

Katinka Clasen

Psychometrische Qualität von Multiplen Mini-Interviews (Masterarbeit)

Reihe: Akademische Qualifikationsarbeiten 2018



GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Zum Geleit

Im Rahmen der gemeinsamen „Qualitätsoffensive Lehrerbildung“ von Bund und Ländern erhält die Universität Erfurt in diesen Jahren Fördermittel des Bundesministeriums für Bildung und Forschung für die Fort- und Weiterentwicklung der Lehrerbildung. Eines der Teilprojekte des Erfurter Vorhabens QUALITEACH ist das „Teaching Talent Center“, das es sich zur Aufgabe gemacht hat, Talente für den Lehrberuf zu identifizieren und zu fördern sowie den Lehrberuf insgesamt für talentierte junge Menschen attraktiv zu machen.

Eine wichtige Basis für diese Arbeit ist die Verfügbarkeit von Diagnose- und Assessmentverfahren, mit denen Persönlichkeitsmerkmale von Lehramtsstudierenden oder gegebenenfalls von Bewerberinnen und Bewerbern für das Lehramtsstudium erfasst, analysiert und in Folge systematisch weiterentwickelt werden können.

Die vorliegende Arbeit ist als Masterarbeit im Studiengang Psychologie entstanden. Unter meiner Betreuung werden fortlaufend Projektstudien, Methodenstudien und Masterarbeiten durchgeführt, in denen Assessmentverfahren entwickelt, erprobt und evaluiert werden.

Der anschließend dargestellte Bericht entspricht dem Original der Masterarbeit von Frau Clasen, der ich für die Erlaubnis danke, den Bericht an die Fachöffentlichkeit weiterzugeben. Für Kolleginnen und Kollegen werden die sehr detailliert dargestellte Vorgehensweise sowie die statistischen Auswertungen eine wertvolle Anregung sein. Ich bitte um Verständnis, dass das *Aufgabenmaterial* und die *Auswertungsbögen* aus dem Anhang entfernt wurden. Da die Materialien weiterhin eingesetzt werden sollen, wäre eine Verbreitung in der Öffentlichkeit kontraproduktiv. Fachkolleg*innen können das Material aber bei uns anfordern.

Die Studie von Frau Clasen bezieht sich auf drei Performance Assessment-Aufgaben, bei denen die Proband*innen innerhalb von wenigen Minuten Handlungsaufgaben zu bearbeiten hatten. Die Tätigkeiten der Proband*innen wurden aufgezeichnet und anschließend von mehreren Beurteiler*innen ausgewertet. Bislang konnten von den umfangreichen Erhebungen jeweils 50 Performances gesichtet werden. Zu unser aller Überraschung erwies es sich als sehr schwierig, bei Einsatz verschiedener Beurteilender zu vergleichbaren Einschätzungen zu gelangen. Nach den derzeitigen Erkenntnissen müssten drei bis fünf Beurteilungen vorliegen, um eine einigermaßen aussagekräftige Einschätzung der Leistung der Teilnehmenden zu erhalten. Allerdings deckt sich dieses Erkenntnis mit den Befunden anderer Autor*innen.

Die Entwicklung und Erprobung der Multiplen Mini-Interviews wird in den kommenden Semestern fortgesetzt. Über die Ergebnisse werden wir berichten.

Prof. Dr. Ernst Hany

Zitiervorschlag für dieses Papier:

Clasen, K. (2018). *Psychometrische Qualität von Multiplen Mini-Interviews. Masterarbeit zur Erlangung des akademischen Grades „Master of Science (M. Sc.)“ im Studiengang Psychologie*. Reihe Akademische Qualifikationsarbeiten aus dem Teaching Talent Center, hrsg. von Ernst Hany. Erfurt: Universität Erfurt, Erfurt School of Education, Projekt QUALITEACH.

Kontaktadresse: Katinka Clasen, M.Sc., Fachgebiet Psychologie, Universität Erfurt, Nordhäuser Str. 63, D-99089 Erfurt. E-Mail: katinka.clasen@uni-erfurt.de

Universität Erfurt

Erziehungswissenschaftliche Fakultät

Fachgebiet Psychologie

Lehrstuhl für Pädagogisch-psychologische Diagnostik und Differentielle Psychologie

Masterarbeit

zur Erlangung des akademischen Grades „Master of Science (M. Sc.)“

im Studiengang Psychologie

Psychometrische Qualität von Multiplen Mini-Interviews

Vorgelegt im Wintersemester 2017/18 von:

Katinka Clasen

Thälmannstraße 23

99085 Erfurt

Matrikelnummer: 39442

Erstprüfer:

Prof. Dr. Ernst Hany

Zweitprüferin:

Nastassja Simeth (M. Sc.)

Zusammenfassung

Das Ziel der vorliegenden Arbeit ist es, die psychometrische Qualität des Verfahrens der Multiplen Mini-Interviews im Anwendungsbereich der Diagnostik lehrberuflicher Kompetenzen zu untersuchen. Hierzu wurden sechs konzipierte MMI-Aufgaben mit Lehramtsstudierenden der Universität Erfurt erprobt und die Leistungen mit Hilfe von zwei bis drei Beurteilern über erstellte Beurteilungsbögen ausgewertet. Drei der sechs MMI-Aufgaben, welche Berufswahl-, Instruktions- und Planungskompetenzen zu messen beabsichtigen, gingen in die Analyse der vorliegenden Arbeit ein, um die psychometrischen Gütekriterien Reliabilität, Objektivität, Validität und Ökonomie anhand einer Stichprobe von 50 Studierenden zu bestimmen. Überprüfungen der Beurteilerübereinstimmung und Skalen der Beurteilungsbögen weisen darauf hin, dass die Reliabilität und Objektivität des MMIs stark von der Konzeption der einzelnen Beurteilungsbögen abhängt. Zur Validität liefert die vorliegende Arbeit unter Einbezug von Globaleinschätzungen durch Testleiter und Fähigkeits- und Persönlichkeitsselbsteinschätzungen der Testpersonen heterogene Befunde. Die grundsätzlich gute Ökonomie der Anwendung des Verfahrens wird durch den enorm hohen Aufwand der Konzipierung und Überprüfung der MMI-Aufgaben vermindert. Die psychometrische Qualität des MMI als Gesamtverfahren lässt sich vor dem Hintergrund der uneinheitlichen Ergebnisse für die drei Aufgaben nicht global bewerten.

Inhaltsverzeichnis

1 Einleitung	1
2 Theoretischer Hintergrund	4
2.1 Anforderungen an den Lehrberuf	4
2.1.1 Kompetenzmodelle	5
2.1.2 Hervorgehobene Kompetenzbereiche ohne explizierte Modelleinbettung	11
2.2 Kompetenzentwicklung in der Lehrerbildung	18
2.2.1 Kompetenzen von Lehramtsstudierenden	20
2.2.2 Diagnostik und Förderung von Kompetenzen in der Lehrerbildung	25
2.3 Verfahren zur Eignungsdiagnostik und Leistungsbeurteilung	29
2.3.1 Berufseignungsdiagnostik	30
2.3.2 Performance Assessment	33
2.3.3 Selbsteinschätzungsverfahren	35
2.3.4 Multiple Mini-Interviews	36
3 Fragestellungen der vorliegenden Arbeit	42
4 Methode	45
4.1 Gestaltung und Ablauf der Erhebungen	45
4.2 Erhebungsverfahren	47
4.2.1 Multiple Mini-Interviews	48
4.2.2 Globaleinschätzungen	54
4.2.3 Online-Fragebogen	56
4.3 Stichprobe	57
4.4 Auswertungsmethodik	59
5 Ergebnisse	63
5.1 Item- und Reliabilitätsanalysen	63
5.1 Validitätsanalysen	77
6 Diskussion	85
Literaturverzeichnis	94
Abbildungsverzeichnis	104
Tabellenverzeichnis	105
Anhang	106

Anmerkung:

Aus Gründen der leichten Lesbarkeit wird in der vorliegenden Arbeit die gewohnte männliche Sprachform bei personenbezogenen Substantiven und Pronomen verwendet. Dies impliziert jedoch keine Benachteiligung des weiblichen Geschlechts, sondern soll im Sinne der sprachlichen Vereinfachung als geschlechtsneutral zu verstehen sein.

1 Einleitung

Der Lehrberuf – er unterliegt dem ständigen Wandel der gesellschaftlich verankerten Erwartungen an das Gelingen des Bildungs- und Schulsystems. Lehrkräfte befinden sich zunehmend im Spannungsfeld unterschiedlicher Erwartungsträger, seien dies Eltern, Kollegen oder die Schüler selbst (Sieland, 2001). Angehende Lehrer und Lehrerinnen müssen heutzutage einem zunehmend vielschichtigen Anforderungsprofil genügen als dies noch vor dreißig Jahren der Fall war. Doch die charakteristische Unbestimmbarkeit und Unsicherheit insbesondere der pädagogischen Tätigkeiten im lehrberuflichen Alltag gestalten die hohen Ansprüche an professionelles Lehrerhandeln als eine große Herausforderung (Baumert & Kunter, 2006; Combe & Kolbe, 2008; Weiß, Schramm & Kiel, 2014). Es existiert mittlerweile eine Vielzahl an Versuchen in der Forschung zum Lehrberuf, Anforderungen und benötigte Kompetenzen von Lehrkräften herauszuarbeiten und daraus Maßnahmen zur Qualitätssicherung abzuleiten. „Was ist eine gute Lehrperson?“ oder „Was macht eine gute Lehrkraft aus?“ lauten die leitenden Überschriften vieler Forschungsarbeiten – die Antworten auf diese Fragen zeigen sich nicht selten als Modelle und Strukturen diverser Kompetenzfacetten, Expertisen oder Professionscharakteristika von Lehrpersonen (Mulder & Gruber, 2011).

Doch wie erreichen Lehrkräfte in der Praxis die Aneignung dieser unterschiedlichen Fähigkeiten, um kompetent und professionell handeln zu können? Wann beginnt die Entwicklung der erwünschten Könnensbereiche? Die von der Kultusministerkonferenz festgelegten Standards der Lehrerbildung verorten die Kompetenzentwicklung aus politischer Sicht in der Ausbildungszeit von Lehrkräften (KMK, 2004). Die notwendigen Fähigkeiten und Kompetenzbereiche werden hierbei als Standards formuliert, welche im Rahmen der Ausbildung angeeignet und nach Abschluss dieser in wünschenswerten Ausprägungen vorhanden sein sollten. Die Ausformulierung von Standards fokussiert jedoch wiederum nur den zu erwartenden Zustand, ohne dabei Ansätze für den zielführenden Prozess zu liefern (Frey & Jung, 2011). Die Aneignung der Standards soll demzufolge über die Lehrerbildung erfolgen, doch führen Befunde darüber, dass das Studienangebot nur kaum zur Entwicklung beruflicher Kompetenz beiträgt (Mayr, 2006a), zu der Frage, inwiefern das Studium tatsächlich auf die berufliche Tätigkeit vorbereitet.

Aus diesem Anlass heraus verlagern sich die Forschungen zum Lehrberuf immer mehr auf die Lehrerbildung und untersuchen Ansätze, wie sich geeignete Lehrer gewinnen, ausbilden und fördern lassen (Weyand, Justus & Schratz, 2012). Kompetenzdiagnostik, -feedback und entsprechende Entwicklungsangebote bereits während der Lehrerbildung scheinen ein möglicher Versuch zu sein, berufliche Kompetenzen angehender Lehrkräfte durch eine direkte und transparente Ausrichtung des Studiums im erwünschten Ausmaß herauszubilden (Schratz, 2012). Die vorliegende

Arbeit setzt in diesem Kontext bei der Kompetenzdiagnostik in der Lehrerbildung an. Um die Kompetenzentwicklung in Richtung der gewünschten Ziel- oder Sollausprägungen der erforderlichen Fähigkeiten anzuregen, bleibt es unumgänglich, zuvor den realen Ist-Zustand zu identifizieren. Nur so kann überhaupt das Entwicklungspotential festgestellt und für weitere Interventionen genutzt werden. Die Kompetenzdiagnostik stellt somit ein entscheidendes Bindeglied dar, um anhand der fähigkeitsbezogenen Voraussetzungen und Potentiale eine optimale Passung zu den Anforderungen des Lehrberufs zu erreichen.

Um die besonders handlungsorientierten Kompetenzfacetten und deren Vorläuferfähigkeiten für das Gelingen der Lehreraarbeit zuverlässig diagnostizieren zu können, bedarf es eines Diagnoseinstruments, welches starke Bezüge zur tatsächlichen Tätigkeit aufweist. Auf diese Weise werden die eigentlich geforderten Kompetenzfacetten auch für die Studierenden bereits während ihrer Ausbildung transparent. Denn so ist es im Lehramtsstudium in Deutschland aktuell noch der Fall, dass Studierende in der Regel ihre berufliche Eignung oder eben Nicht-Eignung erst am Ende ihres Studiums begreifen (Rothland & Terhart, 2011). Diagnostikverfahren, welche bereits während der Ausbildung eingesetzt werden und welche die tatsächlichen Anforderungen des Lehrberufs authentisch simulieren, ermöglichen auch für Studierende eine frühzeitigere Überprüfung der eigenen Passung für den Lehrberuf.

Die Ansprüche an eine solche Kompetenzdiagnostik sind jedoch nicht nur inhaltlich hoch, sondern erfordern auch eine entsprechende Qualität der Messung als solche. Soll den Studierenden ein realistisches Bild ihrer eigenen Kompetenzen zurückgemeldet werden, so müssen die zu messenden Fähigkeiten zum einen zuverlässig und genau erfasst werden, um zwischen den interindividuellen Ausprägungen differenzierte Aussagen treffen zu können. Zum anderen sollte das Verfahren überhaupt auch in der Lage sein, die zu messen vorgegebenen Kompetenzen tatsächlich abzubilden. Außerdem dürfen Messung und Auswertung der Kompetenzen nicht von subjektiven Urteilen der Durchführenden abhängen. Es bestünde sonst die Gefahr, dass das Verfahren als solches von den Studierenden überhaupt nicht akzeptiert und ernst genommen wird. Von Seiten der durchführenden Hochschule muss das Instrument aufgrund der Vielzahl an Lehramtsstudierenden außerdem ressourcensparend einsetzbar sein, denn so stellt es zum regulären Prüfungsaufwand nur eine spezifische Ergänzung dar, welche weder Unmengen an Zeit noch an Personal in Anspruch nimmt.

Als ein potentiellcs Verfahren der Kompetenzdiagnostik im Zuge der Lehrerbildung wird dem besonders simulationsorientierten Verfahren der Multiplen Mini-Interviews viel Zuversicht entgegen gebracht. Das Verfahren stammt eigentlich aus dem medizinischen Kontext und findet in der Auswahl geeigneter Studierender seinen Einsatz (Knorr & Hissbach, 2014). Im englischsprachigen

gen Raum wurde das MMI bereits auf seine Qualität hin untersucht und als potentiell messgenaues, zuverlässiges und im Gegensatz zu Einstellungsinterviews ökonomisches Verfahren konstatiert (siehe Pau et al., 2013). Die vorliegende Arbeit soll einen Beitrag dazu leisten, Aussagen über die psychometrische Qualität des Verfahrens im Anwendungskontext von lehrberufsrelevanten Fähigkeiten und Kompetenzen treffen zu können. Das Potential, welches das MMI aufgrund der Anwendung als Auswahlverfahren im medizinischen Bereich aufzuweisen scheint, muss für den Einsatz in der Lehrerbildung zunächst umfassend geprüft werden.

Zuvor soll in dieser Arbeit ein Überblick über die aktuelle Diskussion um die Qualitätssicherung im Lehrberuf auf Grundlage der hervorgehobenen beruflichen Anforderungen gegeben werden. Nach Erläuterung von bedeutsamen Kompetenzbereichen für das erfolgreiche Handeln als Lehrkraft soll der Fokus auf die Lehrerbildung gelegt werden. Hierzu wird herausgearbeitet, dass die Ermöglichung einer Passung zwischen Lehrkräften und den hohen beruflichen Anforderungen bereits in der Ausbildung angestrebt werden sollte und es werden bisherige Interventionen zur Kompetenzentwicklung gezeigt. Daraufhin sollen theoretische Grundlagen der Eignungs- und Leistungsdiagnostik erläutert und das hier bedeutsame Verfahren der Multiplen Mini-Interviews ausführlich dargestellt werden. Der empirische Kern der Arbeit liegt im Anschluss auf der Analyse der psychometrischen Qualität eines entwickelten Satzes von lehrberufsspezifischen MMI-Aufgaben und soll Aufschluss darüber geben, inwiefern sich das Verfahren zur Kompetenzdiagnostik in der Lehrerbildung eignet.

2 Theoretischer Hintergrund

2.1 Anforderungen an den Lehrberuf

Aufgrund der komplexen Struktur des Lehrberufs lassen sich nur schwerlich klare und strukturierte Anforderungen festlegen (Weiß, Schramm & Kiel, 2014). Der berufliche Alltag von Lehrern gestaltet sich besonders im Teilbereich des pädagogischen Handelns als undurchschaubar und ungewiss, so dass die Tätigkeit nur schwer in standardisierte Anforderungsprofile überführbar ist. Die Vielfalt und Komplexität der lehrberuflichen Anforderungen werfen bis heute die Frage danach auf, was eine gute Lehrperson überhaupt ausmacht und welche Grundlagen an Wissen und Können ein Lehrer aufweisen muss, um professionell zu handeln (Baumert & Kunter, 2006; Blömeke, 2011; Combe & Kolbe, 2008; Keller-Schneider, 2011; Mulder & Gruber, 2011). Zwar ist der Hauptgegenstand des Lehrberufs unbestritten das Erteilen von Unterricht und die Initialisierung von Lernprozessen bei den zu unterrichtenden Schülern, jedoch herrschen bezüglich des Erfolgs der lehrenden Tätigkeit nur unsichere und kaum operativ eindeutige Vorgaben vor (Baumert & Kunter, 2006).

In der Forschung existieren mittlerweile zahlreiche Versuche, die Anforderungen an den Lehrberuf zu systematisieren, Voraussetzungen für eine professionelle Ausübung des Berufs festzulegen sowie Merkmale der „guten“ Lehrperson multidimensional zu beschreiben. Dabei haben sich drei Forschungsstränge ergeben, die aus unterschiedlichen Perspektiven entweder die Merkmale der Lehrerprofessionalität, die benötigten Kompetenzen oder die Entstehung von Expertise zu ihrem Gegenstand machen (Mulder & Gruber, 2011). Die vorliegende Arbeit legt das primäre Augenmerk auf die Kompetenzforschung zum Lehrberuf, welche im Rahmen von Voraussetzungsanalysen benötigte Kompetenzen für eine erfolgreiche Berufsbewältigung formuliert. Solche Kompetenzen lassen sich in ihrer Entwicklung grundsätzlich bereits im Lehramtsstudium verorten und werden in der Praxis über so genannte Lehrbildungsstandards von der Kultusministerkonferenz (2004) umgesetzt. Das Erforschen und Festlegen von erwünschten Kompetenzen dient letztendlich dem übergeordneten Ziel, das lehrberufliche Handeln zur Zufriedenheit aller zu optimieren, verbindliche Anforderungen an das Lehren aufzustellen und somit den Lehrberuf zu professionalisieren (Hascher, 2004; Weiß et al., 2014).

In den folgenden Kapiteln soll zunächst auf ausgewählte Kompetenzmodelle aus der Fachliteratur zum Lehrberuf eingegangen und anschließend in diesem Rahmen häufig betonte Kompetenzfacetten ohne explizite Modelleinbettung erläutert werden. Aufgrund von Veränderungen im Belastungserleben des schulischen Alltags werden weiterhin Faktoren und Kompetenzen aus dem Bereich der beruflichen Gesundheit und Zufriedenheit als Anforderungen an den Lehrberuf erläutert.

2.1.1 Kompetenzmodelle

Im Kontext der Debatte darum, was einen guten und professionell handelnden Lehrer ausmacht, sind verschiedene Entwürfe zu berufsbezogenen Kompetenzen entstanden, die angeben sollen, welche Fähigkeiten und Einstellungen eine Lehrperson für das erfolgreiche Ausüben des Berufs aufweisen muss (Baumert & Kunter, 2006; Messner, 2007). Häufig versuchen die Autoren, ihre Entwürfe zu den Komponenten der Lehrerkompetenzen anhand eines übergeordneten Kompetenzmodells zu strukturieren (Frey & Jung, 2011). Doch bevor zwei dieser Modelle exemplarisch dargestellt werden, muss zunächst einmal definiert werden, wann und inwiefern eine Person überhaupt kompetent handelt. So lässt sich grundsätzlich festhalten, dass nicht nur spezifisches Wissen oder bestimmte Fähigkeiten Kompetenz ausmachen, sondern immer auch das Können im Sinne einer aktiven Anwendung des Wissens von Bedeutung ist (Baumert & Kunter, 2006; Frey & Jung, 2011).

