzung entspricht.

99.	Lehrerin S möchte hilfsbereites Verhalten ihrer SchülerInnen fördern. Lehrerin S sagt: „Ihr sollt euch hilfsbereit verhalten.“	1	2	3	4	5	6
100.	Lehrer Z möchte in der nächsten Stunde in ein neues Thema einführen. Er überlegt, wie er vorgehen soll. Lehrer Z denkt: „Ich gebe den SchülerInnen ein Problem vor und lasse sie nach einer	1	2	3	4	5	6
101.	Die Schüler B und C hinken im Mathematikunterricht etwas hinterher und haben insbesondere Schwierigkeiten mit Textaufgaben. Lehrerin K gibt einen Arbeitsauftrag.	1	2	3	4	5	6
102.	Lehrer C hat einen wissenschaftlichen Artikel über effektive allgemeine Unterrichtsmethoden gelesen. Lehrer C denkt: „Dies lässt sich in meine Praxis nicht übertragen.“	1	2	3	4	5	6

**Herzlichen Dank für Ihre Mitarbeit!**

## Anhang I: Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	$r_{it}$	$p^{**}$
1: Lehrerinnen und Lehrer planen Unterricht fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.  ( $\alpha = .435$ )	<i>Lehrerin B beginnt den Englisch-Unterricht in ihrer Klasse. Lehrerin B sagt: „Wir behandeln heute das ‚future perfect‘. Wir werden erst den Stoff von gestern wiederholen, dann gemeinsam arbeiten und zuletzt die Stunde zusammenfassen.“</i>	4,12	1,202	.165	.624
1: Lehrerinnen und Lehrer planen Unterricht fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.  ( $\alpha = .435$ )	<i>Lehrer B beginnt den Englisch-Unterricht in seiner Klasse. Lehrer B sagt: „Wir werden heute gut voran kommen. Lisa, was war die Hausaufgabe? Martin hat noch eine Frage zum gestrigen Stoff? Dann stelle sie sofort.“</i>	4,06	1,179	.188	.612
1: Lehrerinnen und Lehrer planen Unterricht fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.  ( $\alpha = .435$ )	<i>Lehrer G plant den Unterricht in Klasse 11c. Diese Klasse kennt er noch nicht. Lehrer G denkt: „Zur Unterrichtsplanung orientiere ich mich an den Erfahrungen in meiner eigenen Schulzeit, da ich über die Methoden, die ich selbst erlebt habe, das beste Urteil treffen kann.“</i>	3,75	1,182	.028	.550

Mittelwerte, Standardabweichungen, Trennschärpen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
1: Lehrerinnen und Lehrer planen Unterricht fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.  ( $\alpha = .435$ )	<i>Lehrerin A plant den Unterricht für die Klassen 5c und 8a. Sie überlegt dabei, welche Methoden sie anwenden soll. Lehrerin A denkt: „Welche Methoden und Medien eigenen sich für welche Klasse?“</i>	4,86	1,077	.256	.772
1: Lehrerinnen und Lehrer planen Unterricht fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.  ( $\alpha = .435$ )	<i>Lehrerin E lässt die Schüler in der 10. Klasse ihrer Realschule Aufgaben in Kleingruppen bearbeiten. Als sie durch die Reihen geht, bemerkt sie, dass die Aufgaben kaum gelöst werden. Lehrerin E sagt: „Arbeitet bitte konzentriert weiter.“</i>	4,79	0,994	.262	.758
1: Lehrerinnen und Lehrer planen Unterricht fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.  ( $\alpha = .435$ )	<i>Lehrerin E lässt die Schüler in der 10. Klasse ihrer Realschule Aufgaben in Kleingruppen bearbeiten. Als sie durch die Reihen geht, bemerkt sie, dass die Aufgaben kaum gelöst werden. Lehrerin E sagt: „Notiert euch aufkommende Fragen, die wir dann zusammen besprechen.“</i>	4,51	1,108	.265	.702

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
<b>1: Lehrerinnen und Lehrer planen Unterricht fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.</b>  ( $\alpha = .435$ )	<i>Lehrer G plant den Unterricht in Klasse 11c. Diese Klasse kennt er noch nicht.</i> <i>Lehrer G denkt: „Zur Unterrichtsplanung orientiere ich mich an didaktischem Wissen.“</i>	3,97	0,912	.217	.594
<b>1: Lehrerinnen und Lehrer planen Unterricht fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.</b>  ( $\alpha = .435$ )	<i>Lehrerin A plant den Unterricht für die Klassen 5c und 8a. Sie überlegt dabei, welche Methoden sie anwenden soll.</i> <i>Lehrerin A denkt: „Welche Methoden und Medien haben mir in meiner Schulzeit gut gefallen?“</i>	3,23	1,110	.028	.446
<b>2: Lehrerinnen und Lehrer unterstützen durch die Gestaltung von Lernsituationen das Lernen von Schülerinnen und Schülern. Sie motivieren Schüler und befähigen sie, Zusammenhänge herzustellen und Gelerntes zu verknüpfen.</b>  ( $\alpha = .548$ )	<i>Lehrer Z stellt seiner Klasse eine Frage und erhält von einem Schüler eine falsche Antwort.</i> <i>Lehrer Z sagt: „Nein, das stimmt nicht. Wer kann es richtig sagen?“</i>	4,50	1,177	.303	.700



Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
2: Lehrerinnen und Lehrer unterstützen durch die Gestaltung von Lernsituationen das Lernen von Schülerinnen und Schülern. Sie motivieren Schüler und befähigen sie, Zusammenhänge herzustellen und Gelerntes zu verknüpfen.  ( $\alpha = .548$ )	Lehrerin A möchte, dass ihre Schüler das Prinzip des Dividierens gut verstehen. Lehrerin A sagt: „Rechnet bitte: 10 durch 2.“	4,59	1,084	.360	.788
	Lehrer K möchte im Physik-Unterricht die Elektrik-Kenntnisse seiner SchülerInnen vertiefen. Lehrer K sagt: „Wir wollen zusammen ein Radio bauen. Was brauchen wir dafür?“	4,24	1,268	.188	.848
2: Lehrerinnen und Lehrer unterstützen durch die Gestaltung von Lernsituationen das Lernen von Schülerinnen und Schülern. Sie motivieren Schüler und befähigen sie, Zusammenhänge herzustellen und Gelerntes zu verknüpfen.  ( $\alpha = .548$ )					

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
<b>2: Lehrerinnen und Lehrer unterstützen durch die Gestaltung von Lernsituationen das Lernen von Schülerinnen und Schülern. Sie motivieren Schüler und befähigen sie, Zusammenhänge herzustellen und Gelerntes zu verknüpfen.</b>  (α= .548)	<i>Lehrerin C hat die achte Klasse ihres Gymnasiums in das Prinzip der linearen Funktionen eingeführt.</i>				
	<i>Dann sagt Lehrerin C: „Ich rechne eine Aufgabe vor, dann rechnen wir gemeinsam mit einem Schüler an der Tafel.“</i>	3,77	1,035	.046	.554
<b>2: Lehrerinnen und Lehrer unterstützen durch die Gestaltung von Lernsituationen das Lernen von Schülerinnen und Schülern. Sie motivieren Schüler und befähigen sie, Zusammenhänge herzustellen und Gelerntes zu verknüpfen.</b>  (α= .548)	<i>Lehrer Z stellt seiner Klasse eine Frage und erhält von einem Schüler eine falsche Antwort.</i>				
	<i>Lehrer Z sagt: „Nein, das stimmt nicht. Wer kann erklären, warum?“</i>	4,10	1,076	.034	.620

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
2: Lehrerinnen und Lehrer unterstützen durch die Gestaltung von Lernsituationen das Lernen von Schülerinnen und Schülern. Sie motivieren Schüler und befähigen sie, Zusammenhänge herzustellen und Gelerntes zu verknüpfen.  ( $\alpha = .548$ )	<i>Lehrerin C hat die achte Klasse ihres Gymnasiums in das Prinzip der linearen Funktionen eingeführt. Dann sagt Lehrerin C: „Jeder rechnet nun bitte bis zum Ende der Stunde die Beispiele 1 bis 5 aus dem Buch.“</i>	4,08	1,099	.275	.616
2: Lehrerinnen und Lehrer unterstützen durch die Gestaltung von Lernsituationen das Lernen von Schülerinnen und Schülern. Sie motivieren Schüler und befähigen sie, Zusammenhänge herzustellen und Gelerntes zu verknüpfen.  ( $\alpha = .548$ )	<i>Lehrerin B übt mit ihrer Klasse das Übersetzen aus dem Englischen ins Deutsche. Sie möchte die SchülerInnen motivieren und vergibt die Hausaufgaben für den nächsten Tag. Sie sagt: „Übersetzt bis morgen den Text auf Seite 46 ins Deutsche.“</i>	3,97	1,050	.102	.594

**Mittelwerte, Standardabweichungen, Trennschärpen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests**

<b>Skala (Cronbachs Alpha)</b>	<b>Item*</b>	<b>M</b>	<b>SD</b>	<b>r<sub>it</sub></b>	<b>p**</b>
<b>2: Lehrerinnen und Lehrer unterstützen durch die Gestaltung von Lernsituationen das Lernen von Schülerinnen und Schülern. Sie motivieren Schüler und befähigen sie, Zusammenhänge herzustellen und Gelerntes zu verknüpfen.</b>  (α= .548)	Lehrerin möchte, dass ihre Schüler das Prinzip des Dividierens gut verstehen. Lehrerin A sagt: „Jonas hat einen Freund zum Kuchenessen eingeladen. Vier Kuchenstücke hat er noch. Wie viele Kuchenstücke bekommt jeder der beiden?“	5,10	0,983	.570	.820
	Lehrerin B übt mit ihrer Klasse das Übersetzen aus dem Englischen ins Deutsche. Sie möchte die SchülerInnen motivieren und vergibt die Hausaufgaben für den nächsten Tag. Sie sagt: „Übersetzt bis morgen einen Song ins Deutsche.“	4,78	1,035	.399	.756
(α= .548)					