Häufig ist hier auch die Rede von Handlungskompetenz, denn so befähigen Wissen und Fertigkeiten eine Person innerhalb einer Situation auftretende Probleme oder Anforderungen aktiv zu bewältigen, sprich kompetent und verantwortungsvoll zu handeln (Frey, 2004, 2008; Terhart, 2007). Kompetenz wird somit denjenigen zugesprochen, die in der Lage sind, „so tätig zu werden, dass sie eine Absicht, ein Ziel oder ein Zweck unter Beachtung von Handlungsprinzipien, Werten, Normen und Regeln, mit Bezug auf konkrete, die jeweilige Handlungssituation bestimmende Bedingung, zu erreichen vermag“ (Frey, 2006, S. 31). Ganzheitliche Kompetenz besteht also aus der multidimensionalen Zusammenkunft von psychischen und physischen Fähigkeiten, die einer erfolgreichen Bewältigung von Anforderungen zugrunde liegen (Terhart, 2007). In den meisten Kompetenzmodellen wird sich häufig auf diese grundlegenden Voraussetzungen für kompetentes Handeln fokussiert (Frey, 2008).

Die nötige Qualität von kompetenten Handlungen versuchen Standardmodelle im Sinne von ethischen Normen und Ausbildungszielen, also erwarteten Kompetenzausprägungen, abzumachen. Sie entspringen in erster Linie der Bildungspolitik und beziehen sich dabei vordergründig auf Expertenmeinungen und weniger auf theoretische Kompetenzmodelle aus der Forschung (Frey & Jung, 2011; Helmke, 2009; Oser, Curcio & Duggeli, 2007). Anforderungen über die Kompetenzen von angehenden Lehrern postuliert etwa die Kultusministerkonferenz (2004) durch ihre Standards für die Lehrerbildung in den vier Kompetenzbereichen „Unterrichten“, „Erziehen“, „Beurteilen“ und „Innovieren“. Insgesamt definieren elf spezifische Kompetenzen die Zielsetzungen der Lehrerbildung in den theoretischen und praktischen Abschnitten. Trotz der häufig bemängelten fehlenden theoretischen Fundierung von Standardmodellen weisen sie viele Parallelen zu Kompetenzmodellen und deren grundlegenden Elementen auf (Frey & Jung, 2011).

Das Hierarchische Strukturmodell der Handlungskompetenz nach Frey (2004) teilt die generelle Handlungskompetenz einer Lehrperson auf der höchsten Ebene in vier Kompetenzklassen: Die Fach-, Methoden-, Sozial- und Personalkompetenz, welche strukturell voneinander abhängig und miteinander verwoben sind. Innerhalb der Fachkompetenzklasse sind jene Fähigkeiten subsumiert, die stark mit der spezifischen Berufsdisziplin zusammenhängen. Dies betrifft das jeweilige Fachwissen, Kenntnisse der pädagogisch-psychologischen Grundlagen, der Didaktik und allgemeinen Unterrichtsmethoden sowie Wissen über Lehrpläne und der Diagnostik (Frey, 2006; Frey & Jung, 2011). Elemente aus der Methodenkompetenzklasse beziehen sich auf die Fähigkeiten, zielorientiert und effizient agieren zu können, wozu es an Selbstorganisation, Flexibilität, Ressourcenmanagement, Zielorientierung und Routine bedarf (Frey, 2008; Frey & Jung, 2011). Fähigkeiten, welche die Lehrkraft für Anforderungen in kooperativ-interaktiven Settings benötigt, werden der Sozialkompetenzklasse zugeschrieben. Hierzu zählen Strategien der Kommunikation, Kooperation und Interaktion, ebenso wie durch Empathie geleitete Konflikt- und Motivationsfähigkeiten (Frey, 2006, 2008; Frey & Jung, 2011). Die letzte Kompetenzklasse umfasst schließlich Fähigkeitskonzepte, die sich stark auf die eigene Person und deren Funktionieren beziehen lassen. So erhalten in diesem Bereich Ausprägungen der Selbstreflektion, des Selbstbilds, unterschiedlicher Werthaltungen sowie Bewältigungsstile beruflicher Belastungen durch Gelassenheit, Geduld und Anpassbarkeit ihre Bedeutung (Frey, 2008; Frey & Jung, 2011).

Wie bereits durch die Beschreibungen der einzelnen Kompetenzklassen sichtbar wird, definieren sich die übergeordneten Kompetenzen über verschiedene Fähigkeitsdimensionen. Somit weist jede Kompetenzklasse ein Bündel aus verschiedenen Fähigkeiten auf, welche gemeinsam festgelegte Fähigkeitsdimensionen repräsentieren. Als Beispiel setzt sich etwa die Sozialkompetenz aus den Fähigkeitsdimensionen der Kommunikation, Kooperation, Interaktion, Empathie, Wertschätzung und der Motivierung zusammen. Jede dieser einzelnen Fähigkeiten gestaltet sich als ein übergeordnetes Konstrukt, das ein relativ hohes Abstraktionsniveau besitzt. Erst durch das Zuweisen von konkreten Verhaltensweisen, also dem bestimmbaren Können, erhält eine Fähigkeitsdimension greifbaren Handlungsbezug. Auf dieser untersten Ebene spricht Frey (2004) in diesem Kontext von einzelnen Fertigkeiten, die durch routinierte Ausübung auch ohne bewusste Steuerung ausgeführt werden können und eine Fähigkeitsdimension ausmachen. Sich klar und präzise auszudrücken stellt beispielsweise eine Fertigkeit dar und kann der Fähigkeitsdimension der Kommunikation zugeordnet werden, welche wiederum ein Teil der Sozialkompetenz einer Lehrkraft ausmacht. Handlungskompetenz bedarf schlussendlich auf der untersten Ebene des hierarchischen Strukturmodells distinkter Fertigkeiten einer Person. Ergänzt wird dieses eigene Können immer auch durch die individuellen Selbstkonzepte dieser einzelnen Fertigkeiten, also der Einschätzung des eigenen Könnens durch die Person selbst (Frey, 2004).

Fasst man also konkrete Fertigkeiten zu Fähigkeiten zusammen, bündelt diese in Fähigkeitsdimensionen, ergeben sich die vier Kompetenzklassen, welche gemeinsam die Handlungskompetenz einer Person repräsentieren. Insgesamt weist das Modell somit vier Ebenen auf, die hierarchisch aufeinander aufgebaut sind. Abbildung 1 verdeutlicht noch einmal die einzelnen Ebenen und führt diese exemplarisch aus.

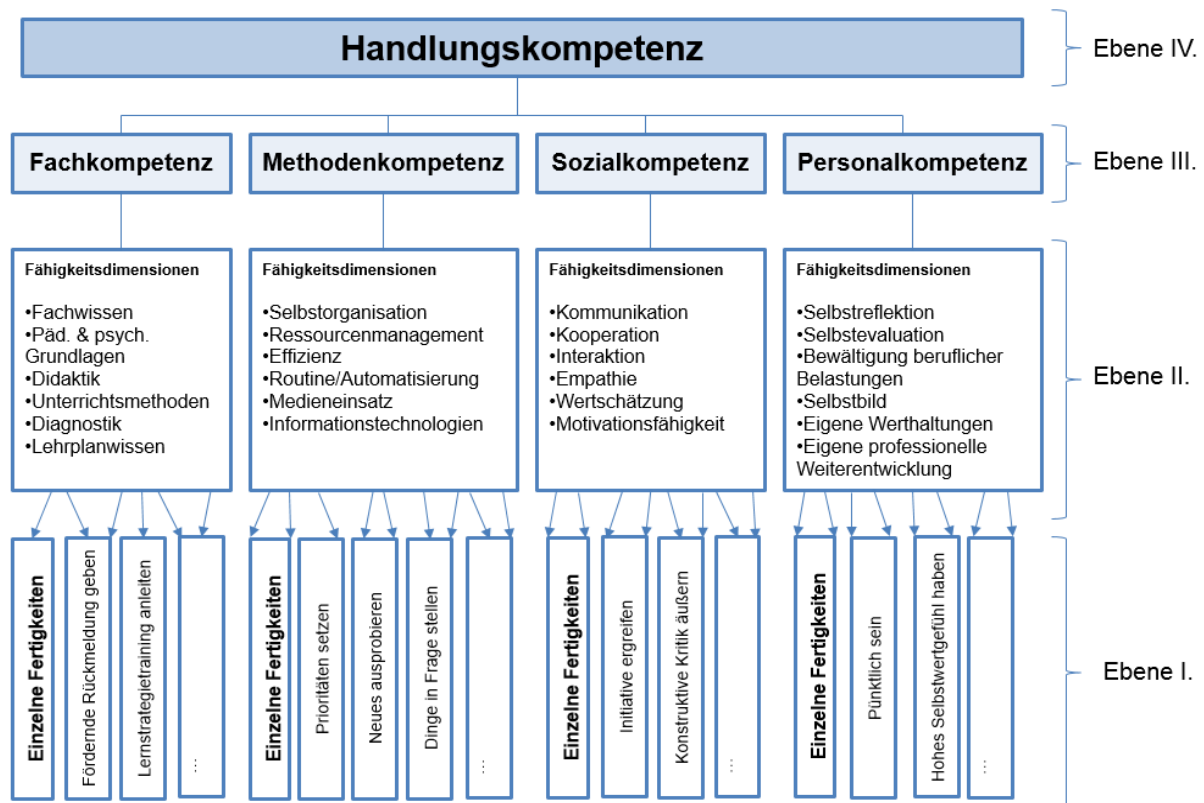


Abbildung 1. Hierarchisches Strukturmodell von Handlungskompetenz nach Frey (2008). Darstellung des Modells ergänzt durch Zusammenfassungen aus Frey & Jung (2011).

Ein weiteres Modell, das die Handlungskompetenz von Lehrern mittels verschiedenen Kompetenzfacetten aufzugliedern versucht, ist das generische Modell professioneller Handlungskompetenz von Baumert und Kunter (2006). In diesem Ansatz liegt der Fokus weniger auf spezifischen Fähigkeiten, als auf Aspekten des berufsspezifischen Wissens und Könnens und verfolgt einen sehr umfassenden Ansatz. Durch ein Zusammenspiel der folgenden vier Elemente ergibt sich als Kern der Theorie die professionelle Handlungskompetenz: Professionswissen, Überzeugungen und Werthaltungen, motivationale Orientierungen sowie selbstregulative Fähigkeiten. Eine Übersicht zu der Zusammensetzung der Handlungskompetenz gibt Abbildung 2.



Abbildung 2. Modell professioneller Handlungskompetenz nach Baumert & Kunter (2006).

Das Professionswissen ist hierbei das am stärksten untergliederte Element und besteht wiederum aus verschiedenen Bereichen des pädagogischen, fachlichen und fachdidaktischen Wissens, sowie Aspekten des Organisations- und Beratungswissens. Es lässt sich aufgrund seiner Zusammensetzung aus Wissen und Expertise auch als so genannte Sachkompetenz der Lehrperson subsummieren (Helmke, 2009). Jeder einzelne Wissensbereich des Professionswissens weist wiederum eine Vielzahl von spezifischen Wissensfacetten auf; so ist etwa das pädagogische Wissen auf den Inhalten zu konzeptuellen bildungswissenschaftlichen Grundlagen, allgemeindidaktischer Konzeption und Planung, Unterrichtsführung und Orchestrierung von Lerngelegenheit sowie fachübergreifende Prinzipien des Diagnostizierens, Prüfens und Bewertens fundiert (Baumert & Kunter, 2006, S. 485). Das pädagogische Wissen fußt dabei auch auf entwicklungspsychologischen Grundlagen sowie Verständnis von Lernprozessen und befähigt die Lehrkraft dazu, Lernende anzuleiten, sowie ein lernförderliches Klima und eine entsprechende Arbeitskultur zu schaffen (Keller-Schneider, 2011). Als besonderen Kern der professionellen Kompetenz im Bereich des Professionswissens erweist sich das fachliche Wissen, auf dessen Grundlage fachdidaktische Handlungen erst ihre Wirkung zeigen können (Baumert & Kunter, 2006). Beratungswissen gewinnt insbesondere im Zuge der Kommunikation mit Eltern oder Schülern an Bedeutung. Es befähigt die Lehrperson dazu, Schüler in ihrem Lernprozess dann zu unterstützen, wenn diese auf Hürden beim Verstehen treffen. Dabei liefert das Beratungswissen nötige Lehr- und Lernstrategien, um das Lernen der Schüler erfolgreich zu begleiten (Keller-Schneider, 2011).

Das Element der Werthaltungen und Überzeugungen wird kategorial vom Professionswissen abgegrenzt, denn es enthält weniger Kompetenzfacetten des Wissens, sondern zeichnet sich durch individuelle Vorstellungen und indirekte subjektive Theorien der Lehrperson aus. Im Gegen-

satz zum sachlichen Wissen über Unterricht und Didaktik sind Überzeugungen zum Lehren und Lernen eher durch einen klaren Wertbezug geprägt (Lipowsky, 2006). Zu diesem Bereich zählen allgemeine Wertbindungen, so genannte epistemologische Überzeugungen und Weltbilder, welche Einfluss auf die Planung und Gestaltung von Unterricht nehmen, sowie Annahmen über das Lehren und Lernen beinhalten. Bezüglich dieser Überzeugungen lassen sich Lehrkräfte in zwei grundlegende Arten der Vorstellungen vom Unterrichten unterscheiden: Lehrkräfte mit transmissionsorientierten Überzeugungen gestalten ihren Unterricht stärker lehrerzentriert und auf Grundlage von Übungen. Konstruktionsorientierte Überzeugungen sprechen Schülern ein hohes Maß an eigener Kontrolle über ihre Lernprozesse zu, so dass eine konstruktionsorientierte Lehrkraft viel eher eigenständige Lerngelegenheiten herstellen wird (Biedermann, Brühwiler & Krattenmacher, 2012; Blömeke, 2011). Werthaltungen und Überzeugungen sind aber nicht völlig losgelöst vom Professionswissen, denn es konnte festgestellt werden, dass stärker ausgebildetes Fachwissen sowie höhere Kenntnisse in fachdidaktischen Bereichen für das Unterrichtsfach Mathematik häufiger mit konstruktivistischen Überzeugungen einhergeht. Fällt das Fachwissen eher niedrig aus, zeigten die Lehrkräfte vermehrt transmissionsorientierte Vorstellungen zum Mathematikunterricht (Blömeke, 2011). Darüber hinaus lassen sich auch Leistungsunterschiede der Schüler anhand des Ausprägungsgrades konstruktivistischer Orientierungen seitens der Lehrer erklären (Staub & Stern, 2002).

Motivationale Orientierungen sowie selbstregulative Fähigkeiten fokussieren in erster Linie die personenbezogene psychologische Funktionsfähigkeit. Hierunter fallen einerseits Erwartungen der eigenen Selbstwirksamkeit, also dem Vertrauen in die eigenen Fähigkeiten, trotz möglicher Schwierigkeiten erfolgreich handeln zu können, andererseits aber auch Überzeugungen darüber, inwiefern Kontrolle über Situationen und Verhalten ausgeübt werden kann. Motivationale Orientierungen geben Auskunft über die intrinsischen Motive für die Ausübung des Lehrberufs und steuern maßgeblich den Enthusiasmus, der dem Beruf entgegen gebracht wird (Baumert & Kunter, 2006). In diesem Kontext beeinflussen sie außerdem die Bedeutung des eigenen unterrichtlichen Handelns und stehen im Zusammenhang mit den Vorstellungen zur persönlichen und generellen Lehrereffizienz. Die Erwartungen zur eigenen Effizienz hängen nicht nur mit dem generellen Verhalten der Lehrkraft im Unterricht zusammen, sondern bestimmen auch maßgeblich, wie bemüht Zielsetzungen verfolgt und wie hoch diese Ziele überhaupt gesteckt werden. Lehrer, die von ihrer eigenen Effizienz überzeugt sind, zeigen sich offener für neue Ideen bezüglich Methoden oder Herangehensweisen, um Schüler in ihren Lernprozessen optimal zu fordern und fördern (Tschannen-Moran, Woolfolk Hoy & Hoy, 1998). Wirksamkeitserwartungen im Allgemeinen bedingen also maßgeblich, an welchen Maßstäben und Ansprüchen eine Lehrkraft ihren Unterricht ausrichtet und beeinflussen letztendlich auch die Lernleistung und –Entwicklung der Schüler (Lipowsky, 2006). Selbstregulative Fähigkeiten sorgen dem übergeordnet dafür, dass diese motivationalen, physischen und psychi-

schen Ressourcen auf verantwortungsvolle und situationsangemessene Art und Weise eingesetzt werden (Baumert & Kunter, 2006).

Neben diesen beiden Modellen zur Kompetenz von Lehrkräften existieren in der Forschung noch viele weitere Versuche, Kompetenzfacetten zu identifizieren und strukturieren. Die gesamten Forschungsbeiträge im Rahmen der Kontroverse über die Anforderungen an den Lehrberuf können hier nicht umfassend, sondern lediglich vereinzelt und exemplarisch abgebildet werden. Dennoch lassen sich einige Kompetenzelemente erkennen, welche trotz nicht explizit formulierter Modelleinbettung in der Fachliteratur häufig wiederkehren und von verschiedenen Autoren stark betont werden. Grundlegend können sie in die von Frey (2004) vorgenommene Aufteilung der Kompetenzklassen integriert und in der Regel entweder der Fach-, Methoden-, Sozial- oder Personalkompetenz zugeordnet werden. Darüber hinaus zeichnen sich vier grobe Felder ab, in deren Kontext die Anforderungen an die Kompetenzen einer Lehrkraft jeweils spezifiziert werden. Das Handlungsfeld Unterrichtsmethodik und -management umfasst etwa in erster Linie Bereiche der Unterrichtsplanung, des Zeitmanagements und der allgemeinen Lernprozesse. Hierfür nötige Kompetenzen werden in der Forschung häufig als fachliche und soziale Kompetenzen umschrieben. Leistungsmessungen, Förderungen und die Motivierung von Schülern gehören dem Handlungsfeld der Diagnostik und Förderung an und bedürfen sowohl fachlicher als auch methodischer Kompetenz. Geht es um die schulische Umwelt, Öffentlichkeits- und Elternarbeit sowie Beratung werden von der Lehrkraft im Feld der Schulentwicklung und des schulischen Umfelds häufig Sozial- und Personalkompetenzen untersucht. Das Klassenmanagement stellt ein eigenes Handlungsfeld dar. Es umfasst die Lehrer-Schüler-Beziehung, Disziplin, Krisen und Konflikte sowie die Förderung des sozialen Verhaltens. In der Forschungsliteratur liegt das Augenmerk in erster Linie auf methodischen und personalen Kompetenzen (Frey & Jung, 2011).

Trotz der zahlreichen Versuche, die Kompetenzstruktur von Lehrkräften zu erfassen und systematisieren, trifft kaum ein Modell Aussagen darüber, wie die einzelnen Kompetenzen zu gewichtet sind. Üblicherweise liegen sie auf einer Ebene und beeinflussen gleichermaßen die Handlungskompetenz. Dennoch existieren vereinzelt Forschungsbeiträge, die jeweils nur eine Facette der Kompetenz stärker betonen und diese ohne Modelleinbettung systematisch untersuchen. In den folgenden Kapiteln soll beispielhaft auf einzelne, besonders häufig hervorgehobene Kompetenzen eingegangen werden.

2.1.2 Hervorgehobene Kompetenzbereiche ohne explizierte Modelleinbettung

In den letzten 30 Jahren unterliegen die Schule an sich als auch die Anforderungen an die Lehrkräfte einem starken Wandel, der zunehmend hohe Erwartungen an die Rolle der Lehrer mit sich bringt. Die Erziehungs- und Unterrichtsaufgaben eines Lehrers bieten insbesondere im Zuge der Interaktion mit Schülern, Eltern, Kollegen, Schulbehörden sowie der Öffentlichkeit ein erhöhtes Belastungspotential, welchem nur mit effizienten Ressourceneinsatz begegnet werden kann (Sieland, 2001). Durch den Anstieg der Erwartungen sowie der zunehmenden Komplexität der Erziehungsaufgaben sind Lehrer nicht selten vorschneller Kritik ausgeliefert ohne eine nötige Wertschätzung der geleisteten Arbeit. Dieser Umstand erfordert auf Seiten der Lehrkräfte immer mehr individuelle Voraussetzungen für den Umgang und die Beeinflussung wahrgenommener Belastungen (Schaarschmidt & Kieschke, 2007). Fehlen der Lehrkraft die nötigen Ressourcen zur Bewältigung der Belastungen oder ist die Lehrkraft nicht in der Lage, diese Ressourcen angemessen zu organisieren, können schließlich Überlastungen und Stressfaktoren zur emotionalen Erschöpfung und Erkrankungen wie Burnout führen (Bromme & Haag, 2008). Die Berufsgruppe der Lehrer ist mittlerweile am häufigsten betroffen von diesem Phänomen Burnout (Bauer, 2009).