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
2: Lehrerinnen und Lehrer unterstützen durch die Gestaltung von Lernsituationen das Lernen von Schülerinnen und Schülern. Sie motivieren Schüler und befähigen sie, Zusammenhänge herzustellen und Gelerntes zu verknüpfen.  (α= .548)	Lehrer K möchte im Physik-Unterricht die Elektrik-Kenntnisse seiner SchülerInnen vertiefen. Lehrer K sagt: „Ich erkläre euch nun, wie ein Radio funktioniert.“	3,77	1,133	.360	.554
3: Lehrerinnen und Lehrer fördern die Fähigkeiten von Schülerinnen und Schülern zum selbstbestimmten Lernen und Arbeiten.  (α= .273)	Lehrer Z möchte in der nächsten Stunde in ein neues Thema einführen. Er überlegt, wie er vorgehen soll. Lehrer Z denkt: „Ich beginne mit einem Vortrag, der das Thema darstellt.“	3,82	1,194	.056	.564
3: Lehrerinnen und Lehrer fördern die Fähigkeiten von Schülerinnen und Schülern zum selbstbestimmten Lernen und Arbeiten.  (α= .273)	Lehrerin G übt mit den Schülern ihrer siebten Klasse das Verstehen von Sachtexten. Sie gibt der Klasse Arbeitsanweisungen. Lehrerin G sagt: „Lest den Text durch und streicht das Wichtigste an. Dann sammeln wir in Kleingruppen wichtige Punkte und jeder fertigt ein Mind-Map an.“	4,38	1,084	.170	.676

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
3: Lehrerinnen und Lehrer fördern die Fähigkeiten von Schülerinnen und Schülern zum selbstbestimmten Lernen und Arbeiten. (α= .273)	<i>Ein Schüler hat bei einer Klassenarbeit schlecht abgeschnitten. Lehrer H spricht mit ihm über das Ergebnis. Lehrer H sagt: „Du hast nur wenig gelernt? Dann konntest du natürlich die Aufgaben nicht lösen.“</i>	2,68	0,998	- .402	.336
3: Lehrerinnen und Lehrer fördern die Fähigkeiten von Schülerinnen und Schülern zum selbstbestimmten Lernen und Arbeiten. (α= .273)	<b><i>Ein Schüler hat bei einer Klassenarbeit schlecht abgeschnitten. Lehrer H spricht mit ihm über das Ergebnis. Lehrer H sagt: „Schade, dass du nicht gut abgeschnitten hast. Dir liegt das Fach vielleicht einfach nicht.“</i></b>	4,88	1,072	.289	.776
3: Lehrerinnen und Lehrer fördern die Fähigkeiten von Schülerinnen und Schülern zum selbstbestimmten Lernen und Arbeiten. (α= .273)	<i>Lehrer Z möchte in der nächsten Stunde in ein neues Thema einführen. Er überlegt, wie er vorgehen soll. Lehrer Z denkt: „Ich gebe den SchülerInnen ein Problem vor und lasse sie nach einer Lösung suchen.“</i>	4,34	0,954	.252	.668
3: Lehrerinnen und Lehrer fördern die Fähigkeiten von Schülerinnen und Schülern zum selbstbestimmten Lernen und Arbeiten. (α= .273)	<i>Schüler P hat bei Klassenarbeiten von Lehrer J wiederholt nicht gut abgeschnitten. Lehrer J sagt: „Ich weiß, dass du es besser kannst.“</i>	3,55	1,158	- .050	.510

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
3: Lehrerinnen und Lehrer fördern die Fähigkeiten von Schülerinnen und Schülern zum selbstbestimmten Lernen und Arbeiten.  (α= .273)	<i>Schüler P hat bei Klassenarbeiten von Lehrer J wiederholt nicht gut abgeschnitten. Lehrer J sagt: „Beim nächsten Mal sollte es besser werden.“</i>	4,23	0,922	.066	.646
3: Lehrerinnen und Lehrer fördern die Fähigkeiten von Schülerinnen und Schülern zum selbstbestimmten Lernen und Arbeiten.  (α= .273)	Lehrer P kündigt eine Klassenarbeit an. Er möchte die SchülerInnen auf das Lernen vorbereiten. Lehrer P sagt: „Es geht um eine ganze Menge Stoff, also besprechen wir nun, wie ihr am besten lernen könnt.“	4,70	1,022	.416	.740
3: Lehrerinnen und Lehrer fördern die Fähigkeiten von Schülerinnen und Schülern zum selbstbestimmten Lernen und Arbeiten.  (α= .273)	<i>Lehrerin G übt mit den Schülern ihrer siebten Klasse das Verstehen von Sachtexten. Sie gibt der Klasse Arbeitsanweisungen zu einem neuen Text. Lehrerin G sagt: „Lest bitte den Text einmal durch. Dann werde ich euch Fragen zum Text stellen.“</i>	3,33	,930	.028	.466
3: Lehrerinnen und Lehrer fördern die Fähigkeiten von Schülerinnen und Schülern zum selbstbestimmten Lernen und Arbeiten.  (α= .273)	<i>Lehrer P kündigt eine Klassenarbeit an. Er möchte die SchülerInnen auf das Lernen vorbereiten. Lehrer P sagt: „Es geht um eine ganze Menge Stoff, also müsst ihr viel lernen.“</i>	4,88	1,029	.277	.776

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
4: Lehrerinnen und Lehrer kennen die sozialen und kulturellen Lebensbedingungen von Schülerinnen und Schülern und nehmen im Rahmen der Schule Einfluss auf deren individuelle Entwicklung.  (α= .373)	<i>In der Klasse 9d werden immer wieder Bemerkungen laut wie: „Alle Muslime sind Terroristen – vor denen muss man Angst haben.“ Lehrer F möchte die Vorurteile abbauen. Lehrer F sagt: „Ich erlaube solche Bemerkungen nicht.“</i>	4,42	1,342	.126	.684
4: Lehrerinnen und Lehrer kennen die sozialen und kulturellen Lebensbedingungen von Schülerinnen und Schülern und nehmen im Rahmen der Schule Einfluss auf deren individuelle Entwicklung.  (α= .373)	<i>Die Schüler B und C hinken im Mathematikunterricht der 9d etwas hinterher und haben insbesondere Schwierigkeiten mit Textaufgaben. Lehrerin K gibt einen Arbeitsauftrag. Lehrerin K sagt: „Bitte bearbeitet die Aufgaben 2-4 in Kleingruppen. B und C, ihr hört dabei bitte genau zu.“</i>	4,20	1,182	.147	.640



Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
4: Lehrerinnen und Lehrer kennen die sozialen und kulturellen Lebensbedingungen von Schülerinnen und Schülern und nehmen im Rahmen der Schule Einfluss auf deren individuelle Entwicklung.  (α= .373)	In der Klasse 9d werden immer wieder Bemerkungen laut wie: „Alle Muslime sind Terroristen – vor denen muss man Angst haben.“ Lehrer F möchte die Vorurteile abbauen. Lehrer F sagt: „Sammelt bis morgen Informationen über die muslimische Kultur, die wir dann gemeinsam besprechen.“	4,92	1,157	.359	.784
	<i>Schülerin J spricht in den Pausen häufig begeistert von der Kfz-Werkstatt ihres Onkels. Lehrer B möchte dieses Interesse fördern. Lehrer B sagt: „Welches ist denn dein Lieblingsauto?“</i>	2,98	1,097	- .139	.396
(α= .373)					

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
<b>4: Lehrerinnen und Lehrer kennen die sozialen und kulturellen Lebensbedingungen von Schülerinnen und Schülern und nehmen im Rahmen der Schule Einfluss auf deren individuelle Entwicklung.</b>  ( $\alpha = .373$ )	<i>Die Schüler B und C hinken im Mathematikunterricht etwas hinterher und haben insbesondere Schwierigkeiten mit Textaufgaben. Lehrerin K gibt einen Arbeitsauftrag. Lehrerin K sagt: „Bitte bearbeitet die Aufgaben 2-4 in Kleingruppen. B und C, ich rechne noch einmal Aufgabe 1 mit euch zusammen.“</i>	4,53	1,081	.261	.706
<b>4: Lehrerinnen und Lehrer kennen die sozialen und kulturellen Lebensbedingungen von Schülerinnen und Schülern und nehmen im Rahmen der Schule Einfluss auf deren individuelle Entwicklung.</b>  ( $\alpha = .373$ )	Schülerin J spricht in den Pausen häufig begeistert von der Kfz-Werkstatt ihres Onkels. Lehrer B möchte dieses Interesse fördern. Lehrer B sagt: „Möchtest du an der Technik-AG teilnehmen?“	5,00	1,038	.353	.800
<b>5: Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern.</b>  ( $\alpha = .709$ )	Ein Wandertag ist in Klasse 9d geplant. Lehrer R bespricht mit den SchülerInnen, wo es hingehen soll. Lehrer R sagt: „Wir sammeln einmal Vorschläge, wo wir hinfahren könnten.“	4,72	1,100	.450	.744

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
5: Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern. (α= .709)	<i>Schüler P scheint im Unterricht niedergeschlagen und abwesend, ohne dass ein offensichtlicher Grund vorliegt. Lehrer M spricht dies in der Pause an. Lehrer M sagt: „Was lenkt dich denn ab? Versuche bitte, dich trotzdem zu konzentrieren.“</i>	3,69	1,099	.021	.538
5: Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern. (α= .709)	<i>Lehrerin S möchte hilfsbereites Verhalten ihrer SchülerInnen fördern. Lehrerin S sagt: „Wer schnell mit seiner Aufgabe fertig ist, fragt bitte den Nachbarn, ob er ihn unterstützen kann.“</i>	4,14	1,200	.155	.628
5: Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern. (α= .709)	<i>Lehrer U möchte seinen Schülern Werte und Normen vermitteln und überlegt, wie er dies umsetzen könnte. Lehrer U denkt: „Ich stelle Regeln auf.“</i>	3,92	1,190	.228	.584
5: Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern. (α= .709)	<i>Lehrer O möchte Umgangsregeln in seiner Klasse einrichten. Lehrer O sagt: „Wir erarbeiten gemeinsam einen Regelkatalog und vereinbaren, was geschehen soll, wenn die Regeln nicht eingehalten werden.“</i>	5,01	1,083	.567	.802