Der krankheitsbedingte Ausfall betrifft nicht nur die erkrankte Lehrkraft, sondern stets auch die Schüler und Kollegen, die mit Unterrichtsausfall und -vertretung konfrontiert werden (Sieland & Tacke, 2000). In diesem Sinne rücken vermehrt belastungsrelevante Ressourcen der Lehrkraft in den Vordergrund der spezifischen Anforderungen an den Lehrberuf und sollten im Rahmen der Erforschung von Lehrerkompetenzen nicht außer Acht gelassen werden. Diese Ressourcen lassen sich in erster Linie als personale Kompetenzen nach Frey (2004) kategorisieren und umfassen Fähigkeiten der Selbststeuerung, des Stressmanagements sowie des Umgangs mit Gefühlen (Sieland & Tacke, 2000). Aber auch Werthaltungen, motivationale Orientierungen sowie selbstregulative Fähigkeiten wird eine hohe Bedeutung im Umgang mit belastenden Anforderungen zugesprochen und können im Modell professioneller Handlungskompetenz nach Baumert und Kunter (2006) verortet werden.

Da Überlastungen oder Burnout im Lehrberuf kein seltenes Phänomen darstellen, wurden die Faktoren für eine erfolgreiche Bewältigung der gestiegenen Anforderungen im Lehrberuf ausführlich untersucht. Es lässt sich feststellen, dass Lehrer mit spezifischen Kompetenzen weniger anfällig für Belastungen sind und selbst bei erhöhten Anforderungen selten ihre eigenen Kapazitäten der Bewältigung überschreiten (Sieland & Tacke, 2000). Immer häufiger ist die Rede von so genannten persönlichen Voraussetzungen, die eine Lehrkraft dazu befähigen, mit den Anforderungen realitätsnah und effizient umzugehen. Hierzu zählen sozial-kommunikative Kompetenzen, ho-

hes Selbstvertrauen sowie eine generelle psychische Widerstandskraft (Schaarschmidt & Kieschke, 2007).

Die Anforderungen an den Lehrberuf haben sich aufgrund der zunehmenden Burnout-Erkrankungen der Lehrkräfte sowie den häufigen Überlastungssymptomen verstärkt hin zu psychischen Bewältigungsstrategien weiterentwickelt, so dass die Eignung für den Lehrberuf nun auch Merkmale der psychischen Stabilität voraussetzt (Schaarschmidt, 2012). Anforderungsanalysen mit Einbezug von Lehrern, Studenten und Lehramtsausbildern ergeben die vordergründige Bedeutung von psychischer Stabilität als spezielle Eignungsvoraussetzung, um dem beruflichen und häufig hektischen Alltag trotz möglicher Misserfolgserlebnisse oder besonderen Perioden starker Belastungen etwa durch Klassenarbeitskorrekturen produktiv und gerüstet entgegen zu können. Hierzu zählen laut Schaarschmidt (2012) die Merkmale der Frustrationstoleranz, Stabilität bei emotionalen Belastungen sowie einer allgemeinen Stressresistenz. Misserfolge müssen zudem offensiv bearbeitet werden, um auch mögliche Enttäuschungen, Ärger oder Niederlagen akzeptieren zu können. Das Aufweisen von spezifischen personalen Kompetenzen dient also nicht nur dem bloßen Gerechtfertigen der tätigkeitsbezogenen Anforderungen und somit einer professionellen und qualitativ hochwertigen Ausübung des Berufs, sondern bedingt auch maßgeblich die berufliche Gesundheit sowie das zufriedene und engagierte Verweilen im Beruf. Kontrollüberzeugungen, selbstregulative Fähigkeiten sowie Werthaltungen dürfen als wichtige Kompetenzfacetten der Lehrkraft nicht in den Hintergrund der Forschungen zum Lehrberuf geraten. Ebenso wichtig sind generelle Einstellungen zum Unterrichten sowie dem beruflichen Ehrgeiz beziehungsweise Engagement. Besonders bei letzterem zeigen die Ergebnisse aus der Belastungsforschung, dass nicht immer ein Maximum der Eigenschaft am erfolgversprechendsten ist (Helmke, 2009). Unterricht und Erziehung können nur dann erfolgreich gelingen, wenn Lehrkräfte ein gewisses Mindestmaß an beruflicher Zufriedenheit trotz möglicher Belastungsfaktoren aufweisen (Bromme & Haag, 2008).

Das Belastungserleben der Lehrkraft beeinflusst nicht zuletzt auch die Ebene der Schülerleistung, denn so weist der Erfolg auf Seiten der Lernenden auch eine Verbindung zu der subjektiven Belastung der Lehrkraft sowie deren Enthusiasmus auf (Bromme & Haag, 2008; Frey, 2008). In einer groß angelegten Onlinebefragung zur Berufszufriedenheit deutsch-schweizer Lehrer konnte Landert (2007) zeigen, dass ebenso Elemente der Arbeitsbelastung einen maßgeblichen Anteil an der Höhe der Berufszufriedenheit bilden. Gleichfalls wirken sich Überzeugungen der eignen Selbstsicherheit im beruflichen Kontext als auch Erfolge in den Erziehungsaufgaben über Werthaltungen und Selbstkompetenzen positiv auf die Berufszufriedenheit aus (Landert, 2007), weshalb sie im weiteren Sinne auch als wichtige Bedingungsfaktoren für den erfolgreichen Umgang mit Belastungsfaktoren genannt wurden. Die Bedeutung der Selbstwirksamkeit sowie Selbstbestimmung ist

im Kontext der allgemeinen Zufriedenheit und Gesundheit nicht zu unterschätzen. Das Vertrauen in die eigenen Kompetenzen beeinflusst die Motivation und Freude der Lehrkraft im alltäglichen Berufsleben und wirkt sich von Seiten des Lehrers ebenfalls auf die Entwicklung von Selbstwirksamkeit und Selbstbestimmung der Schüler aus (Jerusalem et al., 2007).

Eine weitere, in der Forschung häufig thematisierte Kompetenz von Lehrkräften stellt die Sozialkompetenz dar. Diese lässt sich zwar eindeutig in das hierarchische Strukturmodell von Handlungskompetenz nach Frey (2004) eingliedern, wird aber von vielen Autoren in der Fachliteratur als alleinstehende Kompetenz stark hervorgehoben und eine übergeordnete Bedeutung für den Lehrberuf zugesprochen. So sieht Rothland (2010) in der sozialen Kompetenz von Lehrkräften die Kernvoraussetzung für den Tätigkeitsbereich des Unterrichtens sowie der Zusammenarbeit mit Kollegen. Sozial-kommunikative Fähigkeiten erscheinen unabdingbar, schließlich wird das Lernen im Klassenraum primär über Interaktionen und zwischenmenschliche Beziehungen initiiert (Schaarschmidt, 2012). Der Lehrberuf wird demnach auch als „Beziehungsberuf“ bezeichnet in dessen Kontext die Lehrkraft für das Gelingen von Unterricht im eigentlichen Sinne als „Beziehungskünstler“ gesehen wird (Bauer, 2009, S. 254). Die soziale Kompetenz spielt eine entscheidende Rolle in Bezug auf die Anforderungen an die Lehrkraft, als auch als Einflussfaktor für die sozialen Kompetenzen und deren Entwicklung auf Seiten der Schülerschaft. Das Ausmaß an sozialer Verantwortung und Selbstständigkeit geht mit der jeweiligen Ausprägung der Sozialkompetenz der Lehrkraft einher (Frey, 2008). Das Vorherrschen von Kooperationsbereitschaft erweist sich nicht nur als förderlich für die Lernenden, sondern befähigt die Lehrkraft auch dazu, in einem Kollegium zu interagieren (Rothland, 2010). Rückwirkend lässt sich die Kooperation im Kollegium wiederum als schuldifferenzierender Faktor für die Lernleistungen der Schüler feststellen (Frey, 2008).

Als spezielle Fähigkeit beziehungsweise Facette der sozialen Kompetenz ist die kommunikative Kompetenz zu nennen, denn so erweist sich das erfolgreiche Führen von Gesprächen mit Eltern, Schülern oder etwa Kollegen als eine wichtige Aufgabe im lehrberuflichen Alltag (Gartmeier, Bauer, Fischer, Karsten & Prenzel, 2011). Letztendlich kann das Unterrichten an sich als grundlegende kommunikative Tätigkeit beschrieben werden, denn so agiert eine Lehrkraft in Abhängigkeit von Situationen und Adressaten über kommunikative Botschaften (Girmes, 2006). Die kommunikativen Fähigkeiten der Lehrperson üben außerdem inhaltsunabhängig einen nicht zu vernachlässigenden Einfluss auf die Lernleistung von Schülern aus, sofern das Sprachverhalten angenehm und frei von Aggressivität wahrgenommen wird (Frey, 2008). Nicht zuletzt müssen Lehrkräfte im Unterricht stets auf ein wichtiges Instrument ihrer Tätigkeiten zurückgreifen und vertrauen können: ihre eigene Stimme. Eine kräftige Stimme reicht hierbei nicht aus, auch die Ausdrucksfähigkeit ist für

verständliche und klare Kommunikation von Nöten, die auch dazu befähigt, unterschiedliche Adressaten zu erreichen (Schaarschmidt, 2012).

Die Kommunikationskompetenz lässt sich nicht nur bei Frey (2004) finden, sondern zeigt sich im generischen Modell professioneller Handlungskompetenz von Baumert und Kunter (2006) im Beratungswissen als Aspekt des Professionswissens von Lehrkräften. Auch in den Standards für die Lehrerbildung der Kultusministerkonferenz spielen „Kommunikation, Interaktion und Konfliktbewältigung als grundlegende Elemente der Lehr- und Erziehungstätigkeit“ eine wichtige Rolle (KMK, 2004, S. 5). Von Gartmeier et al. (2011) wird die Kommunikationskompetenz besonders im Lehrer-Elterngespräch hervorgehoben. Anforderungsanalysen des Lehrberufs betonen die Wichtigkeit von Stärken im sozial-kommunikativen Kompetenzbereich, die vor allem in Bereichen von prosozialen Einstellungen, sozialer Rücksichtnahme und der Selbstbehauptung von Nöten sind (Schaarschmidt, 2012). Weitere eignungsspezifische Merkmale für den Lehrberuf sind laut Schaarschmidt (2012) das Durchsetzungsvermögen, die soziale Sensibilität, Sicherheit im öffentlichen Auftreten und Freundlichkeit beziehungsweise Warmherzigkeit.

In der Praxis herrscht ebenfalls Einigkeit darüber, dass den sozialen Kompetenzen im lehrberuflichen Alltag eine wichtige Rolle zugeschrieben wird. Lehrkräfte mit und ohne Ausbildungsfunktion nennen in Anlehnung an Frey (2004) alle vier Dimensionen der Fach-, Methoden-, Sozial- und Personalkompetenzen, wobei insbesondere die personale und soziale Kompetenzen am ausführlichsten durch unterschiedliche Fähigkeiten definiert wurden (Weiß, et al., 2014). Ein wichtiger Aspekt der Sozialkompetenz im Unterrichtskontext stellt die Führungskompetenz dar, welche sich in den Fähigkeiten widerspiegelt, Entscheidungen zu treffen und diese auch konsequent geltend zu machen. Dies wird als Ausleben der Rolle des Lehrers beschrieben und bringt unter anderem die Notwendigkeit mit sich, die Beziehung zu den Lernenden im Gleichgewicht zu halten. Innerhalb des sozialen Handlungsfeldes wird außerdem von der Lehrkraft gefordert, authentisch und glaubwürdig zu agieren, um von den Schülern als Interaktionspartner wertgeschätzt zu werden. Im Einklang mit den Befunden von Gartmeier et al. (2011) entnehmen auch Weiß et al. (2014) einer Gruppendiskussionsanalyse von praktisch tätigen Lehrkräften, dass die berufliche Tätigkeit heutzutage verstärkt beratende Aufgaben beinhaltet, folglich Fähigkeiten der Konfliktlösung und Kommunikation unabdingbare Eigenschaften einer Lehrkraft darstellen. Dabei heißt es den Befunden zufolge aber auch, im Konflikt mit Schülern, Eltern oder Kollegen die lehrende Rolle zeitweise abzulegen, um Bedürfnisse aller Gesprächspartner sachlich wahrzunehmen und einen Konflikt als solches auszuhalten.

Die Befunde von Weiß et al. (2014) weisen zudem auf einen Bereich der sozialen Kompetenz hin, der insbesondere für den Umgang mit Schülern gebraucht wird. Hier zielen Anforderungen auf die Orientierung am Kind, das soziale Feingefühl sowie dem Einfühlungsvermögen ab, wie sie

sonst in der Fachliteratur verstärkt unter dem Aspekt der Empathie oder Wertschätzung subsumiert werden (Frey 2004, 2008). Von Lehrkräften im Grundschullehramt wird das Hineinversetzen in die Gefühlswelt der Kinder stark betont, um darauf aufbauend einen bedürfnisorientierten Unterricht aufzubauen. Kindorientierung und Einfühlungsvermögen wirken dabei als Voraussetzungen für die gelingende Umsetzung von Fachwissen und Unterrichtsmethoden. Besonders Gymnasiallehrer begegnen in ihrem schulischen Alltag in der Regel Schülern unterschiedlichsten Alters- und Entwicklungsstufen und bedürfen für die Unterrichtung dieser verschiedenen Interessengruppen eines entsprechenden Ausmaßes an sozialem Feingefühl.

Zwischenmenschliche Fähigkeiten sind im lehrberuflichen Kontext stets in den Interaktionsbereichen mit Schülern, Eltern oder dem Kollegium erforderlich und entscheidend. In den letzten Jahrzehnten hat sich die Beziehungsgestaltung aber besonders zu den Schülern durch zunehmende Verhaltensauffälligkeiten zu einem potentiell belastenden Faktor entwickelt (Sieland, 2001). Viele Lehrer stehen vor der Herausforderung, trotz der widersprüchlichen oder nicht immer zu vereinbarenden Erwartungen in Bezug auf das energische Führen sowie der sensiblen Berücksichtigung der Lernenden eine funktionierende pädagogische Beziehung aufrecht zu erhalten (Bauer, 2009; Schaarschmidt, 2012). Auch soziale Anforderungen stehen im Verdacht, Risiken für Burnout und Überlastungen im Lehrberuf darzustellen, folglich bedingen soziale Kompetenzen nicht nur allgemeinen beruflichen Erfolg sondern beeinflussen auch die psychische Gesundheit der Lehrkraft (Schaarschmidt, 2012). Im Bereich der Interaktion mit Schülern erweisen sich soziale Kompetenzen als wichtige Elemente für den Umgang mit den Forderungen an die Lehrer-Schüler-Beziehung. In der Belastungsforschung lassen sich unter psychisch erschöpften Lehrkräften häufig auch zwei Typen der gestörten Beziehungsgestaltung zu der Schülerschaft herausarbeiten: Einerseits können Lehrkräfte das Beziehungsgleichgewicht in Richtung eines stark führungsorientierten und primär auf den Stoff fokussierten Unterricht verlagern, und so die Notwendigkeit von sozial-kommunikativen Handlungen auf ein Minimum zu reduzieren. Andererseits kann das Gleichgewicht auch auf Seiten der Schüler verschoben sein, so dass die Lehrkraft ausschließlich versucht, ihren Unterricht durch empathische Bemühungen auf Wünsche und Bedürfnisse der Schüler auszurichten, ohne dabei eine klare Führungsrolle einzunehmen (Bauer, 2009). Um diesen gestörten Varianten der Lehrer-Schüler-Beziehungen entgegenzuwirken, sind sozial-kommunikative Kompetenzen auch aus Perspektive der Belastungsforschung eine entscheidende Anforderung an die Lehrkraft.

Wie die vorherigen Ausführungen zeigen konnten, wurden in der Forschung und Fachliteratur insbesondere personale und soziale Kompetenzen der Lehrkraft in den Vordergrund gerückt und besondere Bedeutung zugesprochen. Sie bilden dabei die zwei am stärksten gewichteten Faktoren für eine gelungene Eignung im Sinne der Anforderungen an den Lehrberuf. Darüber hinaus

stößt man in der Literatur auf weitere Kompetenzfacetten und Fähigkeitsstrukturen, die von einzelnen Forschungsgruppen verstärkt untersucht wurden. Darunter zählen besondere Wissens Elemente wie fachdidaktisches oder pädagogisches Wissen, aber auch motivationale Aspekte, die durch Einstellungen und Werte beeinflusst werden und als Gesamtheit in Bezug auf das Lehrerhandeln ebenso berücksichtigt werden sollten wie personale und soziale Kompetenzen (Weiß et al., 2014). Dieses häufig untersuchte Konglomerat von Lehrermerkmalen weist starke Parallelen zu dem Modell professioneller Handlungskompetenz nach Baumert und Kunter (2006) auf, denn dort werden ebenfalls Überzeugungen und Werthaltungen, motivationale Orientierungen und das Professionswissen als entscheidende Komponenten des kompetenten Handelns von Lehrkräften beschrieben.

Das pädagogische Wissen erweist sich als ebenso einflussreich auf die Schülerleistung wie das individuelle Fachwissen einer Lehrkraft (Blömeke, Olsen & Suhl, 2016). Unter pädagogischem Wissen versteht man fachunspezifisches Wissen über Gestaltungsmöglichkeiten des Unterrichts, also die Strukturierung von Unterrichtszeit und -abläufen, sowie das Klassenmanagement und Anwendung von methodischen Kenntnissen zum Lernen und den Sozialformen (Bromme & Haag, 2008). Es ist deutlich mit der Methodenkompetenz gekoppelt und steuert Arbeits- und Lernprozesse der Schüler, um das Fachwissen adressatengerecht zu vermitteln. Das Wissen stellt dabei die nötigen Grundlagen über Lehr- und Lernprozesse dar, um darauf fußend den Unterricht zeitlich in Bezug auf Methoden-, Material- und Medieneinsatz zu planen (Girmes, 2006). Auch aus der Perspektive von Lehrkräften und Lehramtsausbildern in der Praxis wird die Methodenkompetenz im Zuge wichtiger Qualifikationen für den Lehrberuf genannt. Hier ist sie zwar der Fachkompetenz deutlich untergeordnet, bezieht aber ebenso didaktische und methodische Fähigkeiten mit ein, welche erst durch eine Organisationskompetenz erfolgreich geordnet und eingesetzt werden können. Außerdem kann nur durch einen Perspektivwechsel in Richtung der Schüler die Angemessenheit der Methoden beurteilt werden (Weiß et al., 2014). Auch Aussagen zum idealen Lehrerbild von selbst aktiven Lehrern betonen nicht nur die Wissensorientierung, sondern auch eine didaktische Ausrichtung der idealen Lehrertätigkeit. Außerdem wird aus Lehrerperspektive dem idealen Lehrer am häufigsten ein hohes Maß an pädagogischen Kompetenzen zugesprochen, ebenso wie Einfühlungsvermögen, Geduld und eine sorgfältige Unterrichtsvorbereitung (Wilbert & Gerdas, 2007).

Stronge und Hindmann (2003) fassen in ihrem Artikel erwünschte Charakteristika von effektiven Lehrkräften zusammen, um damit bei der Personalauswahl fundierte Variablen zur erfolgreichen Einstellung von Lehrern bereitzustellen. Die Autoren erstellen hierfür sechs spezifische Dimensionen, die bei der Personalauswahl zu berücksichtigen sind: Voraussetzungen für effektives

Lehren, Lehrperson bezogene Merkmale, Klassenmanagement und Organisation, Instruktionsorganisation, Instruktionsdurchführung sowie die Überwachung des Schülerfortschritts und -potentials. Neben dem auch hier betonten lehr- und lernbezogenen Wissen als Voraussetzung für effektives Lehren postulieren die Autoren gleichermaßen organisatorische Kompetenzen auf Seiten der Lehrkraft. Diese werden in erster Linie im Zuge des Klassenmanagements eingesetzt. In anderen Forschungsbeiträgen bisher unerwähnt bleibt die Rolle der Instruktionen im unterrichtlichen Kontext, welche von Stronge und Hindmann (2003) aber durch zwei Bereiche betont werden: Die Instruktionsdurchführung verlangt das Wissen über instruktionstheoretische Strategien sowie deren differenzierten Einsatz bezüglich der Komplexität und den jeweiligen Adressaten. Instruktionen und Aufgaben sollten dabei stets an den Bedürfnissen der Schüler angepasst sein, um Lernende individuell zu fördern. Diese Differenzierung findet sich auch in der Dimension der Überwachung des Schülerfortschritts und -potentials wieder.

Der Erfolg von Lehrern, auf Schüler entsprechend ihrer individuellen Bedürfnisse mittels didaktischen und methodischen Kenntnissen eingehen zu können, hängt aber nicht zuletzt auch von subjektiven Überzeugungen und Motivation ab und moderiert je nach deren Ausprägungen sogar die Leistungen der Lernenden (Blömeke, 2011; Blömeke et al., 2016). Im Sinne von Einstellungen und Überzeugungen betonen auch Anforderungsanalysen, dass die Lehrkraft in Richtung von Mitgestaltungs- und Veränderungsbereitschaft hohe Ausprägungen aufweist und auch zur Verantwortungsübernahme motiviert ist. Außerdem sind für eine Lehrberufseignung auch Begeisterungs- und Motivierungsfähigkeit von Nöten, um Schüler zur Teilhabe am Unterricht motivieren zu können.