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
5: Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern.  ( $\alpha = .709$ )	Ein Wandertag ist in Klasse 9d geplant. Lehrer I bespricht mit den SchülerInnen, wo es hingehen soll. Lehrer R sagt: „Wir fahren zur Burg, das gefällt euch sicher am besten.“	4,96	1,014	.492	.792
5: Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern.  ( $\alpha = .709$ )	Lehrer U möchte seinen Schülern Werte und Normen vermitteln und überlegt, wie er dies umsetzen könnte. Lehrer U denkt: „Ich gebe den Schülern ein moralisches Dilemma vor, über das wir dann diskutieren.“	4,61	1,025	.439	.722
5: Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern.  ( $\alpha = .709$ )	Schüler P scheint im Unterricht niedergeschlagen und abwesend, ohne dass ein offensichtlicher Grund vorliegt. Lehrer R spricht dies in der Pause an. Lehrer R sagt: „Möchtest Du einmal mit mir darüber sprechen, was dich bedrückt?“	5,03	0,996	.528	.806
5: Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern.  ( $\alpha = .709$ )	Lehrer O möchte Umgangsregeln in seiner Klasse einrichten. Lehrer O sagt: „Ich habe einen Regelkatalog ausgearbeitet, an den ihr euch halten sollt.“	4,68	1,118	.492	.736

**Mittelwerte, Standardabweichungen, Trennschärpen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests**

<b>Skala (Cronbachs Alpha)</b>	<b>Item*</b>	<b>M</b>	<b>SD</b>	<b>r<sub>it</sub></b>	<b>p**</b>
<b>5: Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern.</b>	<b>Lehrerin S möchte hilfsbereites Verhalten ihrer SchülerInnen fördern.</b>	4,40	1,109	.319	.680
( $\alpha = .709$ )					
<b>6: Lehrerinnen und Lehrer finden Lösungsansätze für Schwierigkeiten und Konflikte in Schule und Unterricht.</b>	An einer benachbarten Schule werden seit einiger Zeit immer wieder SchülerInnen von anderen gewaltsam angegriffen. Lehrer Bs Klasse ist nicht direkt betroffen.	4,72	1,101	.389	.744
( $\alpha = .603$ )	Lehrer B denkt: „Ich lasse die Schüler über die möglichen Ursachen von Gewalt sowie die Lage der Opfer diskutieren.“				
<b>6: Lehrerinnen und Lehrer finden Lösungsansätze für Schwierigkeiten und Konflikte in Schule und Unterricht.</b>	<b>Lehrerin S plant den Deutsch-Unterricht in ihrer achten Klasse und überlegt, welche Methoden sie vor allem anwenden soll.</b>	4,33	1,015	.403	.666
( $\alpha = .603$ )	Lehrerin S denkt: „Ich lasse am Besten die Schüler in Stillarbeit Aufgaben lösen.“				

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
6: Lehrerinnen und Lehrer finden Lösungsansätze für Schwierigkeiten und Konflikte in Schule und Unterricht. ( $\alpha = .603$ )	<i>In Lehrer Ps Klasse haben sich zwei Schülergruppen gebildet, die sich feindlich gegenüber stehen. Lehrer P thematisiert die Situation. Lehrer P sagt: „In der Klasse haben sich zwei Lager gebildet, die sich bekämpfen. Dieses Verhalten kann ich nicht dulden und werde Sanktionen aussprechen.“</i>	4,08	1,127	.174	.616
6: Lehrerinnen und Lehrer finden Lösungsansätze für Schwierigkeiten und Konflikte in Schule und Unterricht. ( $\alpha = .603$ )	<i>Schüler T wird in Lehrer As sechster Klasse ausgegrenzt und lächerlich gemacht. Lehrer A möchte thematisieren. Lehrer A sagt: „Heute sprechen wir über euer unfaires Verhalten T gegenüber.“</i>	4,05	1,280	.174	.610
6: Lehrerinnen und Lehrer finden Lösungsansätze für Schwierigkeiten und Konflikte in Schule und Unterricht. ( $\alpha = .603$ )	<i>Lehrerin L plant den Deutsch-Unterricht in ihrer achten Klasse und überlegt, welche Methoden sie vorwiegend anwenden soll. Lehrerin L denkt: „Ich lasse am Besten die Schüler in Gruppen Aufgaben diskutieren.“</i>	3,90	,904	.240	.580

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
6: Lehrerinnen und Lehrer finden Lösungsansätze für Schwierigkeiten und Konflikte in Schule und Unterricht. ( $\alpha = .603$ )	An einer benachbarten Schule werden seit einiger Zeit immer wieder SchülerInnen von anderen gewaltsam angegriffen. Lehrer Bs Klasse ist nicht direkt betroffen. Lehrer B denkt: „Da unsere Schule nicht betroffen ist, möchte ich das Thema im Moment nicht ansprechen.“	5,35	1,042	.594	.870
6: Lehrerinnen und Lehrer finden Lösungsansätze für Schwierigkeiten und Konflikte in Schule und Unterricht. ( $\alpha = .603$ )	Schüler T wird in Lehrerin As siebter Klasse ausgegrenzt und lächerlich gemacht. Lehrer A möchte dies thematisieren. Lehrerin A sagt: „In Klasse X wird eine Schülerin von ihren Kameraden ausgegrenzt. Wir wollen diese Situation einmal mit wechselnden Rollen durchspielen.“	4,65	1,160	.343	.730
6: Lehrerinnen und Lehrer finden Lösungsansätze für Schwierigkeiten und Konflikte in Schule und Unterricht. ( $\alpha = .603$ )	In Lehrer Zs Klasse haben sich zwei Schülergruppen gebildet, die sich feindlich gegenüber stehen. Lehrer Z thematisiert die Situation. Lehrer Z sagt: „In der Klasse haben sich zwei Lager gebildet, die sich bekämpfen. Lasst uns darüber sprechen, wie wir das Problem lösen können.“	4,68	1,140	.276	.736

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
7: Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern.  (α= .632)	<i>Schülerin G stört häufig den Unterricht von Lehrer T und gehört zu den eher schwachen Schülern.</i> <i>Lehrer T überlegt, wie er weiter vorgehen soll.</i> <i>Lehrer T denkt: „Ich werde mit den Eltern zusammenarbeiten.“</i>	4,10	1,075	.227	.620
7: Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern.  (α= .632)	<b>Familie M kommt bezüglich ihrer Tochter T zum Beratungsgespräch zu Lehrerin F.</b> <b>Lehrer F sagt: „Ts Leistungen liegen unter dem Klassenschnitt. Sie muss also mehr lernen.“</b>	4,56	1,065	.458	.712
7: Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern.  (α= .632)	<i>Lehrerin As Klasse schneidet in einem standardisierten Rechentest eher schlecht ab. Lehrerin A überlegt, was sie tun sollte.</i> <i>Lehrerin A denkt: „Der Test stimmt mit meinem Eindruck von der Klasse nicht überein, deshalb sehe ich keinen Handlungsbedarf.“</i>	4,68	1,105	.288	.736



Mittelwerte, Standardabweichungen, Trennschärpen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
7: Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern.  ( $\alpha = .632$ )	<i>Lehrerin E möchte ihren Schüler S beurteilen. Lehrerin E denkt: „Die Klassenarbeiten von S waren gut. Ich werde ihn also gut beurteilen.“</i>	3,85	1,111	.211	.570
7: Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern.  ( $\alpha = .632$ )	<i>Schüler B gehört zu den schwächeren Mathematik-Schülern in Lehrer Hs Klasse. In einer Stillarbeitsphase geht Lehrer H durch die Reihen und bemerkt Bs Arbeit. Lehrer H sagt zu B: „Das machst du richtig. Das ist schon besser als in der letzten Klassenarbeit.“</i>	4,22	1,102	.239	.644
7: Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern.  ( $\alpha = .632$ )	<i>Familie M kommt bezüglich ihrer Tochter T zum Beratungsgespräch zu Lehrer F. Lehrer F sagt: „Ich stelle Ihnen einige Lernstrategien vor, mit denen T Erfolg haben kann.“</i>	4,58	1,075	.377	.716

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
7: Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern.  ( $\alpha = .632$ )	<i>Schüler B gehört zu den schwächeren Mathematik-Schülern in Lehrer Hs Klasse. In einer Stillarbeitsphase geht Lehrer H durch die Reihen und bemerkt Bs Arbeit. Lehrer H sagt zu B: „Das machst du richtig. Aber du musst die Anderen noch einholen.“</i>	4,30	1,176	.206	.660
7: Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern.  ( $\alpha = .632$ )	<i>Schülerin G stört häufig den Unterricht von Lehrer T und gehört zu den eher schwachen Schülern. Lehrer T überlegt, wie er weiter vorgehen soll. Lehrer T denkt: „Ich werde strenger sein müssen, damit der Unterricht nicht aufgehalten wird.“</i>	4,05	1,067	.252	.700
7: Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern.  ( $\alpha = .632$ )	Lehrerin E möchte ihren Schüler S beurteilen. Lehrerin E denkt: „Die schriftlichen Leistungen von S waren gut, allerdings hat er mündlich kaum mitgearbeitet und kein Referat gehalten. Ich werde das in mein Urteil mit einbeziehen.“	4,57	1,132	.338	.714

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
<b>7: Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern.</b>  ( $\alpha = .632$ )	Lehrerin A schneidet in einem standardisierten Rechentest eher schlecht ab. Lehrerin A überlegt, was sie tun sollte. Lehrerin A denkt: „Ich werde problematische Aufgabenstellungen noch einmal behandeln.“	4,90	,818	.460	.780
<b>8: Lehrerinnen und Lehrer erfassen Leistungen von Schülerinnen und Schülern auf der Grundlage transparenter Beurteilungsmaßstäbe.</b>  ( $\alpha = .386$ )	<i>Schüler B zeigt im Vergleich zur Klasse eher schwache Leistungen. In einer Klassenarbeit schneidet B nicht gut ab. Lehrer I hat das erwartet. Lehrer I sagt zu B: „Ich weiß, dass Du es besser kannst.“</i>	3,94	1,131	.248	.588
<b>8: Lehrerinnen und Lehrer erfassen Leistungen von Schülerinnen und Schülern auf der Grundlage transparenter Beurteilungsmaßstäbe.</b>  ( $\alpha = .386$ )	<i>Lehrer K möchte die SchülerInnen seiner Klasse beurteilen. Lehrer K denkt: „Ich bespreche auch mit den Kollegen, die diese Klasse kennen, wie sie die einzelnen SchülerInnen einschätzen.“</i>	4,23	1,335	.187	.646