Die Grundhaltung zum Beruf sollte auch den Entschluss mit sich bringen, den Lehrberuf typischen Auslagerungen von Korrektur- und Vorbereitungsaufgaben in den privaten Bereich angestrengt und entbehrungsbereit entgegen zu treten (Schaarschmidt, 2012). Auch die Begeisterung für das Fach ist eine nennenswerte Eigenschaft aus Perspektive von Lehrkräften in der Praxis (Weiß et al., 2014). Doch hier bleibt zu erwähnen, dass das reine Interesse an einem Fach nicht ausreichend für die professionelle Ausübung des Lehrberufs ist. Berufswahlmotive sollten stets die Arbeit mit Kindern und Jugendlichen berücksichtigen, sowie die Absicht, am Schulalltag teilzuhaben (Schaarschmidt, 2012). Intrinsische Motivlagen, wie etwa das Interesse an pädagogischer und erzieherischer Arbeit mit Heranwachsenden üben einen Einfluss auf das Erfolgserleben im Studium und Beruf aus (Affolter, Hollenstein & Brühwiler, 2015; Mayr, 2009). Dies zeigen auch Befunde zur Berufszufriedenheit und Berufswiederwahl, welche den unterrichtlichen und erzieherischen Umgang mit Kindern, somit pädagogische Elemente, als Hauptgründe herausstellen. Das Verbleiben im Beruf ist nicht zuletzt davon abhängig, inwiefern das eigene Unterrichten als erfolgreich wahrgenommen wird. Gerade intrinsische Berufswahlmotive bieten hierfür eine günstige Bedingung (Ma-

yr, 2009). Besonders in Ländern, in denen der Lehrberuf mit einer Verbeamtung einhergehen kann, äußern Studierende jedoch häufig auch Aspekte der Arbeitsplatzsicherheit als Gründe für die Berufswahl (Sliwka & Klopsch, 2012). Solche extrinsische Motive können mit grundlegend eher pragmatisch verankerten Berufsüberzeugungen einhergehen, reichen aber in der Regel nicht für eine gewissenhafte Ausübung des Lehrauftrags aus (Affolter et al., 2015). Im Gegensatz dazu können jedoch auch ausgeprägte intrinsische Berufswahlmotive die Gefahr bergen, sehr überzogene und unrealistische Erwartungen an den Lehrberuf zu hegen. Ein solcher Idealismus sollte zur Vermeidung von überraschend eintretenden Überlastungen im realen Berufsalltag stets mit realistischen Vorstellungen vom Berufsbild gekoppelt sein (Schaarschmidt, 2012; Weyand, 2012).

2.2 Kompetenzentwicklung in der Lehrerbildung

Kapitel 2.1 lässt sich entnehmen, dass die Forschung zu Anforderungen an den Lehrberuf einige theoretisch fundierte Modelle zu berufsrelevanten Kompetenzen hervorgebracht hat und somit einen Versuch wagt, Kriterien für die professionelle und kompetente Berufsausübung festzulegen. Kompetenzmodelle sind jedoch nicht in der Lage, qualitative Aussagen über das Ausprägungsminimum der jeweiligen Kompetenzfacetten und Fähigkeiten zu tätigen, ebenso wenig wie sie den postulierten Konglomerat an Kompetenzen Prioritäten zuordnen (Frey, 2008). Das Modell der Handlungskompetenz nach Frey (2004) legt zwar über die Fach-, Methoden-, Sozial- und Personalkompetenz vier Bereiche fest, stellt diese aber als gleichermaßen einflussreich für die generelle Handlungskompetenz dar. Qualitätsansprüche, die durch Kompetenzmodelle nicht artikuliert werden, lassen sich jedoch in Standardmodellen wiederfinden, welche festgelegte Normen und nötige Ausprägungen der geforderten Kompetenzen formulieren (Frey & Jung, 2011). Im Fokus der Standards ist trotz der häufigen Beanstandung der fehlenden Theoriefundierung ganz klar der „Output“ der Lehrerbildung zu verorten (Frey & Jung, 2011, S. 564), wobei bislang generell die Frage nach den Voraussetzungen und Bedingungen für eine zielführende Entwicklung der durch Kompetenzmodelle festgelegten Fähigkeitsbereiche zu den in Standards formulierten Ausprägungsgraden ungeklärt bleibt. Hochschulen stehen also allein vor der Herausforderung, wie sie die professionelle Handlungskompetenz der Lehramtsstudierenden im Rahmen der Lehrerbildung aufbauen und Studierende als adäquat ausgebildete Fachleute in die Berufspraxis entlassen (Messner, 2007).

Nach wie vor herrscht Uneinigkeit darüber, wann und vor allem welche Kompetenzen in der Lehrerbildung entwickelt werden müssen, um anhand dessen ein Qualifizierungsangebot zu konzipieren, das die Studierenden dazu befähigt, eine entsprechend kompetente Handlungsbasis für den Beruf auszubilden (Frey, 2008; Kolbe & Combe, 2008). Da sich die Kompetenz einer Person nicht nur durch Wissen oder theoretische Fähigkeiten definiert, sondern immer erst dann von Kompetenz gesprochen wird, wenn Wissen und Fähigkeiten erfolgreich angewendet werden, ist auch in

der Ausbildung bereits ein praxisorientierter Handlungsbezug unabdingbar. Professionelle Handlungskompetenz entsteht demzufolge neben dem Aufbau von grundlegenden Wissensstrukturen nur durch die ergänzende Beschäftigung mit Aufgaben der beruflichen Praxis (Messner, 2007). Dass dies in der Realität häufig nicht an den Hochschulen der Fall ist, wird durch die klare Bemängelung der Theorie-Praxis-Verzahnung seitens der Studierenden deutlich. Häufig fokussieren Lehramtsstudiengänge vieler Länder verstärkt curriculares, fachliches Wissen und Inhalte der Erziehungswissenschaften und vernachlässigen hierbei häufig kompetenzorientierte Ausbildungsformen (Bromme & Haag, 2008; Girmes, 2006). Der Übertrag von theoretischem Wissen zu praktischen Handeln erscheint vielen Studierenden durch mangelnden Praxisbezug des Studiums kaum leistbar (Kolbe & Combe, 2008). Die Potentiale der Studierenden für die Ausbildung von kompetenten beruflichen Könnens bedürfen im Sinne von Gelegenheiten zur Übung und Erfahrungssammlung ein verstärktes Maß an Unterstützung (Frey, 2008; Messner, 2007).

Grundsätzlich führt ein stärkerer Praxisbezug zwar zu der Sammlung von Erfahrungen, jedoch reicht dies noch nicht aus, um auch ein dementsprechend professionelleres Können auszubilden. Ein wichtiger moderierender Faktor zur Entfaltung von Handlungskompetenz stellt die Reflexion dar (Bromme & Haag, 2008; Messner, 2007). Neben der transparenten Festlegung von Kompetenzklassen durch die jeweiligen Ausbildungsstätten, um die Komplexität der späteren Berufsanforderungen deutlich zu machen, gilt es ebenso, den Studierenden im Laufe des Studiums immer wieder Feedback zu ihrem eigenen Entwicklungsstand zu gewährleisten (Frey, 2008; Girmes, 2006; Sliwka & Klopsch, 2012). Durch solche Kompetenzrückmeldungen sind die Studierenden in der Lage, ihre Fähigkeiten und somit die Erfüllung von Zielsetzungen des Studiums zu reflektieren, so dass gegebenenfalls ihr Vorgehen im Studium angepasst werden kann. Hierzu sollte es im Rahmen der Ausbildung genügend Möglichkeiten geben, erforderliche Kompetenzen aufbauen zu können (Frey, 2008; Girmes, 2006; Sliwka & Klopsch, 2012). Rückmeldungen kommen dabei nicht nur den einzelnen Studierenden zu Gute, sondern treiben auch insgesamt die Qualität der Lehrerbildung voran, was wiederum aus bildungspolitischer Perspektive hoch wünschenswert erscheint (Mayr, 2012; Sliwka & Klopsch, 2012).

Ein weiterer Grund für eine stärker kompetenzorientierte Lehramtsausbildung mit wiederkehrendem und fest verankertem Feedback als Unterstützungssystem der Lehrberufsspezifischen Kompetenzentwicklung, stellt mit Sicherheit die in vielen Ländern offene oder mit nur wenig Beschränkungen behaftete Zulassung zum Lehramtsstudium dar (Mayr, 2012; Rothland & Terhart, 2011). In der Regel basiert die Zulassungsentscheidung für ein Lehramtsstudium auf der Durchschnittsnote des Abiturs, so dass ein Numerus Clausus als festgelegte Zahl die Anzahl der aufgenommenen Bewerber reguliert. Dies geschieht jedoch nicht aus konzeptuellen Erwägungen, son-

dern stellt rein rechnerisch eine Obergrenze für die Anzahl der zugelassenen Studierenden dar (Rothland & Terhart, 2011). Durch diesen Umstand gelangen jedoch auch immer wieder Studierende in das Lehramtsstudium, die über kaum fundierte Berufswahlmotive verfügen und in der Regel ohne vorherige Gedanken oder Absichten diesen Ausbildungsweg einschlagen. Die fehlenden Zulassungsbeschränkungen fördern bei unentschlossenen Studierenden die Annahme, der Lehrberuf sei eine Tätigkeit, die sich ohne besondere Voraussetzungen grundsätzlich als von jedem ausübbar erweise. Häufig entscheiden sich diese Studierenden im Laufe der Ausbildung durch mangelnde Berufsmotivation dazu, vorzeitig aus dem Studium auszuschneiden (Rothland & Terhart, 2011).

In der Praxis wird häufig darauf vertraut, dass Studierende bereits zu Beginn des Studiums in der Lage sind, ihre Eignung und Berufswahlmotive realitätsgetreu und objektiv reflektieren zu können. Dazu bieten manche Hochschulen so genannte Selbsterkundungsverfahren vor Studienbeginn an, so dass Bewerber Anregungen erhalten, ihre Passung zum Lehrberuf selbstständig zu beurteilen und somit eine eigengesteuerte Selektion zu vollziehen (Bieri Buschor & Schuler Braunschweig, 2011). Es ist jedoch mehr als fragwürdig, ob solche Verfahren der Selbstselektion überhaupt eine Wirkung zeigen. In der Regel zeigen angehende Lehramtsstudierende stark optimistische Erwartungen und Einschätzungen zu ihren eigenen Fähigkeiten und begründen ihre Eignung vorrangig mit subjektiven Erfahrungen und Vorstellungen zum Lehrberuf (Bieri Buschor, Schuler Braunschweig & Strinemann Wolf, 2006). Es bedarf also zu Beginn und während des Studiums an Unterstützungsmaßnahmen im Sinne von Feedback und Evaluation zur Kompetenzentwicklung, um nicht nur auf die reflektierte Weiterentwicklung von geeigneten Studierenden abzielen, sondern um ebenfalls Hilfestellungen dazu zu bieten, die berufliche Eignung zu überdenken und gegebenenfalls eine andere, besser passende Studienrichtung einzuschlagen (Mayr, 2012).

In den folgenden Kapiteln soll zunächst beleuchtet werden, inwiefern Studierende des Lehramts bereits über erforderliche Kompetenzen für den Lehrberuf verfügen und in welchem Ausmaß die Notwendigkeit vorliegt, die Ausbildung stärker auf die Kompetenzentwicklung auszurichten. In diesem Kontext wird ebenso hervorgehoben, welche Kompetenzen bereits im Studium gefördert werden sollten. Zum Schluss soll auf bislang eingesetzte Verfahren zur Erfassung von berufseignungsrelevanten Kompetenzen vor und während des Studiums eingegangen werden.

2.2.1 Kompetenzen von Lehramtsstudierenden

In Anbetracht der Vielzahl an eignungsrelevanten Fähigkeiten und Kompetenzen, welche sich aus den Kompetenz- und Standardmodellen ableiten lassen, bleibt die Frage offen, wann sich diese in der beruflichen Laufbahn hin zu den erstrebenswerten Ausprägungen entwickeln (Baumert & Kunter, 2006; Rothland, 2010). Es ist keinesfalls zu erwarten, dass Lehramtsstudierende bereits zu Be-

ginn ihres Studiums mit den nötigen Kompetenzen vollständig ausgestattet sind (Schaarschmidt, 2012). Vielmehr lassen sich bei Studierenden zunächst mehr oder weniger günstige Voraussetzungen für die Eignung zum Lehrberuf feststellen, anhand derer Entwicklungspotentiale festgesteckt werden können (Rothland, 2010). Dieses Veränderungsprozesse bezüglich der generellen Handlungskompetenz sowie die mögliche Erlernbarkeit von Kompetenzen machen deutlich, dass die berufsbezogene Eignung kein starres oder unveränderbares Konstrukt darstellt (Baumert & Kunter, 2006; Rothland, 2010; Schaarschmidt, 2012). Da es bislang kaum Kompetenzmodelle gibt, welche Aussagen über die Entwicklung der Kompetenzen tätigen, forciert die Lehrerforschung immer stärker den Bereich der Lehrerbildung, denn so stellt diese unbestritten die Ausgangsbasis für das Gelingen einer professionellen Berufsausübung dar (Schratz, 2012). Studien zu den Kompetenzen und deren Einschätzung von Lehramtsstudierenden sollen Aufschluss darüber geben, inwiefern Studierende über Entwicklungspotentiale zu den geforderten Kompetenzen verfügen und ob die Lehrerbildung im Allgemeinen der Notwendigkeit unterliegt, diese im Laufe der Ausbildung zu fördern. Hierbei setzen die Autoren je nach Zugehörigkeit einer Forschungsrichtung unterschiedliches Augenmerk auf die einzelnen Kompetenzfacetten. Im Folgenden soll auf ausgewählte Studien eingegangen und deren Schlussfolgerungen für die Lehrerbildung dargestellt werden.

Im Rahmen des von Frey (2004) aufgestellten hierarchischen Strukturmodells von Handlungskompetenz und der Einteilung in die vier Kompetenzklassen der Fach-, Methoden- Sozial- und Personalkompetenz wurde im Rahmen einer internationalen Studie in den Ländern Deutschland, Schweiz, Österreich, Polen und Italien über Selbstbeurteilungsbögen erforscht, wie sich die dortigen Lehramtsstudierenden in ihren Fähigkeiten einschätzen und ob sich diese zu den vier Kompetenzklassen bündeln lassen. Außerdem wurde der Einfluss von verschiedenen Faktoren wie Alter, Geschlecht, Fachsemester, Ausbildungszufriedenheit oder Berufsperspektiven auf die Kompetenzen geprüft. Die Stichprobe bestand aus 1841 Lehramtsstudierenden, wobei die Mehrzahl aus Deutschland stammte. Die Untersuchung wurde später auf 2150 Studierende sowie Referendare erweitert und den gleichen Untersuchungsfragen unterzogen (Frey, 2008). Insgesamt konnte gezeigt werden, dass sich die Fähigkeiten der Lehramtsstudierenden der verschiedenen Länder allesamt in Fach-, Methoden-, Sozial- und Personalkompetenzen aufteilbar sind. Die höchsten Ausprägungen der Kompetenzklassen wiesen die österreichischen Studierenden auf, die niedrigsten Kompetenzwerte zeigte die polnische Stichprobe. Deutsche Lehramtsstudierende lassen sich im mittleren Bereich der Kompetenzausprägungen verorten. Die Studierenden berichteten außerdem darüber, dass sich insbesondere solche Fähigkeiten leichter in der Praxis anwenden ließen, welche schon in der Hochschule intensiv vermittelt und eingesetzt werden konnten. Schwieriger fällt den Studierenden hingegen der Einsatz von weniger in der Ausbildung fokussierten Kompetenzbereichen, welche Elemente wie die Selbstorganisation, Konflikt-, Führungs- und Analysefähigkeit sowie der Reflektion

beinhalten. Eine signifikante Verbindung der Fähigkeitsausprägungen zu möglichen externen Einflussvariablen ließ sich insgesamt nicht feststellen, so dass das Heranziehen von personenspezifischen Kriterien, Ausbildungsfortschritt und -zufriedenheit nicht ausreichend ist, um die berufsrelevanten Kompetenzen von Lehramtsstudierenden vorherzusagen.

In der Studie von Rothland (2010) wurde ausschließlich die Sozialkompetenz von Lehramtsstudierenden untersucht, da diese laut Aussagen des Autors aufgrund der lehrberufsspezifischen vielfältigen Interaktionsgegebenheiten mit Schülern, Eltern und Kollegen eine hervorzuhebende Anforderung an den Lehrberuf darstellt (eine ausführliche Darstellung der Bedeutung von sozialer Kompetenz findet sich in Kapitel 2.1). Die Ausprägungen der sozialen Kompetenz von Lehramtsstudierenden wurden hierbei nicht nur isoliert empirisch untersucht, sondern auch mit den Ausprägungen von Studierenden der Humanmedizin und Rechtswissenschaften verglichen, denn es gelten für Ärzte und Juristen ebenfalls hohe Anforderungen an soziale Fähigkeiten für die gelingende Ausübung des Berufs. In diesem Rahmen sollte ebenfalls geklärt werden, in welchem Ausmaß die universitäre Ausbildung zur Entwicklung der Sozialkompetenz beiträgt und inwiefern ein Entwicklungsbedarf zu verorten ist. Die Stichprobe setzt sich aus 977 Lehramtsstudierenden unterschiedlichen Ausbildungsgrades und aus verschiedenen Bundesländern, 210 Studierenden der Humanmedizin sowie 135 Studierende der Rechtswissenschaften zusammen. Über Fragebögen wurden die Studierenden gebeten, verschiedenste Auskünfte zu Bereichen wie unter anderem der Kooperationsfähigkeit, der sozialen Verantwortung, der Kommunikations- und Konfliktfähigkeit sowie des situationsgerechten Auftretens zu geben. Es konnte gezeigt werden, dass Lehramtsstudierende ihre sozial-kommunikativen Kompetenzen durchweg als stark ausgeprägt beurteilen und sich ihre Selbsteinschätzungen nicht signifikant von denen der Medizin- und Rechtswissenschaftsstudierenden unterscheiden. Lehramtsstudierende empfinden einen durch das Studium hervorgerufenen Kompetenzzuwachs im Bereich sozialer Fähigkeiten, obgleich diese Steigerung nicht sonderlich herausragend ist.

Die beiden Studien von Frey (2004, 2008) und Rothland (2010) zeigen zusammenfassend, dass die Entwicklung von fachlicher, methodischer, sozialer und personaler Kompetenz bereits im Lehramtsstudium zu verorten ist und in ihrer Entwicklungen grundsätzlich von der Ausbildung tangiert werden können. Aus internationaler Perspektive weisen die Kompetenzausprägungen deutscher Lehramtsstudierende auf die Dringlichkeit hin, die Ausbildung grundsätzlich bezogen auf alle vier Kompetenzdimensionen sowie vermehrt praxisbezogene Fähigkeiten stärker zu fördern. Hinweise für Verbesserungsansätze der Lehramtsausbildung lassen sich in österreichischen und schweizerischen Ausbildungssystemen wiederfinden, denn hier wird den Studierenden vermehrt die Möglichkeit gegeben, ihr theoretisches Wissen in der Praxis zu erproben und somit ihre eigene

Handlungskompetenz ausbilden und reflektieren zu können (Frey 2004, 2008). Die Studie von Rothland (2010) ergänzt dieses generelle Verbesserungspotential der Lehrerbildung durch die Hervorhebung speziell zu fördernder Fähigkeitsaspekte. Zwar fielen die Kompetenzurteile zu sozialen Fähigkeiten grundsätzlich zufriedenstellend aus, doch so konnte insbesondere im männlichen Teil der Stichprobe ein vermehrter Entwicklungsbedarf der Sozialkompetenz festgestellt werden. In Bereichen der Konfliktfähigkeit, der Empfindlichkeit bei sozialer Frustration, der Selbstbehauptung und Konfrontationstendenz zeigten sich überdies geschlechtsunspezifische Entwicklungspotentiale auf. Zwar wiesen Lehramtsstudierende der Ausbildung eine Rolle in Bezug auf ihre Kompetenzentwicklung im Laufe des Studiums zu, doch so geschieht dies eher durch studiengangunspezifische und generell an Hochschulen geltenden Arbeitsformen. Es zeigen sich hier also Ansatzpunkte für spezifische Fördermaßnahmen zur Kompetenzentwicklung und -reflektion innerhalb der Ausbildung.