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
8: Lehrerinnen und Lehrer erfassen Leistungen von Schülerinnen und Schülern auf der Grundlage transparenter Beurteilungsmaßstäbe.  (α= .386)	<i>Lehrer X hat Schüler Bs Deutsch-Klassenarbeit korrigiert. Lehrer X schreibt einen kurzen Kommentar unter die mit 3,5 benotete Arbeit. Lehrer X schreibt: „Du hast deine Rechtschreibung verbessert.“</i>	3,51	1,204	.057	.502
8: Lehrerinnen und Lehrer erfassen Leistungen von Schülerinnen und Schülern auf der Grundlage transparenter Beurteilungsmaßstäbe.  (α= .386)	<i>Lehrerin U kündigt eine Deutsch-Klassenarbeit mit dem Thema „Erörterung“ an. Lehrerin U sagt: „Wer mehr als 30 Punkte erreicht, bekommt eine 1.“</i>	4,28	1,117	.079	.656
8: Lehrerinnen und Lehrer erfassen Leistungen von Schülerinnen und Schülern auf der Grundlage transparenter Beurteilungsmaßstäbe.  (α= .386)	<i>Lehrerin E bespricht zu Schuljahrsbeginn in ihrer Klasse den Englisch-Unterricht der nächsten Monate. Lehrerin E sagt: „Ich bin zufrieden, wenn ihr englische Texte gut ins Deutsche übersetzen könnt.“</i>	3,15	,961	- .103	.430
8: Lehrerinnen und Lehrer erfassen Leistungen von Schülerinnen und Schülern auf der Grundlage transparenter Beurteilungsmaßstäbe.  (α= .386)	<i>Lehrerin U kündigt eine Deutsch-Klassenarbeit mit dem Thema „Erörterung“ an. Lehrerin U sagt: „Ich werde vor allem darauf achten, ob ihr eure Gedanken klar und differenziert darstellt.“</i>	4,45	0,996	.286	.690

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
8: Lehrerinnen und Lehrer erfassen Leistungen von Schülerinnen und Schülern auf der Grundlage transparenter Beurteilungsmaßstäbe.  (α= .386)	Lehrer X hat Schüler Bs Deutsch-Klassenarbeit korrigiert. Lehrer X schreibt einen kurzen Kommentar unter die mit 3,5 benotete Arbeit. Lehrer X schreibt: „Du befindest dich im Leistungsmittelfeld der Klasse.“	4,49	1,070	.324	.698
8: Lehrerinnen und Lehrer erfassen Leistungen von Schülerinnen und Schülern auf der Grundlage transparenter Beurteilungsmaßstäbe.  (α= .386)	Schüler B zeigt im Vergleich zur Klasse eher schwache Leistungen. In einer Klassenarbeit schneidet B nicht gut ab. Lehrer I hat das erwartet. Lehrer I sagt zu B: „Schade, das habe ich leider befürchtet.“	5,21	1,015	.322	.842
8: Lehrerinnen und Lehrer erfassen Leistungen von Schülerinnen und Schülern auf der Grundlage transparenter Beurteilungsmaßstäbe.  (α= .386)	<i>Lehrer K möchte die SchülerInnen seiner Klasse beurteilen. Lehrer K denkt: „Ich kenne diese Klasse schon lange, deshalb kann ich mich ganz auf mein Urteil verlassen.“</i>	4,20	0,988	.163	.640
8: Lehrerinnen und Lehrer erfassen Leistungen von Schülerinnen und Schülern auf der Grundlage transparenter Beurteilungsmaßstäbe.  (α= .386)	<i>Lehrerin E bespricht zu Schuljahrsbeginn in ihrer Klasse den Englisch-Unterricht der nächsten Monate. Lehrerin E sagt: „Ich bin zufrieden, wenn wir gut vorankommen.“</i>	3,50	1,014	.079	.500

Mittelwerte, Standardabweichungen, Trennschärpen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
<b>9: Lehrerinnen und Lehrer sind sich der besonderen Anforderungen des Lehrerberufs bewusst. Sie verstehen ihren Beruf als ein öffentliches Amt mit besonderer Verantwortung und Verpflichtung.</b>	Die LehrerInnen einer Schule diskutieren über die rechtlichen Rahmenbedingungen des Lehrberufs. Lehrerin W sagt: „Es ist wichtig, die rechtlichen Rahmenbedingungen zu kennen, um sicher handeln zu können.“	4,83	0,961	.551	.766
( $\alpha = .749$ )					
<b>9: Lehrerinnen und Lehrer sind sich der besonderen Anforderungen des Lehrerberufs bewusst. Sie verstehen ihren Beruf als ein öffentliches Amt mit besonderer Verantwortung und Verpflichtung.</b>	<i>Lehrer Es Klasse hat seine hochgesteckten Ziele in diesem Schuljahr nicht erreicht.</i> <i>Lehrer E denkt: „Habe ich zu hohe Erwartungen?“</i>	3,87	1,097	.187	.574
( $\alpha = .749$ )					
<b>9: Lehrerinnen und Lehrer sind sich der besonderen Anforderungen des Lehrerberufs bewusst. Sie verstehen ihren Beruf als ein öffentliches Amt mit besonderer Verantwortung und Verpflichtung.</b>	<i>Lehrerin D erwartet von ihren SchülerInnen Pünktlichkeit. Sie selbst kommt häufig zu spät, worauf sie die SchülerInnen ansprechen.</i> <i>Lehrerin D sagt: „Ich werde in Zukunft darauf achten.“</i>	3,54	1,256	.074	.508
( $\alpha = .749$ )					

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
9: Lehrerinnen und Lehrer sind sich der besonderen Anforderungen des Lehrerberufs bewusst. Sie verstehen ihren Beruf als ein öffentliches Amt mit besonderer Verantwortung und Verpflichtung.  ( $\alpha = .749$ )	<i>An der Schule von Lehrer Y sollen die LehrerInnen stark an der Schulentwicklung beteiligt werden. Lehrer Y denkt: „Ich möchte nicht so häufig teilnehmen, da ich es wichtiger finde, mich voll und ganz mit meinem Unterricht und den SchülerInnen zu befassen.“</i>	4,22	1,076	.266	.644
9: Lehrerinnen und Lehrer sind sich der besonderen Anforderungen des Lehrerberufs bewusst. Sie verstehen ihren Beruf als ein öffentliches Amt mit besonderer Verantwortung und Verpflichtung.  ( $\alpha = .749$ )	Lehrer A und Lehrer J diskutieren über ihre Wertvorstellungen bezüglich ihres Berufs. Beide vertreten unterschiedliche Meinungen. Lehrer J denkt: „Ich weiß, dass meine Werte die richtigen sind.“	4,96	1,070	.500	.792
9: Lehrerinnen und Lehrer sind sich der besonderen Anforderungen des Lehrerberufs bewusst. Sie verstehen ihren Beruf als ein öffentliches Amt mit besonderer Verantwortung und Verpflichtung.  ( $\alpha = .749$ )	Die LehrerInnen einer Schule diskutieren über die rechtlichen Rahmenbedingungen des Lehrberufs. Lehrerin W sagt: „Ich kenne die Rahmenbedingungen nicht genau, da es selbstverständlich ist, dass ich mich korrekt verhalte.“	4,78	1,141	.545	.756

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
9: Lehrerinnen und Lehrer sind sich der besonderen Anforderungen des Lehrerberufs bewusst. Sie verstehen ihren Beruf als ein öffentliches Amt mit besonderer Verantwortung und Verpflichtung.	An der Schule von Lehrer Y sollen die LehrerInnen stark an der Schulentwicklung beteiligt werden. Lehrer Y denkt: „Ich nehme gern teil, da ich es wichtig finde, die Schule mit gestalten zu können.“	5,05	1,018	.664	.810
( $\alpha = .749$ )					
9: Lehrerinnen und Lehrer sind sich der besonderen Anforderungen des Lehrerberufs bewusst. Sie verstehen ihren Beruf als ein öffentliches Amt mit besonderer Verantwortung und Verpflichtung.	Lehrer Ws Klasse hat seine hochgesteckten Ziele in diesem Schuljahr nicht erreicht. Lehrer W denkt: „Die Klasse muss sich mehr anstrengen.“	4,49	0,954	.421	.698
( $\alpha = .749$ )					
9: Lehrerinnen und Lehrer sind sich der besonderen Anforderungen des Lehrerberufs bewusst. Sie verstehen ihren Beruf als ein öffentliches Amt mit besonderer Verantwortung und Verpflichtung.	Lehrerin D erwartet von ihren SchülerInnen Pünktlichkeit. Sie selbst kommt häufig zu spät, worauf sie die SchülerInnen ansprechen. Lehrerin D sagt: „Das lässt sich leider nicht vermeiden, da ich sehr viel zu tun habe.“	5,22	1,079	.541	.844
( $\alpha = .749$ )					