Ein überwiegender Teil der Studien zu berufsrelevanten Kompetenzen von Lehramtsstudierenden nutzen als Methode Selbsteinschätzungsinstrumente und verlassen sich somit darauf, dass Studierende in der Lage sind, ihre eigenen Fähigkeiten realistisch einzuschätzen. Ob eine hohe Einschätzung in einem Kompetenzbereich auch damit einhergeht, dass die entsprechende Person hier auch tatsächlich kompetent handeln kann, bleibt in der Regel offen. Rauin und Meier (2007) befragten 900 Lehramtsstudierende dreier Hochschulen in Baden-Württemberg über einen Zeitraum von acht Jahren zu Berufswahlmotiven, Studienbedingungen und Kompetenzerwerb, um schlussfolgern zu können, ob die Ausbildung durch regelmäßige Kompetenzrückmeldungen dazu beigetragen hat, eine realistische Einschätzung der Kompetenzentwicklung zu ermöglichen. Anhand der Selbstauskünfte der Studierenden ließ sich kein einheitliches Bild der Kompetenzen erstellen, da eine objektive Messgrundlage fehlte. Die Urteile zeugten von selbstwertdienlichen Tendenzen und ließen den Schluss nahe, dass Selbsteinschätzungen weniger den qualitativen Ausprägungsgrad einer Kompetenz wiedergeben, sondern widerspiegeln, welche Inhalte häufig Gegenstand der Beschäftigung waren. Die Konfrontation mit der Realität erfolgt schlussendlich erst durch den Wechsel ins Referendariat, wobei viele Referendare erst durch die Auseinandersetzung mit verschiedenen Belastungen ihre fehlende Vorbereitung für den tatsächlichen Umgang mit einer Klasse erkennen (Bauer, 2009).

Es wird deutlich, dass es auch in der Verantwortung der Hochschule und der Ausbildung liegt, den Studierenden transparent zu kommunizieren, auf welchem Stand der Kompetenzentwicklung sie sich befinden und welche Mindestausprägung erforderlich ist, um in der praktischen Lehrsituation erfolgreich handeln zu können. Studierende sollten generell durch ein realistisches Kompetenzfeedback davor geschützt werden, sich gegenwärtig überfordernden Situationen und somit starken Belastungen auszusetzen (Girmes, 2006). Auch unter Berücksichtigung der Belastungsfor-

sung und der zunehmenden Erkrankung an Burnout oder Symptomen der psychischen Überlastung ist es von großer Bedeutung, den Berufseinstieg nicht sogleich mit Überforderungen und Versagensängsten beginnen zu lassen (Schaarschmidt & Kieschke, 2007). Wichtige Handlungskompetenzen, auf die ein Lehramtsstudium zur Belastungsprävention vorbereiten sollte, gehören in erster Linie dem sozial-kommunikativen Bereich an und umfassen Fähigkeiten wie Sensibilität, Rücksichtnahme, Ausgeglichenheit, Belastbarkeit, aber auch Durchsetzungskraft, Selbstbehauptung und eine grundlegende Kontaktfähigkeit (Abujatum, Arold, Knispel, Rudolf & Schaarschmidt, 2007; Schaarschmidt & Kieschke, 2007; Sieland & Tacke, 2000). Wichtige Ansätze der Belastungsprävention, welche im Studium bereits etabliert sein sollen, umfassen zum einen generelle Präventionsmaßnahmen, die sich auf eine Verbesserung der Ausbildung von berufsrelevanter Kompetenzen stützen, und zum anderen spezifische Präventionsmaßnahmen, die gezielte Konflikt-, Kommunikations- und Stressbewältigungstrainings in die Lehrerbildung integrieren (Sieland, 2001).

Von Studierenden, die einen Lehrberuf anstreben, kann in keiner Weise erwartet werden, dass diese bereits alle berufsrelevanten Merkmale und Kompetenzen mit ins Studium bringen. Viel eher lässt sich bei den individuellen Voraussetzungen von Lehramtsstudierenden eine große Heterogenität feststellen, welche mit Blick auf die für den Beruf erforderlichen Kompetenzen im Zuge der Ausbildung zielgeleitet und kompetenzorientiert geschult und gefördert werden müssen (Abujatum et al., 2007; Sieland, 2001). Bei den verschiedenen Ausgangsqualifikationen ist es ratsam, diese nicht als defizitäre Strukturen zu sehen, sondern viel eher in der Lehrerbildung dafür zu sorgen, dass die Voraussetzungen der Studierenden mit den Anforderungen des Lehrberufs in Einklang gebracht werden (Oser, 2006). Dies bedeutet jedoch, dass Hochschulen ihre Ausbildungsziele transparent darstellen müssen, um bereits zu Studienbeginn eine Reflexion über die Eignung im Zusammenhang mit den Anforderungen des Lehrberufs zu unterstützen und damit auch Gelegenheiten schaffen, Berufsalternativen in Erwägung zu ziehen (Abujatum et al., 2007; Mayr, 2012; Schaarschmidt, et al., 2000; Sieland, 2001). Darüber hinaus ist es von Nöten, den Studierenden im Laufe ihrer Ausbildung durch Potentialanalysen immer wieder die eigenen Leistungsfähigkeiten und Entwicklungsstände zurückzumelden, so dass sie ihren Entwicklungsbedarf erkennen und ihre Ausbildung in Richtung der benötigten Kompetenzen vorantreiben können (Mayr, 2012; Sieland, 2001). Grundlegend besteht der Vorschlag, die Ausbildung weg von einer reinen Wissensvermittlung hin zu einer kompetenzorientierten Lehrerbildung umzugestalten (Oser, 2006; Schratz, 2012). Jedoch besteht noch lange nicht die nötige Offenheit gegenüber einer in das Studium eingebetteten Kompetenzdiagnostik und -rückmeldung (Frey, 2004).

2.2.2 Diagnostik und Förderung von Kompetenzen in der Lehrerbildung

Wie bereits in den vorangegangenen Kapiteln betont, stellt die Lehrerbildung einen entscheidenden Abschnitt in der Entwicklung von berufsrelevanten Kompetenzen dar. Zu Beginn des Studiums bringen Lehramtsstudierende individuelle Qualifikationen und heterogene Voraussetzungen mit, so dass unterschiedlichste Potentiale in die Ausbildung einfließen. Die Lehrerbildung sollte im Rahmen einer kompetenzorientierten Ausbildung als Entwicklungs- und Entfaltungsraum fungieren und den Studierenden Unterstützungen bieten, um die reflektierte Weiterbildung voranzutreiben (Mayr, 2012; Terhart, 2007). Um die Qualität der Ausbildung zu steigern, werden immer mehr Forderungen nach regelmäßiger Kompetenzdiagnostik, Feedback und Entwicklungsförderung geäußert. Viele Hochschulen sind aktuell mit der Frage nach der richtigen Methode zur Erfassung von Kompetenzen konfrontiert, um darauf aufbauend den Kompetenzentwicklungsprozess der Studierenden im Lehramt anzuregen und zu unterstützen (Frey, 2006).

Es gestaltet sich als Herausforderung, geeignete Methoden und Instrumente zur Kompetenzdiagnostik zu konzipieren, die auf standardisierte Weise verlässliche Informationen über Kompetenzen liefern und auch in der Lage sind, zwischen individuellen Ausprägungsgraden genau differenzieren zu können (Lehberger & Lüth, 2012; Terhart, 2007). Die Beurteilung von Kompetenzen gestaltet sich aufgrund des Umfangs von empirisch hervorgehobenen Kompetenzfacetten als kein klar definiertes Konzept, denn so konkurrieren verschiedenste theoretische Ansätze zu eignungsrelevanten Fähigkeiten von Lehrern miteinander und schaffen kein einheitliches Bild zwingend erforderlicher Kompetenzen. Aufgrund der vielfältigen und oft nicht eindeutigen Anforderungsstruktur des Lehrberufs wird häufig argumentiert, dass berufsrelevante Leistung nicht verzerrungsfrei erfasst werden kann. Außerdem muss aus der Vielzahl der komplexen Anforderungsbereichen eine reduzierte Auswahl getroffen werden, um die Kompetenzdiagnostik überhaupt aus Gründen der Durchführbarkeit und des Aufwandes realisieren zu können. Darüber hinaus gibt es noch keine genügend empirisch gestützten Befunde zu der Entwicklung von Kompetenzen, als dass klare Ursachen und Wirkgefüge für ein gelingendes Erreichen der nötigen Fähigkeiten aufgestellt werden können. Unterschiede in den Ausprägungen und individuellen Voraussetzungen können auf multiple Ursachen zurückgeführt werden, ob diese im Rahmen der Ausbildung überhaupt beeinflussbar sind, bleibt noch unklar (Terhart, 2007). Aus methodischer Sicht müssen Verfahren zur Kompetenzdiagnostik außerdem den testtheoretischen Qualitätsindikatoren genügen, also entsprechend reliable, valide und objektiv wie gängige Leistungsbeurteilen ausfallen (Frey, 2006; Schuler, 2004).

Einige Ausbildungsinstitutionen haben bereits Projekte ins Leben gerufen, um Verfahren zur Kompetenzdiagnostik, Beratung und der Potentialentwicklung zu erstellen und zu erproben (Frey, 2006; Nieskens, 2012). Diese Methoden lassen sich in Kategorien einteilen, je nachdem welches Ziel

die entwickelten Instrumentarien verfolgen. Dadurch, dass die Zulassung für das Lehramtsstudium in der Regel nicht beschränkt ist, hat sich an vielen Hochschulen der Trend etabliert, so genannte Selbsterkundungsverfahren vor Studienbeginn anzubieten. Häufig werden diese als Online-Version angeboten und ermöglichen es Studieninteressierten, einen Einblick in die berufsspezifischen Charakteristika und Anforderungen zu erhalten (Nieskens, 2012). Die Bewerber sollen im Sinne eines Self Assessments anhand standardisierter Fragen zu ihren Kompetenzen und Fähigkeiten Einschätzungen treffen, die auf einer festgelegten Antwortskala zu lokalisieren sind (König, Kaiser & Felbrich, 2012).

Diese Selbsterkundungsverfahren sollen die Bewerber dabei unterstützen, ihre Passung zu den Anforderungen des Lehrberufs selbstständig zu beurteilen und verlagern somit die Zuständigkeit für die Selektion auf die Seite der potentiellen Bewerber (Bieri Buschor & Schuler Braunschweig, 2011). Es bleibt jedoch fraglich, inwiefern Personen überhaupt in der Lage sind, ihre eigenen Fähigkeiten und Leistungen retrospektiv zuverlässig zu beurteilen, ohne dass hierbei Verzerrungen durch möglicherweise unbewusste Tendenzen zu sozialer Erwünschtheit die Ergebnisse in ihrer Aussagekraft einschränken (Frey, 2006; Terhart, 2007). Zwar sind Selbsterkundungen prinzipiell mit wenig Aufwand verbunden, doch bleibt auch auf inhaltlicher Ebene unklar, ob die Komplexität der lehrberufsrelevanten Kompetenzfacetten ganzheitlich über sehr begrenzte Selbsterkundungsverfahren erfasst werden können (Oser et al., 2007). Aus diesem Grund koppeln einige Hochschulen die Anwendung von Self Assessments mit einem verpflichtenden Beratungsangebot (Nieskens, 2012).

Der generellen Problematik der subjektiven Verzerrungen von Selbsterkundung und Selbstselektion entgegnen manche Hochschulen mit dem Einsatz von Aufnahmeverfahren, die der Fremdselektion unterliegen (Nieskens, 2012). Aufnahmeverfahren sind dem Studium vorangestellt und messen die aus Anforderungsanalysen gewonnen berufsspezifischen Kompetenzen bei Bewerbern des Lehramts, um deren Eignung diagnostisch objektiv beurteilen zu können (Bieri Buschor et al., 2006). Häufig prüfen Aufnahmeverfahren fachliche, soziale und personale Kompetenzen über direkte Beobachtungen, denn nur so kann ein zulässiger Schluss von in der Situation gezeigten Leistungen zur Ausprägungen der zu untersuchenden Kompetenzen erfolgen (Bieri Buschor et al., 2006; Oser et al., 2007). Viele Hochschulen greifen bei Aufnahmeverfahren auf Konzepte des klassischen Assessment Centers zurück, damit die Messungen anhand von vielfältiger Methoden und dem Einsatz mehrerer geschulter Beurteiler den psychometrischen Gütekriterien Reliabilität, Validität und Objektivität genügen (Bieri Buschor et al., 2006; Terhart, 2007). An der Pädagogischen Hochschule Zürich etwa wird das Assessment Center als verbindliches Aufnahmeverfahren für Studieninteressierte ohne Abitur durchgeführt. Zwar birgt ein solches Verfahren stets auch hohe personale und

zeitliche Kosten, doch so besitzt es eine überzeugende prognostische Validität in Bezug auf den Studienerfolg und wird von Studierenden gemeinhin akzeptiert (Bieri Buschor et al., 2006; Nieskens, 2012; Oser et al., 2007).

Doch neben der Selbst- und Fremdselektion haben einige Hochschulen Verfahren der Diagnostik von lehrberufsrelevanten Kompetenzen prioritär dafür konzipiert, den Studierenden während ihrer Ausbildungslaufbahn hochwertige Impulse zur Kompetenzweiterentwicklung zu bieten. Häufig gehen diese Verfahren auch mit beratenden Angeboten für die Berufslaufbahn einher, sollen jedoch viel eher auf die Kompetenzen der Studierenden abzielen, um eine handlungsbefähigende Passung zwischen Fähigkeitsvoraussetzungen der Studierenden und beruflichen Anforderungen zu unterstützen (Nieskens, 2012). Eignungsdiagnostische Verfahren der Leistungsbeurteilung müssen also nicht zwangsläufig mit einer Entscheidung für die allgemeine Zulassung zum Lehramtsstudium gekoppelt sein, sondern können auch als systematische Feststellung des Entwicklungspotentiales genutzt werden. Darauf aufbauend kann den Studierenden schließlich im Zuge ihrer Ausbildung ein individuelles Feedback erteilt werden und ein Angebot geschaffen werden, um die Kompetenzentwicklung zielgerichtet zu fördern (Bieri Buschor et al., 2006; Päßler, Hell & Schuler, 2011).

Aktuell existieren bereits vielzählige Verfahren zur Erfassung von Kompetenzen, die unterschiedliche Zielrichtungen verfolgen (Selbsterkundung, Fremdselektion, Kompetenzfeedback und -entwicklung) und sich dabei verschiedener Methoden bedienen (Nieskens, 2012). Die gängigsten Formate verwenden Fragebögen zur Selbstbeurteilung, darüber hinaus existieren aber auch Gruppenbeurteilungsverfahren, Assessment Center, Unterrichtsproben, Entwicklungsportfolio oder Bildertests. Betrachtet man Diagnoseinstrumente, welche zwischen 1991 und 2005 veröffentlicht wurden, zeigt sich, dass 79 % (37 von 47) der Instrumente aus Fragebogenformaten bestehen und markieren einen deutlichen Trend für diese Methodik. Nur vier Instrumente nutzen eine Interviewstruktur und nur drei lassen sich dem Assessment Center zuordnen (Frey, 2006). Die Trierer Tagung „Auf den Lehrer kommt es an – Geeignete Lehrer/innen gewinnen, (aus-)bilden und fördern“ (2009) identifizierte insgesamt sechzehn empirisch erprobte Verfahren aus dem deutschsprachigen Raum, welche entweder zu Zwecken der Selektion, Beratung oder Förderung entwickelt wurden und an pädagogischen Hochschulen Verwendung finden. Im Jahre 2011 wurde diese Sammlung überarbeitet und in einem Überblick veröffentlicht (Nieskens, 2012). Tabelle 1 zeigt die sechzehn Verfahren und bestimmt ergänzend die verwendeten Methoden und Instrumente der jeweiligen Verfahren.

Tabelle 1

Sechzehn Verfahren zur Lehrerrekutierung, Lehrereignung und Studierendenauswahl nach Nieskens (2012), ergänzt durch Angaben zur Methodik

Name des Verfahrens	Methodik	Verantwortliche
ABC-Lehramt Assessment, Beratung und Coaching zur Klärung der Berufseignung von Lehramtsstudierenden	Assessment, Beratung, Coaching	Dipl. Päd. Birgit Weyand, Prof. Dr. Helga Schnabel-Schüle, Universität Trier Zentrum für Lehrerbildung
CCT – Career Counselling for Teachers	Fragebogen, Kompetenztest	Prof. Dr. Johannes Mayr, Dr. Birgit Nieskens, Prof. Dr. Eveline Gutzwiller-Helfenfinger, Universität Klagenfurt, Leuphana Universität Lüneburg, Pädagogische Hochschule Zentralschweiz, Luzern
Eignungsabklärung im Rahmen des Grundjahrs an der PHZ Luzern	Hospitation, Gespräche	Marco Wyss, Prof. Dr. Michael Fuchs, Pädagogische Hochschule Zentralschweiz Luzern
EignungsFeedback	Selbst-/Fremdbeobachtung, Beratung, Portfolio	Vanessa Rumpold, Nicole Valdorf, Dr. Lilian Streblow, Mandy Geisler, Anke Schöning, Universität Bielefeld
FIBEL – Self-Assessment der Universität Münster	Self Assessment, Selbstbeobachtung	Prof. Dr. Wolfgang Böttcher, Dr. Detlef Berntzen, Prof. Dr. Uwe P. Kanning (Hochschule Osnabrück), Westfälische Wilhelms-Universität Münster
Förderung von Lehrerkompetenzen der Lernkultur in sprachlich, kulturellen und leistungsmäßig heterogenen Klassen	Workshop	Ute Kienzl, Silvia Bär-Montag, GPV Gestaltpädagogische Vereinigung e. V.
Ganzheitliche Stressprävention	Training	Dr. Julia Kosinár, Universität Bremen
Lehrereignung frühzeitig erkennen und fördern	Selbst-/Fremdbeobachtung, Selbsterkundung, Training	Dr. Elmar Lüth, Prof. Dr. Reiner Lehberger, Prof. Dr. Uwe Schaarschmidt, Dr. Andreas W. Fischer, ZEIT-Stiftung, Universität Hamburg, Zentrum für Lehrerbildung, COPING
PARcours – Eignungsdiagnostisches Verfahren für angehende Lehramtsstudierende	Multimodales Assessment Center	Prof. Dr. Norbert Seibert, Renate Wirth, Universität Passau
Professions- und kompetenzorientierte, personalisierte LehrerInnenausbildung (Sek I+II)	Ausbildungsintegrierte Portfolioarbeit	Dr. Christian Kraller, Universität Innsbruck
Psychosoziale Basiskompetenzen für den Lehrerberuf	Blockseminar, Fremdbeobachtung, Reflexion	Prof. Dr. Heinrich Dauber, Dr. Elke Döring-Seipel, Prof. Dr. Dorit Bosse, Timo Nolle, Universität Kassel
Quer- und Seiteneinsteiger ins Lehramt Physik	Fragebögen	Dr. Friederike Korneck, Jan Lamprecht, Lars Oettinghaus, Goethe-Universität Frankfurt am Main
Studienberatungstests für Lehramtsstudierende (LBT)	Tests, Fragebögen	Dr. Gabriele Hörl, Univ.-Prof. Dr. Ferdinand Eder, Univ.-Prof. Mag. Dr. Franz Hofmann, Universität Salzburg
Teach First Deutschland	Training	Kirsten Altenhoff, Kommunikation, Teach First Deutschland gemeinnützige GmbH, Berlin
Tätigkeits-, Stimmungs- und Zeitanalyse. Trainings zum Arbeits- und Gesundheitsverhalten von Studierenden, Lehrkräften und Schülern	Fragebögen, Zeittagebuch, Training	Prof. Dr. Bernhard Sieland, Leuphana Universität Lüneburg
TEK-LS – Training emotionaler Kompetenzen für Lehrkräfte und Schüler	Fragebögen, Training	Prof. Dr. Bernhard Sieland, Leuphana Universität Lüneburg

Auch hier zeigt sich der häufige Einsatz von Fragebögen, jedoch werden ebenfalls Elemente der Verhaltensbeobachtung in die einzelnen Verfahren integriert. Es wird deutlich, dass Hochschulen, die eine Kompetenzdiagnostik im Zuge von entwicklungsunterstützenden Feedback- und Reflexionsangeboten in die Ausbildung integrieren möchten, sich auch mit methodischen Fragestellungen auseinandersetzen müssen. Es bleibt ihnen offen, ob sie sich bereits veröffentlichte Methoden und Instrumente zu Nutze machen, oder gänzlich neue Verfahren konzipieren (Frey, 2006). Aufgrund der Komplexität und der Verhaltensbezogenheit von lehrberuflichen Kompetenzen liegt es jedoch nahe, direkte Beobachtungen von Performanz heranzuziehen und eignungsdiagnostische Verfahren der Leistungsbeurteilung an den hohen Ansprüchen der Testgütekriterien auszurichten, um zuverlässige und objektive Aussagen über Kompetenzausprägungen tätigen zu können (Bieri Buschor et al., 2006; Oser et al., 2007). In den folgenden Kapiteln sollen allgemeine eignungsdiagnostische Verfahren daher noch einmal detailliert dargestellt werden und auf ihre Stärken und Schwächen hin untersucht werden. Dabei fokussiert sich die vorliegende Arbeit primär auf verhaltensbezogene und simulationsorientierte Verfahren, da diese bis dato nur sehr selten als mögliche Kompetenzdiagnostikverfahren für das Lehramtsstudium in Betracht gezogen wurden.