**Mittelwerte, Standardabweichungen, Trennschärpen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests**

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
<b>9: Lehrerinnen und Lehrer sind sich der besonderen Anforderungen des Lehrerberufs bewusst. Sie verstehen ihren Beruf als ein öffentliches Amt mit besonderer Verantwortung und Verpflichtung.</b>	Lehrer A und Lehrer J diskutieren über ihre Wertvorstellungen bezüglich ihres Berufs. Beide vertreten unterschiedliche Meinungen. Lehrer J denkt: „Ich denke darüber nach, ob meine Werte die richtigen sind.“	4,46	0,933	.511	.692
( $\alpha = .749$ )					
<b>10: Lehrerinnen und Lehrer verstehen ihren Beruf als ständige Lernaufgabe.</b>	<b>Lehrerin S hat von einer neuartigen Unterrichtsmethode gehört, die interessant zu sein scheint.</b>	4,88	1,096	.506	.776
( $\alpha = .805$ )	<b>Lehrerin S denkt: „Ich habe genügend Methoden zur Auswahl.“</b>				
<b>10: Lehrerinnen und Lehrer verstehen ihren Beruf als ständige Lernaufgabe.</b>	Lehrerin D kommt mit Klasse 7a sehr gut zurecht, mit Klasse 9b hat sie leider immer wieder Probleme. Lehrerin D denkt: „In Klasse 9b werde ich anders vorgehen müssen.“	4,61	,855	.480	.722
( $\alpha = .805$ )					
<b>10: Lehrerinnen und Lehrer verstehen ihren Beruf als ständige Lernaufgabe.</b>	Lehrer P plant die Themen für seine Physik-AG. Lehrer P denkt: „Ich informiere mich, ob es neue wissenschaftliche Erkenntnisse gibt.“	5,00	1,101	.564	.800
( $\alpha = .805$ )					

**Mittelwerte, Standardabweichungen, Trennschärpen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests**

<b>Skala (Cronbachs Alpha)</b>	<b>Item*</b>	<b>M</b>	<b>SD</b>	<b>r<sub>it</sub></b>	<b>p**</b>
<b>10: Lehrerinnen und Lehrer verstehen ihren Beruf als ständige Lernaufgabe. (α= .805)</b>	Die Leistungen von Lehrer Bs Klasse sinken, und die Schüler verhalten sich ihm gegenüber respektlos. Lehrer B überlegt, was er tun könnte. Lehrer B denkt: „Ich frage bei meinen KollegInnen nach, was sie tun würden.“	3,93	1,001	.319	.586
<b>10: Lehrerinnen und Lehrer verstehen ihren Beruf als ständige Lernaufgabe. (α= .805)</b>	<b>Lehrerin D kommt mit Klasse 7a sehr gut zurecht, mit Klasse 9b hat sie leider immer wieder Probleme. Lehrerin D denkt: „Schade, dass Klasse 9b nicht wie Klasse 7a ist.“</b>	5,20	0,978	.591	.840
<b>10: Lehrerinnen und Lehrer verstehen ihren Beruf als ständige Lernaufgabe. (α= .805)</b>	<b>Die Leistungen von Lehrer Bs Klasse sinken, und die Schüler verhalten sich ihm gegenüber respektlos. Lehrer B überlegt, was er tun könnte. Lehrer B denkt: „Ich möchte nicht, dass meine KollegInnen von meinen Problemen erfahren.“</b>	5,08	1,058	.581	.816
<b>10: Lehrerinnen und Lehrer verstehen ihren Beruf als ständige Lernaufgabe. (α= .805)</b>	<b>Lehrer P plant die Themen für seine Physik-AG. Lehrer P denkt: „Ich kann die Materialien vom letzten Jahr wieder verwenden.“</b>	4,26	1,002	.564	.652

**Mittelwerte, Standardabweichungen, Trennschärpen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests**

<b>Skala (Cronbachs Alpha)</b>	<b>Item*</b>	<b>M</b>	<b>SD</b>	<b>r<sub>it</sub></b>	<b>p**</b>
<b>10: Lehrerinnen und Lehrer verstehen ihren Beruf als ständige Lernaufgabe.</b>  ( $\alpha = .805$ )	Lehrer C hat einen wissenschaftlichen Artikel über effektive allgemeine Unterrichtsmethoden gelesen.  Lehrer C denkt: „Dies werde ich auf jeden Fall umsetzen.“	4,37	1,005	.480	.674
<b>10: Lehrerinnen und Lehrer verstehen ihren Beruf als ständige Lernaufgabe.</b>  ( $\alpha = .805$ )	Lehrerin S hat von einer neuartigen Unterrichtsmethode gehört, die interessant zu sein scheint.  Lehrerin S denkt: „Ich werde ein Seminar zu dem Thema besuchen.“	5,08	0,934	.607	.816
<b>10: Lehrerinnen und Lehrer verstehen ihren Beruf als ständige Lernaufgabe.</b>  ( $\alpha = .805$ )	<b>Lehrer C hat einen wissenschaftlichen Artikel über effektive allgemeine Unterrichtsmethoden gelesen.</b>  <b>Lehrer C denkt: „Dies lässt sich in meine Praxis nicht übertragen.“</b>	4,80	1,043	.506	.760
<b>11: Lehrerinnen und Lehrer beteiligen sich an der Planung und Umsetzung schulischer Projekte und Vorhaben.</b>  ( $\alpha = .718$ )	<b>Einige Schüler von Lehrer Ks zehnter Klasse möchten eine Klassenzeitung gründen und berichten Lehrer K davon.</b>  <b>Lehrer K sagt: „Also dann: Ich wünsche euch viel Erfolg.“</b>	4,75	1,184	.480	.750

Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
11: Lehrerinnen und Lehrer beteiligen sich an der Planung und Umsetzung schulischer Projekte und Vorhaben. ( $\alpha = .718$ )	Lehrerin M plant ein großes Schulprojekt und fragt Lehrerin L, ob sie mitarbeiten möchte. Lehrerin L sagt: „Ich nehme teil. Ich finde es wichtig, auch außerhalb des Unterrichts mit den SchülerInnen zu arbeiten.“	5,14	1,100	.619	.828
11: Lehrerinnen und Lehrer beteiligen sich an der Planung und Umsetzung schulischer Projekte und Vorhaben. ( $\alpha = .718$ )	Lehrerin N hat ihre SchülerInnen einen Fragebogen zur Unterrichtsevaluation bearbeiten lassen und betrachtet nun die Ergebnisse. Lehrerin N denkt: „Ich behalte meinen Unterricht bei, da das nur eine Momentaufnahme war.“	4,97	1,005	.503	.794
11: Lehrerinnen und Lehrer beteiligen sich an der Planung und Umsetzung schulischer Projekte und Vorhaben. ( $\alpha = .718$ )	Lehrerin M plant eine Projektwoche mit ihrer Klasse. Sie überlegt, wie sie das Projekt gestalten soll. Lehrerin M denkt: „Ich möchte ganz kreativ vorgehen und mich nicht durch zu viel Theorie beeinflussen lassen.“	2,48	1,071	- .407	.296
11: Lehrerinnen und Lehrer beteiligen sich an der Planung und Umsetzung schulischer Projekte und Vorhaben. ( $\alpha = .718$ )	Einige Schüler von Lehrer Ks zehnter Klasse möchten eine Klassenzeitung gründen und berichten Lehrer K davon. Lehrer K sagt: „Ich berate euch bei eurer Arbeit.“	5,04	1,031	.588	.808

**Mittelwerte, Standardabweichungen, Trennschärpen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests**

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
11: Lehrerinnen und Lehrer beteiligen sich an der Planung und Umsetzung schulischer Projekte und Vorhaben.  ( $\alpha = .718$ )	Eine Kollegin von Lehrer H lässt ihre SchülerInnen ihren Unterricht evaluieren. Lehrer H überlegt, ob er dies auch tun soll.  Lehrer H sagt: „Die Leistungen der SchülerInnen reichen aus, im daraus schließen zu können, ob sie mit dem Unterricht zufrieden sind.“	4,59	1,013	.503	.718
11: Lehrerinnen und Lehrer beteiligen sich an der Planung und Umsetzung schulischer Projekte und Vorhaben.  ( $\alpha = .718$ )	Lehrerin N hat ihren SchülerInnen einen Fragebogen zur Unterrichtsevaluation bearbeiten lassen und betrachtet nun die Ergebnisse.  Lehrerin N denkt: „Ich werde die Bereiche verändern, mit denen die SchülerInnen unzufrieden sind.“	4,80	1,094	.541	.760
11: Lehrerinnen und Lehrer beteiligen sich an der Planung und Umsetzung schulischer Projekte und Vorhaben.  ( $\alpha = .718$ )	Eine Kollegin von Lehrer H lässt ihre SchülerInnen ihren Unterricht evaluieren. Lehrer H überlegt, ob er dies auch tun soll.  Lehrer H sagt: „Ja, denn mir ist es wichtig zu wissen, was ich aus Schülersicht verbessern kann.“	5,22	1,027	.634	.844

**Mittelwerte, Standardabweichungen, Trennschärfen sowie Schwierigkeitsindizes der Items des Situational Judgement Tests**

Skala (Cronbachs Alpha)	Item*	M	SD	r <sub>it</sub>	p**
<b>11: Lehrerinnen und Lehrer beteiligen sich an der Planung und Umsetzung schulischer Projekte und Vorhaben.</b>  ( $\alpha = .718$ )	Lehrerin M plant eine Projektwoche mit ihrer Klasse. Sie überlegt, wie sie das Projekt gestalten soll. Lehrerin M denkt: „Ich lese nach, welche Forschungsergebnisse es zur Gestaltung von Projekten gibt.“	4,53	1,072	.619	.706
<b>11: Lehrerinnen und Lehrer beteiligen sich an der Planung und Umsetzung schulischer Projekte und Vorhaben.</b>  ( $\alpha = .718$ )	<b>Lehrerin M plant ein großes Schulprojekt und fragt Lehrerin L, ob sie mitarbeiten möchte.</b> Lehrerin L sagt: „Ich nehme nicht teil. Ich möchte die Zeit lieber in die Gestaltung meines Unterrichts investieren.“	4,38	1,015	.320	.676

\*Kursivdruck: Items mit geringer Trennschärfe < .3, Fettdruck: negativ gepolte, umkodierte Items

\*\* Kursivdruck: Items mit Lösungswahrscheinlichkeit > .8

## Anhang K: Zuordnung der Items zu Kompetenzen durch 8 ExpertInnen

Kompetenz	Item	Übereinstimmung mit Kompetenz*	weitere genannte Kompetenz(en) (Anzahl Nennungen)*
1	Lehrerin B beginnt den Englisch-Unterricht in ihrer Klasse. Lehrerin B sagt: „Wir behandeln heute das ‚future perfect‘. Wir werden erst den Stoff von gestern wiederholen, dann gemeinsam arbeiten und zuletzt die Lehrer Es Klasse hat seine hochgesteckten Ziele in diesem Schuljahr nicht erreicht.	1	2 (7)
9	Lehrer E denkt: „Habe ich zu hohe Erwartungen?“	0	10 (8)
11	Einige Schüler von Lehrer Ks zehnter Klasse möchten eine Klassenzeitung gründen und berichten Lehrer K davon. Lehrer K sagt: „Also dann: Ich wünsche euch viel Erfolg.“	3	5 (5)
7	Schülerin G stört häufig den Unterricht von Lehrer T und gehört zu den eher schwachen Schülern. Lehrer T überlegt, wie er weiter vorgehen soll. Lehrer T denkt: „Ich werde mit den Eltern zusammenarbeiten.“	3	6 (5)
2	Lehrer Z stellt seiner Klasse eine Frage und erhält von einem Schüler eine falsche Antwort. Lehrer Z sagt: „Nein, das stimmt nicht. Wer kann es richtig sagen?“	2	1 (4), 7 (2)

Kompetenz	Item	Übereinstimmung mit Kompetenz*	weitere genannte Kompetenz(en) (Anzahl Nennungen)*
7	<p>Familie M kommt bezüglich ihrer Tochter T zum Beratungsgespräch zu Lehrerin F.</p> <p>Lehrer F sagt: „Ts Leistungen liegen unter dem Klassenschnitt. Sie muss also mehr lernen.“</p>	8	0
8	<p>Schüler B zeigt im Vergleich zur Klasse eher schwache Leistungen. In einer Klassenarbeit schneidet B nicht gut ab. Lehrer I hat das erwartet.</p> <p>Lehrer I sagt zu B: „Ich weiß, dass Du es besser kannst.“</p>	0	2 (8)
8	<p>Lehrer K möchte die SchülerInnen seiner Klasse beurteilen.</p> <p>Lehrer K denkt: „Ich bespreche auch mit den Kollegen, die diese Klasse kennen, wie sie die einzelnen SchülerInnen einschätzen.“</p>	7	7 (1)
11	<p>Lehrerin M plant ein großes Schulprojekt und fragt Lehrerin L, ob sie mitarbeiten möchte.</p> <p>Lehrerin L sagt: „Ich nehme teil. Ich finde es wichtig, auch außerhalb des Unterrichts mit den SchülerInnen zu</p>	7	12 (1)
7	<p>Lehrerin As Klasse schneidet in einem standardisierten Rechentest eher schlecht ab. Lehrerin A überlegt, was sie tun sollte.</p> <p>Lehrerin A denkt: „Der Test stimmt mit meinem Eindruck von der Klasse nicht</p>	1	8 (7)
3	<p>Lehrer Z möchte in der nächsten Stunde in ein neues Thema einführen. Er überlegt, wie er vorgehen soll.</p> <p>Lehrer Z denkt: „Ich beginne mit einem Vortrag, der das Thema darstellt.“</p>	0	1 (5), 2 (3)



Kompetenz	Item	Übereinstimmung mit Kompetenz*	weitere genannte Kompetenz(en) (Anzahl Nennungen)*
3	Lehrerin G übt mit den Schülern ihrer siebten Klasse das Verstehen von Sachtexten. Sie gibt der Klasse Arbeitsanweisungen.	2	2 (6)
3	Lehrerin G sagt: „Lest den Text durch und streicht das Wichtigste an. Dann Ein Schüler hat bei einer Klassenarbeit schlecht abgeschnitten. Lehrer H spricht mit ihm über das Ergebnis. Lehrer H sagt: „Du hast nur wenig gelernt? Dann konntest du natürlich die Aufgaben nicht lösen.“	0	2 (1), 5 (1), 7 (6)
4	In der Klasse 9d werden immer wieder Bemerkungen laut wie: „Alle Muslime sind Terroristen – vor denen muss man Angst haben.“ Lehrer F möchte die Vorurteile abbauen.	3	5 (4), 6 (1)
5	Lehrer F sagt: „Ich erlaube solche Ein Wandertag ist in Klasse 9d geplant. Lehrer R bespricht mit den SchülerInnen, wo es hingehen soll. Lehrer R sagt: „Wir sammeln einmal Vorschläge, wo wir hinfahren könnten.“	6	6 (2)
5	Schüler P scheint im Unterricht niedergeschlagen und abwesend, ohne dass ein offensichtlicher Grund vorliegt. Lehrer M spricht dies in der Pause an. Lehrer M sagt: „Was lenkt dich denn ab? Versuche bitte, dich trotzdem zu	0	2 (1), 4 (6), 6 (1)
5	Lehrerin S möchte hilfsbereites Verhalten ihrer SchülerInnen fördern. Lehrerin S sagt: „Wer schnell mit seiner Aufgabe fertig ist, fragt bitte den Nachbarn, ob er ihn unterstützen kann.“	2	2 (4), 5 (2)

Kompetenz	Item	Übereinstimmung mit Kompetenz*	weitere genannte Kompetenz(en) (Anzahl Nennungen)*
6	An einer benachbarten Schule werden seit einiger Zeit immer wieder SchülerInnen von anderen gewaltsam angegriffen. Lehrer Bs Klasse ist nicht direkt betroffen.	1	4 (2), 5 (5)
6	Lehrer B denkt: „Ich lasse die Schüler Lehrerin S plant den Deutsch-Unterricht in ihrer achten Klasse und überlegt, welche Methoden sie vor allem anwenden soll.	0	1 (7), 2 (1)
11	Lehrerin S denkt: „Ich lasse am Besten die Schüler in Stillarbeit Aufgaben Lehrerin N hat ihre SchülerInnen einen Fragebogen zur Unterrichtsevaluation bearbeiten lassen und betrachtet nun die Ergebnisse. Lehrerin N denkt: „Ich behalte meinen Unterricht bei, da das nur eine	1	10 (7)
5	Lehrer U möchte seinen Schülern Werte und Normen vermitteln und überlegt, wie er dies umsetzen könnte. Lehrer U denkt: „Ich stelle Regeln auf.“	7	6 (1)
7	Lehrerin E möchte ihren Schüler S beurteilen. Lehrerin E denkt: „Die Klassenarbeiten von S waren gut. Ich werde ihn also gut beurteilen.“	0	7 (8)
8	Lehrer X hat Schüler Bs Deutsch-Klassenarbeit korrigiert. Lehrer X schreibt einen kurzen Kommentar unter die mit 3,5 benotete Arbeit. Lehrer X schreibt: „Du hast deine Rechtschreibung verbessert.“	5	7 (3)

Kompetenz	Item	Übereinstimmung mit Kompetenz*	weitere genannte Kompetenz(en) (Anzahl Nennungen)*
9	Lehrerin D erwartet von seinen SchülerInnen Pünktlichkeit. Sie selbst kommt häufig zu spät, worauf sie die SchülerInnen ansprechen.	4	5 (4)
9	Lehrerin D sagt: „Ich werde in Zukunft darauf achten.“ An der Schule von Lehrer Y sollen die LehrerInnen stark an der Schulentwicklung beteiligt werden. Lehrer Y denkt: „Ich möchte nicht so häufig teilnehmen, da ich es wichtiger finde, mich voll und ganz mit meinem Lehrer B beginnt den Englisch-Unterricht in seiner Klasse.	1	11 (7)
1	Lehrer B sagt: „Wir werden heute gut voran kommen. Lisa, was war die Hausaufgabe? Martin hat noch eine Frage zum gestrigen Stoff? Dann stelle Lehrer G plant den Unterricht in Klasse 11c. Diese Klasse kennt er noch nicht.	2	2 (4), 3 (2)
1	Lehrer G denkt: „Zur Unterrichtsplanung orientiere ich mich an den Erfahrungen in meiner eigenen Schulzeit, da ich über die Methoden, die ich selbst erlebt habe, Lehrerin A möchte, dass ihre Schüler das Prinzip des Dividierens gut verstehen.	6	10 (2)
2	Lehrerin A sagt: „Rechnet bitte: 10 durch 2.“	4	1 (2), 3 (2)
2	Lehrer K möchte im Physik-Unterricht die Elektrik-Kenntnisse seiner SchülerInnen vertiefen. Lehrer K sagt: „Wir wollen zusammen ein Radio bauen. Was brauchen wir dafür?“	7	3 (1)

Kompetenz	Item	Übereinstimmung mit Kompetenz*	weitere genannte Kompetenz(en) (Anzahl Nennungen)*
3	Schüler P hat bei Klassenarbeiten von Lehrer J wiederholt nicht gut abgeschnitten. Lehrer J sagt: „Ich weiß, dass du es besser kannst.“	0	2 (2), 4 (1), 7 (3), 8 (2)
4	Die Schüler B und C hinken im Mathematikunterricht der 9d etwas hinterher und haben insbesondere Schwierigkeiten mit Textaufgaben. Lehrerin K gibt einen Arbeitsauftrag. Lehrerin K sagt: „Bitte bearbeitet die Lehrer O möchte Umgangsregeln in seiner Klasse einrichten.	0	1 (2), 2 (3), 3 (1), 7 (1)
5	Lehrer O sagt: „Wir erarbeiten gemeinsam einen Regelkatalog und vereinbaren, was geschehen soll, wenn die Regeln nicht eingehalten werden.“ In der Klasse 9d werden immer wieder Bemerkungen laut wie: „Alle Muslime sind Terroristen – vor denen muss man Angst haben.“ Lehrer F möchte die Vorurteile abbauen.	6	6 (2)
4	Lehrer F sagt: „Sammelt bis morgen In Lehrer Ps Klasse haben sich zwei Schülergruppen gebildet, die sich feindlich gegenüber stehen. Lehrer P thematisiert die Situation.	3	5 (3), 6 (2)
6	Lehrer P sagt: „In der Klasse haben sich zwei Lager gebildet, die sich Lehrerin U kündigt eine Deutsch-Klassenarbeit mit dem Thema „Erörterung“ an.	8	0
8	Lehrerin U sagt: „Wer mehr als 30 Punkte erreicht, bekommt eine 1.“	8	0