2.3 Verfahren zur Eignungsdiagnostik und Leistungsbeurteilung

Verfahren zur Eignungsdiagnostik werden in erster Linie dazu genutzt, Leistungen von Personen zu beurteilen und damit ihre Eignung für einen spezifischen Beruf festzustellen. Doch neben diesem Einsatz im klassischen Bereich der Eignungsdiagnostik existieren noch weitere Funktionen der Beurteilung von Leistungen: Einer Person kann ihr aktueller Leistungsstand zurückgemeldet und somit Ansatzpunkte und Feedback für die eigene Leistungssteuerung geboten werden. Informationen über das derzeitige Leistungsniveau, die Prognose zukünftiger Leistungsentwicklung sowie eine mögliche Beurteilung von Potenzialen können eine Ausgangslage für Beratungen und Entwicklungsförderungen bieten. Nicht zuletzt bieten Leistungsmessungen auch Transparenz darüber, welche Anforderungen in der beruflichen Tätigkeit tatsächlich herrschen. (Schuler, 2004; Stiggins, Arter, Chappuis & Chappuis, 2004). Besonders letztere Funktionen der Eignungsdiagnostik und Leistungsbeurteilung können von Interesse in der Lehrerbildung sein, um den Studierenden ein realistisches Bild über ihre bislang entwickelten Kompetenzen zu verschaffen und durch Feedback die Reflexion des eigenen Könnens und somit die Kompetenzentwicklung zu unterstützen.

Die nächsten Kapitel sollen besonders solche Verfahren der Eignungsdiagnostik erläutern, welche weniger darauf abzielen, festgelegte und spezifische Wissensziele abzufragen. Für die Kompetenzerfassung im Lehramtsstudium legt diese Arbeit das Augenmerk ausschließlich auf simulationsorientierte Methoden, welche die Performanz einer Person beobachtbar machen und diese als Zugang zu den Kompetenzen einer Person nutzen. Dazu erfolgt zuerst eine Einführung in die Grund-

lagen der Berufseignungsdiagnostik, um im Anschluss das Verfahren des Performance Assessment als eine spezielle verhaltensbezogene Variante des Assessment Centers näher zu beleuchten. Da es innerhalb von Assessment Centern gängig ist, die fremdbeurteilten Leistungsbeobachtungen mit Selbsteinschätzungen zu kombinieren und Selbstbeurteilungen von Kompetenzen in der Lehrerbildung häufig bevorzugt werden, soll dieser ergänzende Zugang anschließend kurz in seinen Vor- und Nachteilen angeschnitten werden. Am Ende des Kapitels wird die Methode des Multiplen Mini-Interviews ausführlich dargestellt und diskutiert und als potentiell geeignetes Verfahren zur Kompetenzmessung in der Lehrerbildung vorgeschlagen werden.

2.3.1 Berufseignungsdiagnostik

Ein spezielles Feld der Leistungsbeurteilung ist die Berufseignungsdiagnostik, welches primär Leistungen und Fähigkeiten im beruflichen Kontext fokussiert. Eine Beurteilung der Berufseignung erfolgt über einen präzisen Abgleich zwischen Kenntnissen und Fähigkeiten eines Bewerbers und den tätigkeitsspezifischen Anforderungen des Arbeitsplatzes. Es soll analysiert werden, inwiefern das individuelle Personenprofil des Bewerbers die gewünschte Leistungsfähigkeit des Berufes ermöglicht, so dass eine eignungsbezogene Prognose abschätzbar wird (Kleinmann, 1997; Päßler et al., 2011). Berücksichtigt man dabei weniger die bisherigen Leistungen eines Bewerbers, sondern betrachtet spezielle Fähigkeiten und erfolgsrelevante Eigenschaften als Bedingung für zukünftiges Verhalten, spricht man von einer Potentialanalyse (Schuler, 2004). Im Vordergrund steht hier eine langfristige Zielorientierung: Durch das Erfassen von Entwicklungsmöglichkeiten einer Person, ist diese in der Lage, künftig selbstgesteuert passende Berufswege einzuschlagen. Durch den Abgleich des beobachteten und typischen Verhaltens mit den geschätzten Leistungsmöglichkeiten einer Person kann Feedback darüber gegeben werden, ob gezeigtes Verhalten ober- - beziehungsweise unterhalb der eigenen Möglichkeiten liegt. Somit lässt sich Verhalten durch Informationen über das eigene Leistungspotential gezielt verbessern (Schuler, 2004).

Als Analyseinstrumente für die Beurteilung von Leistungen haben sich drei Modalitäten bewährt: Der biographische Ansatz nutzt diagnostische Informationen über vergangenes Verhalten und gezeigte Leistungsergebnisse für Prognosen zur zukünftigen Passung zwischen Bewerbern und Berufsanforderungen. Gegenstand des eigenschafts- oder konstruktorientierten Ansatzes sind stabile und homogene Eigenschaften des Bewerbers, welche auch als „Traits“ bezeichnet werden. Der simulationsorientierte Ansatz stützt sich hingegen auf inhaltsvalide Verhaltensstichproben relevanten beruflichen Verhaltens und arbeitet realitätsnah, indem Bewerber in Simulationen von tatsächlichen beruflichen Aufgaben geprüft werden (Päßler et al., 2011). Hierzu zählt auch das Performance Assessment, eine besonders verhaltensbezogene Form des Assessment Centers, welches im darauffolgenden Kapitel erläutert wird. Insgesamt unterliegt der eignungsdiagnostische Prozess

hohen Qualitätsansprüchen, denn das jeweilige Instrument muss in der Lage sein, die tatsächliche und individuelle Ausprägung des Bewerbers auf den Anforderungsdimensionen korrekt zu erfassen. Nur bei einer verlässlichen Differenzierung zwischen Merkmalen sowie zwischen Personen kann das Instrument vergleichbare Urteile liefern (Schuler, 2004). Neben einer verzerrungsfreien und genauen Messung ist es daher ebenso wichtig, theoretisch zu klären, was das Instrument in der Anwendung überhaupt misst (Kleinmann, 1997). Nur bei ausführlicher Untersuchung der Validität eines Instruments können die Angemessenheit der Schlüsse aus den Ergebnissen und die Richtigkeit der Interpretation dieser sinnvoll bewertet und Prognosen angefertigt werden (Schuler, 2004).

Anforderungen an ein Beurteilungsverfahren liegen zunächst einmal in der Erfassung leistungsrelevanter Aspekte, die der entsprechenden realen Tätigkeit nahe kommen (Schuler, 2004). Für die Prognose darüber, ob ein Bewerber zum Tätigkeitsprofil anhand seiner Fähigkeiten passt, muss Klarheit darüber herrschen, welche Anforderungen ein Bewerber überhaupt erfüllen muss. Die feste Grundlage bildet hier also ein Anforderungsprofil, welches aus der Analyse der konkreten Verhaltensanforderungen einer spezifischen Tätigkeit entsteht (Schuler, 2004). Eine solche Anforderungsanalyse gibt Aufschluss darüber, welche Kompetenzen in welchem Ausmaß von einem Bewerber leistungs- und erfolgsrelevant einzustufen sind (Kanning, 2009a). Hierzu zählt auch eine theoretisch fundierte Abklärung der möglichen Prädiktoren für beruflichen Erfolg, der individuellen Voraussetzungen für Zufriedenheit im jeweiligen Beruf und der Identifizierung von personalen Ressourcen gegen Belastungen (Päßler et al., 2011).

Die Beurteilung der Leistung kann in diesem Rahmen zum Beispiel anhand von Leistungskriterien oder –zielen erfolgen, so dass a priori theoretische Annahmen festgelegte Beobachtungseinheiten als Erfolgsindikatoren definieren (Schuler, 2004). Für jede beobachtete Leistung ergibt sich ein Leitfaden für deren Beurteilung, wobei spezifische Aspekte der Leistung aufgeführt sind und eine Rubrik mit verschiedenen Dimension erstellt wird, innerhalb derer die Qualität der Leistung eingestuft werden kann (Stiggins et al., 2004). In jedem Fall muss eine transparente und explizite Übersetzung von Verhalten in stimmige Scores erfolgen, welche dann quanti- und qualifizierbar sind (Clauser, 2000). Die Kriterien für eine solche Wertung von Verhaltensaspekten werden in der Praxis häufig von so genannten Experten festgelegt, die sich dabei auf Wissen und Erfahrung in dem spezifischen Themengebiet stützen. Dieses Vorgehen führt jedoch zu erheblichen Einschränkungen der Validität, denn so gewichten Experten möglicherweise stark subjektive, unklare oder obskure Details als Beurteilungskriterien, welche kaum mit den realen Fähigkeiten von Bewerbern zu vereinbaren sind (Clauser, 2000; Wiley & Haertel, 1996).

Theoretisch fundierte Einstufungsskalen liefern in diesem Kontext einen verlässlicheren und deutlich objektiveren Beurteilungsprozess. Verhaltensbeobachtungen werden mehrstufigen Ska-

lenpunkten zugeordnet, die in der Regel durch Zahlenwerte (optimale Skalenlänge zwischen 5 und 9), Adjektive oder adverbiale Bestimmungen festgelegt sind (Schuler, 2004). Bei verhaltensverankerten Einstufungsskalen können alle oder vereinzelte Skalenwerte mit positiven, neutralen oder negativen Verhaltensbeschreibungen beispielhaft verankert werden, was den Vorteil eines deutlich erkennbaren Anforderungsbezugs und Verhaltensorientierung mit sich bringt. Außerdem bieten verhaltensverankerte Einstufungsskalen weniger Raum für subjektive Interpretationen des Verhaltens oder der Urteilsdimension. Häufig finden auch Verhaltensbeobachtungsskalen Anwendung, welche im Likertformat fünf Skalenpunkte besitzen und jeweils bei der oberen und unteren Extremausprägung durch Häufigkeitsbezeichnungen abgrenzt sind (Schuler, 2004). Die Objektivität der Beurteilung kann außerdem durch ein entsprechendes Training der Beurteiler erhöht werden, so dass diese für Fehler sensibilisiert sind und nach Möglichkeit Beurteilerfehler wie etwa den Halo-Effekt (von bereits bekannten Eigenschaften einer Person auf noch unbekannte schließen), Milde-/Strenge-Effekt (durch ein negatives/positives Gesamtbild wird die Beurteilung verzerrt) oder die Tendenz zur Mitte (überwiegend mittlere Skalenwerte werden ausgewählt) vermeiden (Kleinmann, 1997). Werden bei einem Beurteilungsprozess objektive Kriterien etabliert und ein ausreichendes Beurteilertraining durchgeführt, unterscheiden sich die Urteile von Laien und Experten in ihrer Qualität grundsätzlich nicht (Wiley & Haertel, 1996).

Grundsätzlich ist aber davon auszugehen, dass eine vollständige Objektivität, also eine ausnahmslose Unabhängigkeit des Urteils von der Person des Beurteilers kaum zu erreichen ist. Häufig greift man in diesem Kontext auf die Beurteilerübereinstimmung zurück, indem mehrere Beurteiler den gleichen Beurteilungsgegenstand bearbeiten und die Ergebnisse in ihrer Übereinstimmung überprüft werden. Kommen zwei oder mehrere Beurteiler zu demselben Ergebnis, spricht dies für eine hohe Interraterreliabilität (Schuler, 2004). Um die Diskrepanzen in den Urteilen mehrerer Beurteiler zu bereinigen oder zumindest abzumildern, legen Empfehlungen die Verwendung von mindestens zwei unabhängigen Beurteilern pro Beobachtung nahe (Wiley & Haertel, 1996).

Da innerhalb der Leistungsbeurteilung auch Rückmeldungen zu den Leistungsergebnissen der Bewerber gegeben werden, spielen Kriterien der Akzeptabilität durchaus eine Rolle. Beurteilungsverfahren werden als fair und akzeptabel wahrgenommen, wenn der ganzheitliche Prozess sozialverträglich und rücksichtsvoll gehandhabt wird. Die soziale Validität ist erheblich zu steigern, indem Bewerbern im Zuge einer klaren Transparenz ausreichend Informationen zum Verfahren mitgeteilt werden und somit das Gefühl der Partizipation und Kontrolle möglich gemacht wird. Nicht zu vernachlässigen ist außerdem die Urteilkommunikation und das Feedback, um für den Bewerber eine verhaltenssteuernde und lernfördernde Wirkung zu entfalten. Hierbei können dem Bewerber wertvolle Hinweise zu Verbesserungsmöglichkeiten und mögliche Unterstützung in der

Entwicklung zugänglich gemacht werden. Auf der anderen Seite steht natürlich insbesondere für den Anwender der Berufseignungsdiagnostik die Praktikabilität und der Nutzen eines solchen Verfahrens im Vordergrund: Die Leistungsbeurteilung sollte insgesamt kosten- und ressourcensparend durchführbar sein, sich somit ökonomisch, praktikabel und verständlich in der Verwendung auszeichnen (Schuler, 2004).

2.3.2 Performance Assessment

Wie bereits im vorigen Kapitel erläutert, erfassen Leistungsbeurteilungen die Fähigkeiten einer Person nicht nur über spezifische Wissenstests, sondern können auch einen biographischen, eigenschafts- oder simulationsorientierten Ansatz verfolgen. Besonders letzterer erfüllt das Merkmal, auf authentische Art und Weise Leistungen zu messen, indem genau das Verhalten zum Gegenstand der Messung gemacht wird, welches für wichtige berufliche Aufgaben von Bedeutung ist (Höft & Funke, 2006). Durch realitätsnahe Simulationen von beruflich relevanten Situationen erweist sich eine simulations- und verhaltensorientierte Diagnostik, in der Literatur oft als Performance Assessment bezeichnet, als eine potentiell valide Methodik (Linn, Baker & Dunbar, 1991; Palm, 2001). Das Performance Assessment verlangt dabei, dass das zu beobachtende Verhalten eines Bewerbers oder Kandidaten die größtmögliche Übereinstimmung mit dem Verhalten aufweist, das im realen Kontext von Bedeutung ist. Mitunter bedeutet dies die Notwendigkeit von Equipment und Material, um einen realitätsnahen Kontext herzustellen, der einen klaren Bezug zu Problemstellungen und Inhalten des Anwendungsfelds erkennen lässt (Palm, 2001). Dieses Vorgehen steht im klaren Kontrast zu ressourcen- und zeitsparenden Papier-Bleistift Tests oder Multiple Choice Formaten, deren Ergebnisse in der Regel erst dann an Aussagekraft gewinnen, wenn sie als Korrelate anderer Leistungen interpretiert werden (Linn et al., 1991).

Grundlage des Performance Assessments sind stets Beobachtungen von Verhalten, um anhand derer ein Urteil über die Qualität der Leistung zu erstellen (Stiggins et al., 2004). Diese Methode bietet einen viel indirekteren Zugang zu Fähigkeiten einer Person, da erst aus verschiedenen Verhaltensweisen in unterschiedlichen Situationen ein Schluss darüber gezogen werden kann, in welcher Ausprägung eine Fähigkeit vorliegt. Die jeweilige Ausprägung der Fähigkeit oder Kompetenz einer Person bleibt über die Situationen hinweg die stabilste Einflussgröße für das Verhalten (Kanning, 2009a). Dem Performance Assessment wird damit zugeschrieben, verstärkt Fähigkeiten aus den Bereichen der Problemlösung und dem Problemverständnis, kritischem Denken sowie metakognitiven Prozessen abdecken zu können (Linn et al., 1991). Anstelle von fachlichen Qualifizierungen zählt viel mehr die Verhaltenskompetenz als übergeordnete Leistung, welche durch motivationale, kommunikative und soziale Fähigkeiten beeinflusst wird (Kleinmann, 1997). Je nach Aufgabenformat spielen also immer auch kommunikative Kompetenzen eine entscheidende Rolle, etwa

bei Interaktionen mit anderen Personen oder dem Halten einer Präsentation (Stiggins et al., 2004). Soziale Kompetenzen etwa lassen sich fast nur über situative Verhaltensbeobachtungen auf indirekte Art erfassen. Das Performance Assessment wird in diesem Kompetenzbereich als simulationsorientiertes Verfahren zur Methode der Wahl (Kanning, 2009a).

Die Beobachtungssettings in simulations- und verhaltensorientierten Verfahren lassen sich in zwei verschiedene Aufgabenformate unterteilen: Zum einen die individuellen, situationsgebundenen Aufgaben, bei denen Kandidaten in Einzelarbeit mit Problemsituationen aus dem tatsächlichen Tätigkeitsfeld konfrontiert werden. Diese können beispielsweise über Präsentationen, Fallübungen oder Stegreifreden gelöst werden. Zum anderen erfordern interaktive, situationsgebundene Aufgaben das Zusammenspiel mit mindestens einer weiteren Person, um die Aufgabenstellung zu bewältigen. Hierbei werden häufig Rollenspiele eingesetzt und prüfen in erster Linie soziale Kompetenzen (Höft & Funke, 2006; Kanning, 2009a).

In der Praxis wird besonders häufig das klassische Assessment Center angewendet, welches im Rahmen einer groß angelegten Testbatterie häufig viele Elemente des Performance Assessment integriert, doch darüber hinaus noch zusätzliche Übungen hinzufügt, die eher den Charakter von Leistungstests besitzen (Höft & Funke, 2006; Kanning, 2009a). Auf diese Weise können viele Bewerber durch systematische und standardisierte Aufgabentypen zur gleichen Zeit beurteilt werden, indem primär durch Verhaltensbeobachtungen auf die entsprechenden Kompetenzen der Bewerber geschlossen wird (Bieri Buschor et al., 2006; Kleinmann, 1997). Der klare Vorteil eines Assessment Centers liegt in den Grundlagen der Konzipierung: Durch die vielen verschiedenen Übungsformen und Aufgaben zeichnet sich das Verfahren durch seine hohe Methodenvielfalt aus (Bieri Buschor et al., 2006). Außerdem kann bei einem Assessment Center von einem Multitrait-Multimethod-Prinzip gesprochen werden, da stets eine Vielzahl an Eigenschaften und Fähigkeiten anhand ebenso vieler Methoden unabhängig voneinander erhoben werden können (Höft & Funke, 2006; Kanning, 2009a). Eine weitere Stärke liegt im Mehraugenprinzip: Es werden mehrere geschulte Beobachter eingesetzt, um die Verhaltensleistungen zu beurteilen und damit möglichen Beurteilungsfehlern und -verzerrungen vorzubeugen (Schuler, 2004).

Performance Assessment oder das Assessment Center im Allgemeinen erfordert in seiner Entwicklung und Durchführung hohe Kosten, zumal die Anwendung des Assessment Centers in seiner Dauer von einem halben Tag bis zu drei Wochen variieren kann (Bieri Buschor et al., 2006; Kleinmann, 1997). Obwohl es grundsätzlich als prognostisch valides Instrument eingestuft wird, bleiben Untersuchungen zur Konstruktvalidität des Assessment Centers insgesamt unzureichend. Ein gut strukturiertes und nur aus wenigen Dimensionen bestehendes Beurteilungssystem erweist sich als positiv beeinflussende Variable für die Validität. Berücksichtigt man diese Variablen, kann

das Assessment Center in vielen Bereichen der Personalauswahl, -entwicklung sowie der interpersonellen Stärken- und Schwächenanalyse brauchbar eingesetzt werden (Kleinmann, 1997). Bei letzterem kombiniert man die simulationsorientierten und fremdbeurteilten Übungsformate mit Elementen der Selbsteinschätzung, um einen optimalen diagnostischen Zugang zu Stärken und Schwächen zu erhalten (Bieri Buschor & Schuler Braunschweig, 2011). Auf die Verwendung von Selbsteinschätzungen in der Eignungsdiagnostik soll im nächsten Kapitel eingegangen werden.

2.3.3 Selbsteinschätzungsverfahren

Selbsteinschätzungen ergänzen nicht nur simulationsorientierte Verfahren wie das Assessment Center, sondern spielen ebenfalls in der Eignungsabklärung im Lehrberuf eine bedeutende Rolle (siehe Kapitel 2.3.1). Es ist nicht zu bestreiten, dass sich der Vergleich der eigenen Einschätzung mit der von anderen förderlich auf die Selbstreflexion auswirkt und die Erhebung von Selbsteinschätzungen äußerst einfach und wenig aufwändig erfolgen kann (Bieri Buschor & Schuler Braunschweig, 2011; Moser, 2004). Betrachtet man aber Selbsteinschätzungen als eigenständigen Zugang zu den Kompetenzen einer Person, bleibt bislang unklar, was über Selbsteinschätzungen tatsächlich erfasst wird. Selten wurde bislang empirisch untersucht, ob und inwiefern überhaupt eine Übereinstimmung mit anderen Instrumenten der Kompetenz- und Leistungsmessung vorliegt (König et al., 2012).

Wird ein Bewerber gebeten, seine eigenen Fähigkeiten und Kompetenzen anhand einer vorgegebenen Skala einzuschätzen, können subjektive Beurteilungsverzerrungen kaum ausgeschlossen werden. Daten aus Selbsteinschätzungsverfahren lassen sich im Grunde genommen nur dann vergleichen, wenn den sich einschätzenden Personen ein gemeinsamer Vergleichsmaßstab vorgegeben wurde. Andernfalls bleibt nicht die tatsächliche Fähigkeitsausprägung die stärkste beeinflussende Variable, sondern es rücken persönliche Merkmale wie Selbstwirksamkeitserwartungen, motivationale Voreingenommenheit oder soziale Erwünschtheit in den Vordergrund (König et al., 2012). Auch das Selbstvertrauen einer Person kann das Urteil der Selbsteinschätzung moderieren. Bei starker Überzeugung von der eigenen Person und hoher Selbstachtung fallen Urteile über die eigene Leistung, verglichen mit einer Fremdbeurteilung, nicht selten stark überbewertet aus. Gerade bei schwach ausgeprägten Fähigkeiten kann über eine besonders milde und somit selbstwertdienliche Beurteilung der eigenen Leistung der Selbstwert geschützt oder gar erhöht werden (Moser, 2004).

Im Gegensatz zu simulationsorientierten Verfahren, welche Leistungen durch tatsächliche Beobachtung zu beurteilen versuchen, greifen Verfahren der Selbsteinschätzung auf Verhaltensbeschreibungen oder Attributnennungen zurück, welche der Bewerber durch einen Abgleich mit sei-

nem vorherigem Verhalten in ihrer individuellen Ausprägung bewertet. Bei Selbsteinschätzungen bilden also lediglich Erinnerungen an entsprechendes Verhalten und dessen Erfolg die Grundlage für den Urteilsprozess. Hier wird deutlich, dass die Angewiesenheit auf Erinnerungsvorgänge eine starke Quelle für Verzerrungen bietet (Moser, 2004). Ob und vor allem was der Bewerber im Moment der Selbsteinschätzung bezüglich seines eigenen Verhaltens und der Leistung erinnert, muss nicht zwangsläufig mit der Realität übereinstimmen. Auch hier können die oben genannten Variablen beeinflussend wirken. Schlussendlich können Selbsteinschätzungen zwar den Informationsgehalt von simulationsorientierten Verfahren erhöhen, sind aber in ihrer Verlässlichkeit als alleinstehende Methode stark eingeschränkt.

2.3.4 Multiple Mini-Interviews

In den vorangegangenen Kapiteln wurden bereits verschiedene Vorgehensweisen in der Leistungsdiagnostik vorgestellt und auf ihre Vor- und Nachteile hin untersucht. Das Assessment Center scheint eine Vielzahl von Stärken bezüglich der verhaltensbezogenen Beurteilung von Fähigkeiten und Kompetenzen aufzuweisen, ergänzt jedoch Elemente des Performance Assessment mit eher leistungstestähnlichen Übungsformaten. Ein weiteres Verfahren, das sowohl den Multi-Gedanken des Assessment Centers bezüglich Methoden- und Kompetenzvielfalt als auch eine stark am Verhalten ausgerichtete Diagnostik miteinander kombiniert, ist das Multiple Mini-Interview.

Das Multiple Mini-Interview (MMI) wurde ursprünglich im Jahr 2001 an der McMaster Universität in Hamilton, Kanada, entwickelt und fungiert dort noch heute als strukturierte Auswahlmethode für Bewerber des Medizinstudiums (Knorr & Hissbach, 2014; Pau et al., 2013). Mit diesem Verfahren sollen in erster Linie nicht-kognitive Kompetenzen wie etwa moralisches Urteilsvermögen oder Kommunikations- und Kooperationsfähigkeit der Bewerber zuverlässig gemessen werden, um eine optimale Passung zwischen Eingangsvoraussetzungen und Anforderungen des Studiums und Berufs zu garantieren (Reiter, Eva, Rosenfeld & Norman, 2007; Tiller et al., 2013). Die Bewerber durchlaufen hierbei eine Reihe von standardisierten, kurzen Stationen mit verschiedenen Aufgaben und werden dabei von geschulten Beobachtern beurteilt (Bingham & Scharf, 2011). Mittlerweile hat sich das MMI als Aufnahmeverfahren in vielen englischsprachigen und vereinzelt europäischen Ländern verbreitet und findet primär in medizinisch-gesundheitlichen Ausbildungen wie Allgemein-, Zahn- und Tiermedizin sowie Pharmazie seine Anwendung (Knorr & Hissbach, 2014, Pau et al., 2013).

Trotz seiner mittlerweile zahlreichen Anwendung stellt sich das MMI nicht als ein spezifisch definiertes Instrument heraus, das in seinem Inhalt und seiner Qualität zu stets gleichen Ergebnissen führt. Es ist vielmehr ein Rahmenkonzept für den Prozess der Beurteilung von unterschiedlichen

Fähigkeiten (Eva et al., 2012). So kann das MMI in einer Vielzahl von Eigenschaften inhaltlich als auch methodisch modifiziert und den Ansprüchen des jeweiligen Anwendungskontexts angepasst werden. Im Folgenden sollen die wichtigsten Variablen beleuchtet und auf deren Auswirkungen auf die Qualität des MMIs eingegangen werden.

Der ursprüngliche Anspruch an das MMI, ausschließlich nicht-kognitive Fähigkeiten zu erfassen, lässt nichtsdestotrotz den Anwendern relativ großen Freiraum darüber, welche Kompetenzen tatsächlich mit der Methode erhoben werden. Die Entwicklung eines programmspezifischen MMIs setzt insbesondere die Festlegung derjenigen Fähigkeiten voraus, welche das jeweilige Institut zu messen beabsichtigt. Im besten Fall basiert dieser Prozess auf Anforderungsanalysen sowie fundierten theoretischen Recherchen zu entsprechenden Kompetenzen, häufig legen jedoch so genannte Experten nach subjektiven Kriterien fest, welche Kompetenzen erfasst werden sollen (Knorr & Hissbach, 2014). Dabei variiert die Anzahl der getesteten Fähigkeiten von drei bis 19, darunter lassen sich häufig Arten der Kommunikationsfähigkeit wiederfinden (Knorr & Hissbach, 2014). Generell sollen die Stationen jedoch so gestaltet sein, dass bei deren Bearbeitung keinerlei Rückgriff auf spezifisch gelerntes Wissen notwendig ist und somit nur der situative Umgang mit der Aufgabe zur Beurteilung der festgelegten Fähigkeiten herangezogen wird (Bingham & Scharf, 2011; Pau et al., 2013).

Die Flexibilität in der Entwicklung der Stationen lässt je nach Interesse und Absichten des verwendenden Instituts verschiedene Aufgabenformate zu. Dies verhindert nicht zuletzt, dass sich die Bewerber auf die Aufgaben vorbereiten können, indem sie beispielsweise Antworten im Voraus auswendig lernen (Eva, Rosenfeld, Reiter & Norman, 2004c). Typische Aufgabenformen der einzelnen Stationen des MMIs sind beispielsweise Diskussionen zu vorgelegten Themen, Interaktionen mit instruierten Schauspielenden, die Beantwortung standardisierter Fragen oder Bearbeitung festgelegter Aufgabenstellungen (Knorr & Hissbach, 2014). Die jeweiligen Stationen werden a priori festgelegt und in ihrem Inhalt und Ablauf bestmöglich standardisiert, so dass jeder Bewerber den gleichen Satz an Aufgaben bearbeitet und die Leistungen vergleichbar werden (Eva, Rosenfeld, Reiter & Norman, 2004c). Jede Station besteht in der Regel aus einer kurzen Vorbereitungszeit, in der die Bewerber die Szenarien und Instruktionen zu der jeweiligen Station durchlesen, einer Bearbeitungszeit, in der die Äußerungen und Interaktionen des Bewerbers beobachtet und im Anschluss mit einem Beurteilungsbogen bewertet werden (Rosenfeld, Reiter, Trinh & Eva, 2008). Bei einem klassischen MMI findet jede Station in einem separaten Raum statt, so dass mehrere Bewerber durch Zuteilung zu den entsprechenden Räumen zeitgleich mehrere Situationen durchlaufen. Akustische Signale unterteilen den Ablauf deutlich in Vorbereitungs- und Bearbeitungsabschnitte und

geben den Raumwechsel für alle Bewerber an, so dass ein weitgehend automatisch wechselnder Zirkel entsteht (Rosenfeld et al., 2008).

Die Anzahl und das Format der Stationen sowie die Dauer der einzelnen zeitlichen Abschnitte kann ebenso programmspezifisch angepasst werden, wie es bei den zu messenden Fähigkeiten der Fall ist. Doch besonders die Anzahl der Stationen sowie die vorgegebene Bearbeitungszeit pro Station üben erheblichen Einfluss auf die Qualität der Messung und Handhabbarkeit der Methode aus. Pau et al. (2013) fassten in einem Review die Ergebnisse von insgesamt 30 publizierten Studien zusammen, welche das Multiple Mini-Interview als Methode verwendeten. Die verschiedenen MMIs bestanden in der Regel aus vier bis zwölf Stationen, wobei die Zeit pro Station von fünf bis fünfzehn Minuten variierte und jede Station mit einem Beobachter besetzt war. Das durchschnittliche MMI besteht somit aus 10 Stationen, welche jeweils eine Bearbeitungsdauer von acht Minuten aufweisen, wohingegen andere Autoren die optimale Stationenanzahl zwischen 7 und 12 zusammenfassen (Rees et al., 2016).

Das wiederholte Erfassen von Leistungen über multiple Stationen stellt definitiv eine Hauptstärke des MMIs dar und prägt maßgeblich die psychometrische Qualität, insbesondere die Reliabilität der Methode (Eva et al., 2012; Knorr & Hissbach, 2014; Serres & Peeters, 2012). Generell zeichnet sich das MMI durch eine vergleichsweise akzeptable Reliabilität aus, es lassen sich Reliabilitätskoeffizienten zwischen 0.65 und 0.85 finden (Eva et al., 2004c; Hecker et al., 2009; Hissbach, Sehner, Herendza & Hampe, 2014; Pau et al., 2013; Rees et al., 2016; Serres & Peeters, 2012). Dennoch variiert die Messgenauigkeit der MMIs erheblich, Messungen nur einer einzelnen MMI Station weisen eine sehr geringe Generalisierbarkeit auf und sind stark kontext-, beurteiler- und situationsabhängig (Eva, Reiter, Rosenfeld & Norman, 2004; Hissbach et al., 2014; Knorr & Hissbach, 2014).

In der Praxis wurden hierzu versuchsweise Modifikationen bezüglich der Anzahl der Stationen, Dauer pro Station sowie Anzahl der Beobachter/Beurteiler je Station vorgenommen, um Verbesserungen in der Reliabilität zu erzielen (Hissbach et al., 2014). Insgesamt zeigte sich sehr deutlich, dass eine Erhöhung der Stationenanzahl einen größeren Einfluss auf die Reliabilität hat als eine längere Stationsdauer oder höhere Beobachteranzahl. Die Verwendung mehrerer Stationen beugt Verzerrungen durch situative Einflüsse und Zufall am kosteneffektivsten vor, infolgedessen sollte die Stationenanzahl und damit verbundene Verfügbarkeit von Räumlichkeiten im Mittelpunkt der Ressourcenverteilung des MMIs stehen (Eva et al., 2004; Eva et al., 2004c; Hissbach et al., 2014; Knorr & Hissbach, 2014; Rees et al., 2016). Nicht zu vernachlässigen sind gleichwohl Verzerrungen in den Beurteilungen und Effekte durch Interaktionen zwischen Bewerbern und Beurteilern, so dass der Einsatz von mindestens zwei Beurteilern pro Station trotz steigender Personalkosten nicht ver-

nachlässigt werden darf (Hecker et al., 2009; Hissbach et al., 2014). Außerdem erweist sich die Subjektivität von Beurteilern als relativ gewichtige Fehlerquelle, eine genaue Festlegung, wer als Beurteiler fungieren darf, liegt bisher nicht vor. Häufig greifen Institute auf eigenes Personal zurück, seltener werden Studierende oder Beurteiler aus der direkten Berufspraxis rekrutiert (Pau et al., 2013). Nur ein entsprechend umfangreiches Training für alle Beurteiler sowie strukturierte Beurteilungsbögen garantieren einen standardisierten Ablauf für jede Station und eine objektive Bewertung der Leistung des Bewerbers (Bingham & Scharf, 2011; Eva et al., 2004c; Hecker et al., 2009).

Durch die Möglichkeit, die Leistung von Bewerbern über mehrere Situationen hinweg zu erfassen, lassen sich die einzelnen Begegnungen mit den jeweiligen Beobachtern relativ kurz gestalten (Eva et al., 2012). Dodson et al. (2009) zufolge reichen bereits fünf Minuten pro Station aus, ohne starke Einbußen in der Reliabilität oder den Leistungen der Bewerber zu provozieren. Nach bereits 5-6 Minuten haben die meisten Beurteiler bereits einen relativ festen Eindruck von der Leistung des Bewerbers gewonnen, so dass die Zeit darüber hinaus größtenteils redundante Informationen generiert, von der allenfalls etwas leistungsschwächere Bewerber profitieren (Dodson et al., 2009; Eva et al., 2012; Knorr & Hissbach, 2014). Durch Reduktionen in der Stationendauer können nicht zuletzt Ansprüche der kosten- und zeitsparenden Durchführbarkeit umgesetzt werden.

Ogleich die Reliabilität der MMIs häufig Gegenstand von Verbesserungsversuchen in der Forschung ist, bleiben Angaben zur Validität der Methode überwiegend ungeklärt. Es herrscht vielmehr eine Kontroverse darüber, was ein MMI eigentlich misst (Patterson et al., 2016; Tiller et al., 2013). Letztendlich ist die Validität eines jeden MMIs sehr stark davon abhängig, auf welcher Basis und mit welchem Inhalt die einzelnen Stationen gestaltet sind – dies obliegt der jeweiligen Absicht des durchführenden Instituts (Rees et al., 2016). Da das MMI hauptsächlich zur Auswahl von Studierenden in medizinischen Ausbildungen angewendet wird, existieren bislang auch nur vereinzelte Aussagen zur Validität in diesem begrenzten Fachbereich. So korreliert die Leistung im MMI nicht mit traditionellen Auswahlverfahren und weist keine Assoziationen mit akademischen Qualifikationen oder dem Notendurchschnitt vor Beginn des Studiums auf (Hecker et al., 2009; Knorr & Hissbach, 2014; Pau et al., 2013). Jedoch zeigt sich das Abschneiden in einem MMI als bester Prädiktor für die Leistung im OSCE (Objective Structured Clinical Examination)¹ und sagt sowohl die anschließende Leistung im ärztlichen Referendariat als auch in der medizinischen Abschlussprüfung vorher (Pau et al., 2013; Reiter, Eva, Rosenfeld & Norman, 2007). Auch können Informationen aus dem MMI für eine Vorhersage von intra- und interpersonellen Eigenschaften herangezogen werden und sind dabei zuverlässiger als beispielsweise Selbstauskünfte oder Empfehlungen (Patterson et al.,

¹ OSCE ist ein mündlich-praktisches Prüfungsverfahren im Rahmen des Medizinstudiums. Es erfasst praktische Fähigkeiten des medizinischen Alltags.

Siehe: <http://flexikon.doccheck.com/de/OSCE>, abgerufen am 28.12.2017

2016). Studierende, welche durch ein MMI-Auswahlverfahren zum Studium zugelassen wurden, übertrafen die abgelehnten und an anderen Universitäten ausgebildeten Studierenden in ihrer Leistung sogar 4-6 Jahre nach der Auswahlentscheidung, und dies lässt sich nicht allein durch den Lehrplanerfolg der jeweiligen Universität erklären (Eva et al., 2012). Zudem fanden Jerant et al. (2012) einen Zusammenhang zwischen Persönlichkeitsfaktoren und der MMI-Leistung. Sie konnten zeigen, dass Medizinstudierende mit besonders hoch ausgeprägter Extraversion signifikant höhere Punktzahlen im MMI erreichten als Studierende mit niedrigen Werten der Extraversion. Die Persönlichkeitsfaktoren Extraversion und Verträglichkeit standen ebenfalls gemeinsam mit dem Abschneiden im MMI in einem positiven Zusammenhang mit der Annahme an einer Medizinstudie.

Wie sich anhand der wenigen Befunde feststellen lässt, mangelt es dem MMI insgesamt an Befunden zur Validität, besonders aber der prognostischen Validität (Knorr & Hissbach, 2014; Pau et al., 2013). Deutlich wird ein klarer Bedarf an Followup-Studien mit längsschnittlichem Design, um langzeitliche Informationen über das Abschneiden von Studierenden im Studium und Beruf zu erhalten, die durch ein MMI zum Studium zugelassen wurden (Patterson et al., 2016). Einzig und allein die Augenscheinvalidität wird in der Mehrheit der publizierten Studien als hoch eingeschätzt, diese spricht jedoch primär für die allgemeine Akzeptanz der Methode (Eva et al., 2004c; Hecker et al., 2009; Knorr & Hissbach, 2014; Rees et al., 2016). Bewerber, die ein MMI absolviert haben, betonten ausführlich die wahrgenommene Fairness der Methode. Durch die verschiedenen Stationen mit jeweils wechselnden Beobachtern kommt es zu keiner Voreingenommenheit bezüglich Geschlecht, Alter, sozioökonomischen Status oder kulturellen Verschiedenheiten (Pau et al., 2013; Rees et al., 2016). Darüber lassen die multiplen Stationen die Möglichkeit offen, schlechteres Abschneiden bei vorherigen Aufgaben durch höhere Anstrengungen in den folgenden Stationen auszugleichen und somit einen soliden Gesamteindruck zu hinterlassen (Bingham & Scharf, 2011; Eva et al., 2004c; Pau et al., 2013). Andere Studien berichten allerdings auch darüber, dass eine misslungene Aufgabe zu erhöhter Nervosität bei den Bewerbern führte und sich dies durchaus hemmend auf eine effektive und strukturierte Bearbeitung der weiteren Aufgaben auswirken kann (McAndrew & Ellis, 2012). Dass die Leistungen von Station zu Station unterschiedlich ausfallen können, lässt sich demnach zugleich als Schwäche und Stärke interpretieren.

Ein Schwachpunkt des MMIs zeichnet sich gewiss in seinen hohen personalen, logistischen und zeitlichen Kosten ab. Die Verwendung eines MMIs zur Auswahl von geeigneten Studierenden zu einem Studiengang führt zu höheren Kosten als beispielsweise ein schriftlicher Aufnahmetest (Hissbach et al., 2014). Besonders kostenintensiv ist vor allem der Einsatz von Personal mit entsprechender Expertise für die Entwicklung der Stationen, der Betreuung in der Durchführung sowie der Auswertung der gesammelten Daten, welche mindestens zwei Beurteiler verlangt (Rosenfeld et al.,

2008). Auch die Bewerber investieren nennenswerte zeitliche Ressourcen, bei einem durchschnittlichen MMI mit zehn Stationen, jeweils mit einer Bearbeitungsdauer von zehn Minuten, ergibt dies 100 geleistete Minuten pro Kandidat. Die Durchführbarkeit eines MMIs hängt deutlich von den finanziellen und personalen Ressourcen einer Fakultät ab, jedoch müssen auch für einen festgelegten Zeitraum genügend Räumlichkeiten zur Verfügung stehen (Pau et al., 2013). Aus diesem Grund wurde bereits der Versuch gestartet, das MMI über das Internet mit dem Instant-Messaging-Dienst Skype per Video durchzuführen, so dass der Ablauf räumlich entkoppelt werden kann und Bewerber keine weiten Reisen auf sich nehmen müssen. Dieses Vorgehen erprobte sich als ressourcensparend, entsprechend reliabel, valide und von den Bewerbern akzeptiert wie ein vor Ort durchgeführtes MMI (Tiller et al., 2013). Verglichen mit einem Panel Interview, bei dem sich ein Bewerber vor mehreren Interviewpartnern behaupten muss, schneidet das MMI aber durchaus positiver ab: Insgesamt werden nicht mehr Prüfer gebraucht, die Kosten sind insgesamt etwas niedriger und die einzelnen Begegnungen mit den Bewerbern fallen deutlich kürzer aus (Pau et al., 2013; Rees et al., 2016).

Schlussendlich lässt sich das MMI als eine Methode bezeichnen, die durch multiple, unabhängige Beobachtungen in verschiedenen gestalteten Situationen durch einen objektiven Beurteilungsprozess eine reliable Messung unterschiedlicher Fähigkeiten eines Bewerbers ermöglicht. Dabei wird die Interaktion zwischen Bewerber und Beobachter in den Hintergrund gerückt, so dass der Fokus auf den situativen Leistungen in einer Reihe von standardisierten Aufgaben liegt und Zufalls- sowie Beurteilerverzerrungen vermindert werden. Die Methode des MMIs kann dessen ungeachtet nur so gut sein wie die Entwicklung ihres Inhalts. Da das MMI bislang in erster Linie im Kontext medizinischer Ausbildungen erprobt wurde, bleiben Ergebnisse zu Messungen von Fähigkeiten fernab der medizinischen Expertise bislang unterrepräsentiert. Da es sich aber grundsätzlich als potentiell zuverlässiges Instrument erweist, stellt es für die Kompetenzdiagnostik im Rahmen der Lehramtsausbildung ein vielversprechendes Verfahren dar. Hier gilt zu prüfen, inwiefern das MMI auch für andere Fachbereiche adaptierbar ist und ob die psychometrischen Gütekriterien ebenso zufriedenstellend ausfallen können, wenn es im Kontext der Diagnostik von lehrberuflichen Kompetenzen verwendet wird. Ob sich hierbei die gleichen Vorgaben wie in medizinischen Settings zur Konstruktion als erfolgsversprechend erweisen und ob das MMI dazu in der Lage ist, zwischen den Ausprägungsgraden von eignungsrelevanten Kompetenzen des Lehrberufs differenzieren zu können, bleibt empirisch zu überprüfen. Bevor das MMI als ein mögliches Instrument zur Messung von Kompetenzen in der Lehrerbildung etabliert werden kann, muss es zunächst unter inhaltsspezifischer Anwendung und Konzeption dem hohen Anspruch psychometrischer Qualität genügen.