Kompetenz	Item	Übereinstimmung mit Kompetenz*	weitere genannte Kompetenz(en) (Anzahl Nennungen)*
5	Ein Wandertag ist in Klasse 9d geplant. Lehrer I bespricht mit den SchülerInnen, wo es hingehen soll. Lehrer R sagt: „Wir fahren zur Burg, das gefällt euch sicher am besten.“	3	11 (5)
1	Lehrerin A plant den Unterricht für die Klassen 5c und 8a. Sie überlegt dabei, welche Methoden sie anwenden soll. Lehrerin A denkt: „Welche Methoden und Medien eignen sich für welche Klasse?“	8	0
1	Lehrerin E lässt die Schüler in der 10. Klasse ihrer Realschule Aufgaben in Kleingruppen bearbeiten. Als sie durch die Reihen geht, bemerkt sie, dass die Aufgaben kaum gelöst werden.	1	2 (3), 4 (4)
2	Lehrerin E sagt: „Arbeitet bitte Lehrerin C hat die achte Klasse ihres Gymnasiums in das Prinzip der linearen Funktionen eingeführt. Dann sagt Lehrerin C: „Ich rechne eine Aufgabe vor, dann rechnen wir gemeinsam mit einem Schüler an der Schülerin J spricht in den Pausen häufig begeistert der Kfz-Werkstatt ihres Onkels. Lehrer B möchte dieses Interesse fördern.“	3	1 (5)
4	Lehrer B sagt: „Welches ist denn dein Lieblingsauto?“	6	3 (1), 7 (1)
6	Schüler T wird in Lehrer As sechster Klasse ausgegrenzt und lächerlich gemacht. Lehrer A möchte thematisieren. Lehrer A sagt: „Heute sprechen wir über euer unfaires Verhalten T gegenüber.“	8	0

Kompetenz	Item	Übereinstimmung mit Kompetenz*	weitere genannte Kompetenz(en) (Anzahl Nennungen)*
8	Lehrerin U kündigt eine Deutsch-Klassenarbeit mit dem Thema „Erörterung“ an. Lehrerin U sagt: „Ich werde vor allem darauf achten, ob ihr eure Gedanken klar und differenziert darstellt.“	8	0
8	Lehrerin E bespricht zu Schuljahrsbeginn in ihrer Klasse den Englisch-Unterricht der nächsten Monate. Lehrerin E sagt: „Ich bin zufrieden, wenn ihr englische Texte gut ins Deutsche	7	2 (1)
9	Lehrer A und Lehrer J diskutieren über ihre Wertvorstellungen bezüglich ihres Berufs. Beide vertreten unterschiedliche Meinungen. Lehrer J denkt: „Ich weiß, dass meine Werte die richtigen sind.“	2	5 (1), 10 (5)
10	Lehrerin S hat von einer neuartigen Unterrichtsmethode gehört, die interessant zu sein scheint. Lehrerin S denkt: „Ich habe genügend Methoden zur Auswahl.“	7	9 (1)
10	Lehrerin D kommt mit Klasse 7a sehr gut zurecht, mit Klasse 9b hat sie leider immer wieder Probleme. Lehrerin D denkt: „In Klasse 9b werde ich anders vorgehen müssen.“	1	1 (3), 2 (4)
10	Lehrer P plant die Themen für seine Physik-AG. Lehrer P denkt: „Ich informiere mich, ob es neue wissenschaftliche Erkenntnisse gibt.“	7	1 (1)

Kompetenz	Item	Übereinstimmung mit Kompetenz*	weitere genannte Kompetenz(en) (Anzahl Nennungen)*
11	Lehrerin M plant eine Projektwoche mit ihrer Klasse. Sie überlegt, wie sie das Projekt gestalten soll.	0	1 (2), <b>10 (6)</b>
1	Lehrerin M denkt: „Ich möchte ganz kreativ vorgehen und mich nicht durch zu viel Theorie beeinflussen lassen.“ Lehrerin E lässt die Schüler in der 10. Klasse ihrer Realschule Aufgaben in Kleingruppen bearbeiten. Als sie durch die Reihen geht, bemerkt sie, dass die Aufgaben kaum gelöst werden.	0	<b>2 (6)</b> , 3 (2)
5	Lehrerin E sagt: „Notiert euch Lehrer U möchte seinen Schülern Werte und Normen vermitteln und überlegt, wie er dies umsetzen könnte.	8	0
6	Lehrer U denkt: „Ich gebe den Schülern ein moralisches Dilemma vor, über das wir dann diskutieren.“ Lehrerin L plant den Deutsch-Unterricht in ihrer achten Klasse und überlegt, welche Methoden sie vorwiegend anwenden soll.	0	1 (3), <b>3 (5)</b>
7	Lehrerin L denkt: „Ich lasse am Besten die Schüler in Gruppen Aufgaben Schüler B gehört zu den schwächeren Mathematik-Schülern in Lehrer Hs Klasse. In einer Stillarbeitsphase geht Lehrer H durch die Reihen und bemerkt Bs Arbeit.	7	2 (1)
7	Lehrer H sagt zu B: „Das machst du Familie M kommt bezüglich ihrer Tochter T zum Beratungsgespräch zu Lehrer F. Lehrer F sagt: „Ich stelle Ihnen einige Lernstrategien vor, mit denen T Erfolg haben kann.“	6	6 (2)

Kompetenz	Item	Übereinstimmung mit Kompetenz*	weitere genannte Kompetenz(en) (Anzahl Nennungen)*
11	Einige Schüler von Lehrer Ks zehnter Klasse möchten eine Klassenzeitung gründen und berichten Lehrer K davon. Lehrer K sagt: „Ich berate euch bei eurer Arbeit.“	6	3 (2)
2	Lehrerin C hat die achte Klasse ihres Gymnasiums in das Prinzip der linearen Funktionen eingeführt. Dann sagt Lehrerin C: „Jeder rechnet nun bitte bis zum Ende der Stunde die Beispiele 1 bis 5 aus dem Buch.“	1	4 (1), 2 (1)
9	Die LehrerInnen einer Schule diskutieren über die rechtlichen Rahmenbedingungen des Lehrberufs. Lehrerin W sagt: „Ich kenne die Rahmenbedingungen nicht genau, da es selbstverständlich ist, dass ich mich	6	10 (2)
10	Die Leistungen von Lehrer Bs Klasse sinken, und die Schüler verhalten sich ihm gegenüber respektlos. Lehrer B überlegt, was er tun könnte. Lehrer B denkt: „Ich frage bei meinen KollegInnen nach, was sie tun würden.“	8	0
11	Eine Kollegin von Lehrer H lässt ihre SchülerInnen ihren Unterricht evaluieren. Lehrer H überlegt, ob er dies auch tun soll.	1	9 (2), 10 (5)
2	Lehrer H sagt: „Die Leistungen der SchülerInnen reichen aus, im daraus Lehrer Z stellt seiner Klasse eine Frage und erhält von einem Schüler eine falsche Antwort. Lehrer Z sagt: „Nein, das stimmt nicht. Wer kann erklären, warum?“	8	0



Kompetenz	Item	Übereinstimmung mit Kompetenz*	weitere genannte Kompetenz(en) (Anzahl Nennungen)*
2	<p>Lehrerin B übt mit ihrer Klasse das Übersetzen aus dem Englischen ins Deutsche. Sie möchte die SchülerInnen motivieren und vergibt die Hausaufgaben für den nächsten Tag.</p> <p>Sie sagt: „Übersetzt bis morgen den Schüler P scheint im Unterricht niedergeschlagen und abwesend, ohne dass ein offensichtlicher Grund vorliegt.</p>	6	1 (2)
5	<p>Lehrer R spricht dies in der Pause an.</p> <p>Lehrer R sagt: „Möchtest Du einmal mit mir darüber sprechen, was dich Schüler B gehört zu den schwächeren Mathematik-Schülern in Lehrer Hs Klasse. In einer Stillarbeitsphase geht</p>	0	4 (6), 6 (1), 7 (1)
7	<p>Lehrer H durch die Reihen und bemerkt Bs Arbeit.</p> <p>Lehrer H sagt zu B: „Das machst du</p>	7	2 (1)
10	<p>Lehrerin D kommt mit Klasse 7a sehr gut zurecht, mit Klasse 9b hat sie leider immer wieder Probleme.</p> <p>Lehrerin D denkt: „Schade, dass Klasse 9b nicht wie Klasse 7a ist.“</p>	3	1 (1), 6 (4)
11	<p>Lehrerin N hat ihren SchülerInnen einen Fragebogen zur Unterrichtsevaluation bearbeiten lassen und betrachtet nun die Ergebnisse.</p> <p>Lehrerin N denkt: „Ich werde die Bereiche verändern, mit denen die Schülerin G stört häufig den Unterricht von Lehrer T und gehört zu den eher schwachen Schülern. Lehrer T überlegt, wie er weiter vorgehen soll.</p>	1	9 (4), 10 (3)
7	<p>Lehrer T denkt: „Ich werde strenger sein müssen, damit der Unterricht nicht</p>	0	2 (2), 4 (1), 6 (5)

Kompetenz	Item	Übereinstimmung mit Kompetenz*	weitere genannte Kompetenz(en) (Anzahl Nennungen)*
7	<p>Lehrerin E möchte ihren Schüler S beurteilen.</p> <p>Lehrerin E denkt: „Die schriftlichen Leistungen von S waren gut, allerdings hat er mündlich kaum mitgearbeitet und kein Referat gehalten. Ich werde das in An der Schule von Lehrer Y sollen die LehrerInnen stark an der Schulentwicklung beteiligt werden.“</p>	1	8 (7)
9	<p>Lehrer Y denkt: „Ich nehme gern teil, da ich es wichtig finde, die Schule mit gestalten zu können.“</p> <p>Lehrerin möchte, dass ihre Schüler das Prinzip des Dividierens gut verstehen.</p>	0	11 (8)
2	<p>Lehrerin A sagt: „Jonas hat einen Freund zum Kuchenessen eingeladen. Vier Kuchenstücke hat er noch. Wie viele Kuchenstücke bekommt jeder der An einer benachbarten Schule werden seit einiger Zeit immer wieder SchülerInnen von anderen gewaltsam angegriffen. Lehrer Bs Klasse ist nicht direkt betroffen.“</p>	8	0
6	<p>Lehrer B denkt: „Da unsere Schule nicht Lehrer X hat Schüler Bs Deutsch-Klassenarbeit korrigiert. Lehrer X schreibt einen kurzen Kommentar unter die mit 3,5 benotete Arbeit.“</p>	3	4 (3), 5 (2), 6 (3)
8	<p>Lehrer X schreibt: „Du befindest dich im Leistungsmittelfeld der Klasse.“</p> <p>Die Leistungen von Lehrer Bs Klasse sinken, und die Schüler verhalten sich ihm gegenüber respektlos. Lehrer B überlegt, was er tun könnte.</p>	2	7 (6)
10	<p>Lehrer B denkt: „Ich möchte nicht, dass meine KollegInnen von meinen</p>	1	9 (7)

Kompetenz	Item	Übereinstimmung mit Kompetenz*	weitere genannte Kompetenz(en) (Anzahl Nennungen)*
6	Schüler T wird in Lehrerin As siebter Klasse ausgegrenzt und lächerlich gemacht. Lehrer A möchte dies thematisieren.	8	0
3	Lehrerin A sagt: „In Klasse X wird eine Schülerin von ihren Kameraden Schüler P hat bei Klassenarbeiten von Lehrer J wiederholt nicht gut abgeschnitten. Lehrer J sagt: „Beim nächsten Mal sollte es besser werden.“	0	7 (7), 8 (1)
3	Lehrer P kündigt eine Klassenarbeit an. Er möchte die SchülerInnen auf das Lernen vorbereiten. Lehrer P sagt: „Es geht um eine ganze Menge Stoff, also besprechen wir nun, wie ihr am besten lernen könnt.“	8	0
11	Eine Kollegin von Lehrer H lässt ihre SchülerInnen ihren Unterricht evaluieren. Lehrer H überlegt, ob er dies auch tun soll. Lehrer H sagt: „Ja, denn mir ist es wichtig zu wissen, was ich aus	1	10 (7)
10	Lehrer P plant die Themen für seine Physik-AG. Lehrer P denkt: „Ich kann die Materialien vom letzten Jahr wieder verwenden.“	3	1 (4), 9 (1)
2	Lehrerin B übt mit ihrer Klasse das Übersetzen aus dem Englischen ins Deutsche. Sie möchte die SchülerInnen motivieren und vergibt die Hausaufgaben für den nächsten Tag. Sie sagt: „Übersetzt bis morgen einen	7	3 (1)

Kompetenz	Item	Übereinstimmung mit Kompetenz*	weitere genannte Kompetenz(en) (Anzahl Nennungen)*
1	Lehrer G plant den Unterricht in Klasse 11c. Diese Klasse kennt er noch nicht. Lehrer G denkt: „Zur Unterrichtsplanung orientiere ich mich an didaktischem Wissen.“	8	0
8	Lehrerin E bespricht zu Schuljahrsbeginn in ihrer Klasse den Englisch-Unterricht der nächsten Monate. Lehrerin E sagt: „Ich bin zufrieden, wenn wir gut vorankommen.“	5	1 (1), 2 (2)
9	Lehrer Ws Klasse hat seine hochgesteckten Ziele in diesem Schuljahr nicht erreicht. Lehrer W denkt: „Die Klasse muss sich mehr anstrengen.“	0	1 (1), 7 (2), 8 (3), 10 (2)
9	Lehrerin Q erwartet von ihren SchülerInnen Pünktlichkeit. Sie selbst kommt häufig zu spät, worauf sie die SchülerInnen ansprechen. Lehrerin Q sagt: „Das lässt sich leider nicht vermeiden, da ich sehr viel zu tun	2	5 (5), 10 (1)
10	Lehrer C hat einen wissenschaftlichen Artikel über effektive allgemeine Unterrichtsmethoden gelesen. Lehrer C denkt: „Dies werde ich auf jeden Fall umsetzen.“	8	0
7	Lehrerin As Klasse schneidet in einem standardisierten Rechentest eher schlecht ab. Lehrerin A überlegt, was sie tun sollte. Lehrerin A denkt: „Ich werde problematische Aufgabenstellungen	1	8 (6), 10 (1)

Kompetenz	Item	Übereinstimmung mit Kompetenz*	weitere genannte Kompetenz(en) (Anzahl Nennungen)*
6	In Lehrer Zs Klasse haben sich zwei Schülergruppen gebildet, die sich feindlich gegenüber stehen. Lehrer Z thematisiert die Situation.	8	0
4	Lehrer Z sagt: „In der Klasse haben sich zwei Lager gebildet, die sich Schülerin J spricht in den Pausen häufig begeistert von der Kfz-Werkstatt ihres Onkels. Lehrer B möchte dieses Interesse fördern.	6	3 (1), 7 (1)
3	Lehrer B sagt: „Möchtest du an der Technik-AG teilnehmen?“ Lehrerin G übt mit den Schülern ihrer siebten Klasse das Verstehen von Sachtexten. Sie gibt der Klasse Arbeitsanweisungen zu einem neuen Text.	8	0
1	Lehrerin G sagt: „Lest bitte den Text Lehrerin A plant den Unterricht für die Klassen 5c und 8a. Sie überlegt dabei, welche Methoden sie anwenden soll.	8	0
8	Lehrerin A denkt: „Welche Methoden und Medien haben mir in meiner Schulzeit gut gefallen?“ Schüler B zeigt im Vergleich zur Klasse eher schwache Leistungen. In einer Klassenarbeit schneidet B nicht gut ab. Lehrer I hat das erwartet.	1	3 (2), 6 (2), 7 (3)
9	Lehrer I sagt zu B: „Schade, das habe ich leider befürchtet.“ Lehrer A und Lehrer J diskutieren über ihre Wertvorstellungen bezüglich ihres Berufs. Beide vertreten unterschiedliche Meinungen.	6	1 (5), 10 (1)
	Lehrer J denkt: „Ich denke darüber nach, ob meine Werte die richtigen		

Kompetenz	Item	Übereinstimmung mit Kompetenz*	weitere genannte Kompetenz(en) (Anzahl Nennungen)*
10	<p>Lehrerin S hat von einer neuartigen Unterrichtsmethode gehört, die interessant zu sein scheint.</p> <p>Lehrerin S denkt: „Ich werde ein Seminar zu dem Thema besuchen.“</p>	7	5 (1)
11	<p>Lehrerin M plant eine Projektwoche mit ihrer Klasse. Sie überlegt, wie sie das Projekt gestalten soll.</p> <p>Lehrerin M denkt: „Ich lese nach, welche Forschungsergebnisse es zur Gestaltung von Projekten gibt.“</p>	0	5 (1), 10 (7)
5	<p>Lehrer O möchte Umgangsregeln in seiner Klasse einrichten.</p> <p>Lehrer O sagt: „Ich habe einen Regelkatalog ausgearbeitet, an den ihr euch halten sollt.“</p>	7	6 (1)
8	<p>Lehrer K möchte die SchülerInnen seiner Klasse beurteilen.</p> <p>Lehrer K denkt: „Ich kenne diese Klasse schon lange, deshalb kann ich mich ganz auf mein Urteil verlassen.“</p>	7	7 (1)
9	<p>Die LehrerInnen einer Schule diskutieren über die rechtlichen Rahmenbedingungen des Lehrberufs.</p> <p>Lehrerin W sagt: „Es ist wichtig, die rechtlichen Rahmenbedingungen zu kennen, um sicher handeln zu können.“</p>	7	10 (1)
11	<p>Lehrerin M plant ein großes Schulprojekt und fragt Lehrerin L, ob sie mitarbeiten möchte.</p> <p>Lehrerin L sagt: „Ich nehme nicht teil. Ich möchte die Zeit lieber in die Gestaltung meines Unterrichts</p>	7	9 (1)

Kompetenz	Item	Übereinstimmung mit Kompetenz*	weitere genannte Kompetenz(en) (Anzahl Nennungen)*
2	<p>Lehrer K möchte im Physik-Unterricht die Elektrik-Kenntnisse seiner SchülerInnen vertiefen.</p> <p>Lehrer K sagt: „Ich erkläre euch nun, wie ein Radio funktioniert.“</p>	3	3 (5)
3	<p>Lehrer P kündigt eine Klassenarbeit an. Er möchte die SchülerInnen auf das Lernen vorbereiten.</p> <p>Lehrer P sagt: „Es geht um eine ganze Menge Stoff, also müsst ihr viel lernen.“</p>	8	0
3	<p>Ein Schüler hat bei einer Klassenarbeit schlecht abgeschnitten. Lehrer H spricht mit ihm über das Ergebnis.</p> <p>Lehrer H sagt: „Schade, dass du nicht gut abgeschnitten hast. Dir liegt das Fach vielleicht einfach nicht.“</p>	0	7 (7), 8 (1)
5	<p>Lehrerin S möchte hilfsbereites Verhalten ihrer SchülerInnen fördern.</p> <p>Lehrerin S sagt: „Ihr sollt euch hilfsbereit verhalten.“</p>	8	0
3	<p>Lehrer Z möchte in der nächsten Stunde in ein neues Thema einführen. Er überlegt, wie er vorgehen soll.</p> <p>Lehrer Z denkt: „Ich gebe den SchülerInnen ein Problem vor und lasse sie nach einer Lösung suchen.“</p>	6	1 (2)
4	<p>Die Schüler B und C hinken im Mathematikunterricht etwas hinterher und haben insbesondere Schwierigkeiten mit Textaufgaben.</p> <p>Lehrerin K gibt einen Arbeitsauftrag.</p> <p>Lehrerin K sagt: „Bitte bearbeitet die</p>	1	1 (5), 3 (1), 7 (1)

---

Kompetenz	Item	Übereinstimmung mit Kompetenz*	weitere genannte Kompetenz(en) (Anzahl Nennungen)*
10	Lehrer C hat einen wissenschaftlichen Artikel über effektive allgemeine Unterrichtsmethoden gelesen. Lehrer C denkt: „Dies lässt sich in meine Praxis nicht übertragen.“	<b>7</b>	1 (1)

---

\* Fettdruck: Mehrheitliche Übereinstimmung, Kursivdruck: Keine Übereinstimmung mit intendierter Kompetenz