3 Fragestellungen der vorliegenden Arbeit

Im vorherigen Kapitel wurde ausführlich dargestellt, dass das MMI kein fest definiertes Instrument zur Messung von eignungsrelevanten Kompetenzen darstellt, sondern viel mehr als ein generelles methodisches Rahmenkonzept verstanden werden kann. Es ist aber durchaus möglich, Multiple Mini-Interviews neben der bislang häufigen Verwendung im medizinischen Bereich auch im Kontext anderer Fachbereiche zu konzipieren (Eva et al., 2012; Knorr & Hissbach, 2014).

Durch die zunehmende Kompetenzorientierung in der Forschung zum Lehrberuf und der Bedeutung der Entwicklung notwendiger Kompetenzen für eine gelingende und professionelle Ausübung des Berufs, wurde das Augenmerk verstärkt auf die Lehrerbildung verlagert. Die Kompetenzdiagnostik im Zuge des Lehramtsstudiums gewinnt hierbei zunehmend an Popularität und so lassen sich zahlreiche Versuche identifizieren, diagnostische Verfahren zur Eignungsabklärung in das Lehramtsstudium einzubetten (Frey, 2006; Nieskens, 2012). Die Einsatzbereiche, wie etwa als Selektionsverfahren vor dem Studium oder als Ausgangslage für ein gezieltes Kompetenzfeedback zur Entwicklungsunterstützung, variieren derzeit noch stark und weisen unterschiedlich präferierte Instrumente zur Diagnostik auf (Frey, 2006; Nieskens, 2012).

Auch an der Universität Erfurt wird daran gearbeitet, lehrberufsrelevante Kompetenzen diagnostisch erfassen zu können. Dies erfolgte in anfänglichen Phasen noch mit der Absicht, geeignete Studierende für den Lehrberuf eignungsdiagnostisch selektieren zu können, soll nunmehr aber vermehrt für ein Kompetenzfeedback zur Entwicklungsunterstützung während der Lehrerbildung genutzt werden. Als potentiell geeignetes Verfahren zur Messung dieser lehrberufsrelevanten Handlungskompetenzen wurden die aus der Medizin stammenden simulations- und verhaltensorientierten Multiplen Mini-Interviews als ein potentiell geeignetes Verfahren zur Messung von lehrberufsrelevanten Handlungskompetenzen auserwählt. Das MMI soll als neuer Ansatz für die Erstellung eines qualitativ hochwertigen Diagnoseinstruments innerhalb der Lehrerbildung empirisch erprobt und überprüft werden. Hierzu werden lehrberufsrelevante Kompetenzen theoriegeleitet ausgewählt, um diese durch methodisch vielfältige Aufgaben als Bestandteile von Multiplen Mini-Interviews zuverlässig messbar zu machen. Es existieren zwar Forschungsergebnisse zur psychometrischen Qualität der MMIs im medizinischen Kontext, jedoch bleiben Ergebnisse fernab dieser Domäne zum MMI bisher unterrepräsentiert. Aus diesem Grund soll die vorliegende Arbeit die Adaption des Verfahrens dahingehend überprüfen, ob Multiple Mini-Interviews im Anwendungskontext der Diagnostik lehrberufsrelevanter Kompetenzen den Ansprüchen an die psychometrische Qualität standhalten können.

Als Grundlage für diese Arbeit liegt ein Gesamtset von sechs verschiedenen MMI-Aufgaben vor, welches die Ausprägungen einer Auswahl von lehrberufsrelevanten Kompetenzen bei Lehramtsstudierenden zu messen beabsichtigt. Die sechs MMI-Aufgaben erfassen jeweils verschiedene Kompetenzfacetten, darunter unter anderem Kompetenzen des Erklärens und Unterrichtens, der Berufswahl und des Planens, sowie Kommunikationsfähigkeiten, Stressbewältigung und Empathie. Als Aufgabenformate wurde bei zwei Aufgaben jeweils ein kurzes Rollenspiel mit einem instruierten Schauspieler realisiert, die restlichen vier Aufgaben erfassen die Kompetenzen anhand diskursiver Auseinandersetzungen mit verschiedenen Aufgabenstellungen in Monologform.

Aus zeitlichen Gründen kann in der vorliegenden Arbeit jedoch nur ein Teil der sechs MMI-Aufgaben berücksichtigt werden, so dass die Untersuchung hinsichtlich der psychometrischen Qualität des MMIs nur auf Grundlage von drei Aufgaben erfolgen kann, welche im folgenden Kapitel genauer dargestellt werden (siehe Kapitel 4). Ob anhand des konstruierten MMIs die Messabsichten auch im lehrberuflichen Kontext zuverlässig erreicht werden, soll mittels verschiedener Kriterien der Testgüte untersucht werden. Die Betrachtung der Testgütekriterien des MMIs stellt einen bedeutsamen Punkt zur Entscheidung über zukünftige Anwendungsmöglichkeiten des Verfahrens dar und sollte daher unbedingt bereits in der Phase der Verfahrenskonstruktion berücksichtigt werden (Kemper, Ziegler, Krumm, Heene & Bühner, 2015). Die vorliegende Arbeit überprüft den Satz von drei MMI-Aufgaben hinsichtlich der Hauptgütekriterien der Reliabilität, Objektivität und Validität. Zudem sollen Aussagen über das Nebengütekriterium der Ökonomie getroffen werden. Folgende Untersuchungshypothesen werden in dieser Arbeit untersucht:

(1) Die drei MMI-Aufgaben messen die Kompetenzbereiche zuverlässig und genau.

(1a) Die Messung weist eine hohe Reliabilität auf, indem die Urteile der eingesetzten Beurteiler für jede MMI-Aufgabe hoch miteinander übereinstimmen und zu einheitlichen Einschätzungen der Leistungen der Testpersonen führen.

(1b) Die Items der Beurteilungsbögen der drei MMI-Aufgaben sind intern konsistent und produzieren genaue Messergebnisse. Die gemessenen Unterschiede sind nicht durch Messfehler bedingt.

(2) Die Messergebnisse der MMI-Aufgaben sind insgesamt unabhängig von der Person des Testleiters.

(2a) Die Durchführung der MMI-Aufgaben ist objektiv, so dass die Leistungen der Testpersonen nicht von der Testsituation oder dem Testleiter beeinflusst werden.

- (2b) Die Auswertungsergebnisse der MMI-Aufgaben sind unabhängig von der Person des Beurteilers, so dass von einer hohen Auswertungsobjektivität gesprochen werden kann.
- (3) Die drei MMI-Aufgaben erfassen tatsächlich die zu messen beabsichtigten Kompetenzen und Fähigkeiten, so dass die Ergebnisse der Messung generalisiert werden können. Die Messungen der drei MMI-Aufgaben sind konstrukt- und kriteriumsvalide.
- (3a) Die Leistungsbeurteilungen zeigen über die drei MMI-Aufgaben konvergente und divergente Validität. Die jeweilige Leistungsbeurteilung durch die Testbeurteiler in den drei MMI-Aufgaben steht in einem positiven Zusammenhang mit der entsprechenden globalen Einschätzung der Bewältigung der MMI-Aufgabe durch den Testleiter. Die Leistungsbeurteilungen durch die Testbeurteiler in den drei MMI-Aufgaben stehen in keinem relevanten Zusammenhang zueinander.
- (3b) Die Leistungsbeurteilungen sind faktoriell valide. Die Beurteilungssitems der drei MMI-Aufgaben sind voneinander unabhängig und können jeweils dem entsprechenden Aufgabenfaktor eindeutig zugeordnet werden.
- (3c) Die Leistungsbeurteilungen sind über die drei MMI-Aufgaben kriteriumsvalide. Die jeweiligen Leistungsbeurteilungen der drei MMI-Aufgaben stehen in einem positiven Zusammenhang mit den Selbsteinschätzungen der spezifischen aufgabenrelevanten Fähigkeitsbereiche der Testpersonen. Anhand der Selbsteinschätzungen zu den aufgabenrelevanten Fähigkeiten lässt sich die Gesamtleistung über alle drei Aufgaben vorhersagen. Die aufsummierte durch die Testbeurteiler bestimmte Gesamtleistung in allen drei MMI-Aufgaben steht in einem positiven Zusammenhang mit dem globalen Gesamteindruck, den die Testleiter angeben. Items, welche insbesondere die sprachliche Qualität der Äußerungen in den MMI-Aufgaben bewerten, weisen einen Zusammenhang mit globalen Einschätzungen zur Qualität des sprachlichen Ausdrucks und der Argumentation. MMI-Aufgaben mit komplexen Aufgabenstellungen zeigen einen Zusammenhang der Gesamtleistung mit der globalen Auffassungsgabe. In Anlehnung an Jerant et al. (2012) stehen ebenfalls die Persönlichkeitsfaktoren Extraversion und Verträglichkeit mit der MMI-Leistung in Verbindung. Die Aufregung der Testpersonen beeinflusst die Gesamtleistung im MMI nicht.
- (4) Die drei MMI-Aufgaben sind insofern ökonomisch, als dass sie nur geringe finanzielle und zeitliche Ressourcen in Anspruch nehmen. Dies betrifft die Durchführungszeit und die Handhabbarkeit der Auswertung.

4 Methode

Um die Forschungsfragen zur psychometrischen Qualität des MMIs im Kontext der Erfassung lehrberufsrelevanter Kompetenzen empirisch zu überprüfen, wurden sechs MMI-Aufgaben erstmalig in ihrer Zusammensetzung in einer groß angelegten Erhebung in einem Zeitraum von vier Wochen an der Universität Erfurt durchgeführt. Da die vorliegende Arbeit zeitlichen Restriktionen unterliegt, war es nicht möglich, alle sechs MMI-Aufgaben in die Auswertungen miteinzubeziehen. Aus diesem Grund stützt sich diese Arbeit nur auf eine Auswahl von drei Aufgaben, anhand dessen Aussagen über die Qualität des MMIs getroffen werden sollen. Im Folgenden werden die Gestaltung und der genaue Ablauf der gesamten Erhebungen umfassend dargestellt, in der Beschreibung der Erhebungsverfahren wird dann jedoch nur noch auf die drei ausgewählten MMI-Aufgaben eingegangen.

4.1 Gestaltung und Ablauf der Erhebungen

Die Erhebung von sechs MMI-Aufgaben fand insgesamt in einem vierwöchigen Zeitraum vom 27.11.-21.12.2017 in den Räumlichkeiten der erziehungswissenschaftlichen Fakultät der Universität Erfurt statt. Dabei wurde die Durchführung der MMI-Aufgaben inhaltlich in zwei Teile aufgeteilt, so dass vom 27.11.-07.12.2017 zunächst die ersten drei MMI-Aufgaben ohne Rollenspiel durchgeführt wurden und die restlichen zwei Rollenspielaufgaben mit einer monologischen Aufgabe gemeinsam ab dem 11.12.2017 erfolgten. Die Beschreibung der drei ausgewählten MMI-Aufgaben, welche im Rahmen dieser Arbeit ausgewertet werden, erfolgt in Kapitel 4.2.1.

Die Erhebungen fanden zeitgleich in zwei benachbarten Räumen in einem Zeitfenster von vier Stunden statt und wiesen jeder Testperson eine Testdauer von 30 Minuten zu. Innerhalb der 30 Minuten sollten die Testpersonen die drei entsprechenden Aufgaben bearbeiten und hatten hierbei für jede Aufgabe insgesamt zehn Minuten. Jede Testperson musste dementsprechend sowohl zu einem Termin im ersten als auch im zweiten Erhebungszeitraum erscheinen, um insgesamt alle sechs Aufgaben des MMIs vollständig durchlaufen zu haben.

Die Gesamtarbeitszeit der vollständigen MMI-Konzeption bestehend aus sechs Aufgaben beträgt somit pro Person 60 Minuten. Die Terminvergabe erfolgte über einen zuvor an alle Testpersonen versandten Online-Fragebogen, innerhalb dessen zunächst demographische Angaben, Persönlichkeitseinschätzungen und Selbsteinschätzungen abgefragt wurden. Eine vollständige Beschreibung dieses Fragebogens erfolgt in Kapitel 4.2.3. Nach vollständiger Bearbeitung des Fragebogens konnten die Testpersonen zwei verbindliche Erhebungstermine online eintragen.

Um im gesamten Erhebungszeitraum für beide Räume jeweils einen Testleiter für die Durchführung bereitstellen zu können, wurden zwölf Personen aus unterschiedlichen Fachbereichen (Mitarbeitende des Teaching Talent Centers und des Lehrstuhls Pädagogisch-psychologische

Diagnostik und Differentielle Psychologie, studentische Hilfskräfte und Masterstudierende der Psychologie) sowie zuzüglich drei Schauspieler*innen für den zweiten Erhebungszeitraum mit Rollenspielaufgaben eingesetzt. Als Testpersonen konnten Studierende des Lehramts im ersten Mastersemester der Universität Erfurt gewonnen werden, indem das Absolvieren der MMIs als Leistung einer teilnehmerstarken Veranstaltung angesetzt wurde.

An jedem Erhebungstag durchliefen maximal sechzehn Testpersonen die drei angesetzten MMI-Aufgaben, wobei stets zwei Testpersonen in den separaten Räumen gleichzeitig erhoben werden konnten. Vor jedem Termin wurden die Testpersonen gebeten, zunächst im Flur zu den Räumlichkeiten zu warten, bis sie paarweise hineingebeten und jeweils einem Raum und Testleiter zugewiesen wurden. Nach einer kurzen Begrüßung und dem Notieren wichtiger Personendaten sollten die Testpersonen die MMI-Aufgaben nacheinander bearbeiten. Hierfür erhielten sie stets zwei Minuten Vorbereitungszeit pro Aufgabe, um sich die ausgeteilte Aufgabenstellung durchzulesen und sich Antworten zu überlegen. Anschließend folgte eine achtminütige Bearbeitungsdauer, in der sich die Testpersonen zu der Aufgabe äußern konnten. Somit brauchte eine Testperson pro Termin maximal 30 Minuten, bestehend aus jeweils zwei Minuten Vorbereitungs- und acht Minuten Bearbeitungszeit pro Aufgabe.

Nach erfolgreicher Bearbeitung der drei Aufgaben wurden die Testpersonen entlassen und die nächsten zwei in die Räumlichkeiten gebeten. Im ersten Erhebungsteil vom 17.11.-07.12.2017 durchliefen die Testpersonen die drei Aufgaben immer in der gleichen Reihenfolge. Im zweiten Erhebungsteil vom 11.12.-21.12.2017 fanden die zwei Rollenspielaufgaben als erstes und letztes Element jeweils pro Raum im Wechsel statt, da nur ein Schauspieler*in pro Rollenspiel zur Verfügung stand und dieser nicht zeitgleich in beiden Räumlichkeiten agieren konnte. Ebenso erfolgte die Zuordnung von Testperson zu Testleiter per Zufall beziehungsweise Verfügbarkeit der verschiedenen Testleiter pro Erhebungstag und -uhrzeit. Somit konnte eine Testperson die beiden Erhebungstermine bei unterschiedlichen Testleitern absolviert haben oder zu beiden Terminen denselben Testleiter antreffen.

Bereits vor Ort erfolgte nach Abschluss eines Termins für jede Testperson eine globale Einschätzung durch den Testleiter, welche sich sowohl auf die subjektive Bewertung der Leistung zu jeder MMI-Aufgabe bezieht, also auch generelle Bewertungen zu Ausdruck, Aufgabenverständnis und Auftreten der jeweiligen Testperson umfasst. Diese Globaleinschätzungen werden in Kapitel 4.2.2 ausführlich dargestellt. Um die Auswertung der insgesamt sechs Aufgaben aus zeitlichen Gründen nach den Erhebungen stattfinden lassen zu können, wurden die Äußerungen der Testpersonen unter vorheriger Abklärung des Einverständnisses bei monologischen Aufgabenformaten mit einem Diktiergerät und bei Rollenspielaufgaben mit einer Videokamera aufgezeichnet. Jeder Test-

person wurde hierzu ein fester Testpersonencode zugewiesen, um die insgesamt sechs Aufnahmen im Zuge der Auswertungen einer Person zuordnen zu können. Abbildung 3 verdeutlicht noch einmal den Ablauf der MMI-Aufgaben in beiden Erhebungsteilen.



Abbildung 3. Ablauf der MMI-Aufgaben zu beiden Erhebungsteilen.

4.2 Erhebungsverfahren

Für die Erhebungen kamen insgesamt sechs MM-Aufgaben zum Einsatz, die unterschiedliche lehrberufsrelevante Kompetenzfacetten zu messen beabsichtigen. Wie bereits erwähnt wurde, können in der vorliegenden Arbeit lediglich drei Aufgaben auf ihre psychometrische Qualität hin untersucht werden. Daher sollen im Folgenden auch nur die drei MMI-Aufgaben umfassend dargestellt werden, welche auch in die Auswertungen eingegangen sind. Für jede Aufgabe wird das fokussierte Messkonstrukt theoretisch begründet und die Aufgabenbeschaffenheit detailliert dargestellt. Da es sich bei den produzierten Daten der MMI-Aufgaben um Verhaltensaufzeichnungen handelt, findet die Auswertung über entsprechende Beurteilungsbögen statt, welche das beobachtbare Verhalten durch eine festgelegte Anzahl von Variablen erfassen.

Zusätzlich zu den MMI-Aufgaben kamen Globaleinschätzungen zum Auftreten und den Leistungen der Testpersonen bezüglich der Erhebungstermine zum Einsatz. Außerdem nahmen die

Testpersonen bereits vor dem Durchlaufen der MMI-Aufgaben Selbsteinschätzungen anhand eines Online-Fragebogen vor, über den ebenfalls demographische Angaben zu den Teilnehmern gewonnen werden konnten. Im Folgenden werden die Multiplen Mini-Interviews, die Globaleinschätzungen und der Online-Fragebogen vorgestellt.

4.2.1 Multiple Mini-Interviews

Aus den sechs MMI-Aufgaben wurden die drei folgenden Aufgaben ausgewählt: *Berufswahl*, *Bildbeschreibung* und *Ausflug in den Zoo*. Diese drei ausgewählten MMI-Aufgaben stellen in ihrer Zusammensetzung das Aufgabenset des ersten Erhebungstermin der zweigeteilten Gesamterhebung dar (siehe Kapitel 4.1, Abbildung 3). Tabelle 2 gibt einen Überblick über die Gesamtheit der sechs durchgeführten MMI-Aufgaben und ordnet den einzelnen Formaten die jeweils fokussierte Kompetenzfacette zu, so dass die in diesem Ansatz berücksichtigten lehrberufsspezifischen Kompetenzen deutlich werden.

Tabelle 2

Durchgeführte Aufgabentypen der MMI-Erhebung im Winter 2017

Aufgabentitel	Erfasste Kompetenz	Aufgabentyp	Kurzbeschreibung
Berufswahl	Berufswahlkompetenz	Monolog	Erläuterungen zu Motiven der Lehrberufswahl, der persönlichen Eignung, dem Umgang mit Herausforderungen sowie berufliche Alternativen
Bildbeschreibung	Instruktionsklarheit, Erklärungskompetenz	Monolog	Beschreibung eines komplexen Bildes, so dass jemand dieses nachzeichnen könnte
Ausflug in den Zoo	Planungskompetenz, Pädagogisches Grundverständnis	Monolog	Planung und Strukturierung eines jahrgangsübergreifenden Ausflugs in den Zoo
Kollegengespräch	Konfliktfähigkeit, Problemlösung	Rollenspiel	Gespräch mit einem in der Zusammenarbeit unzuverlässigen Kollegen, gemeinsame Lösung des Problems
Forenbeitrag	Sachliche Konfliktlösung	Monolog	Sachliche Stellungnahme zu einer polemisch geschriebenen Kritik des Lehrberufs
Hausmeister	Gelassenheit, Geduld, Stressresistenz	Rollenspiel	Ausleihe eines CD-Players unter Zeitdruck bei einem sehr gemächlichen Hausmeister

Ausflug in den Zoo

Die Aufgabe zielt darauf ab, die planerischen Kompetenzen sowie das pädagogisch-erzieherische Grundverständnis von Lehramtsstudierenden über eine verhaltensorientierte Aufgabenstellung messbar zu machen. Es handelt sich bei dieser MMI-Aufgabe um ein monologisches Aufgabenformat, so dass die Testperson ihre Ausführungen selbstständig und ohne Interaktion mit dem Testleiter darstellen soll. Sie basiert dabei auf Annahmen zur Methodenkompetenz nach Frey (2004, 2008), welche die Analysefähigkeit und das planmäßige Vorgehen in Bezug auf die Organisation von Arbeitsprozessen als wichtige Elemente einer generellen Handlungskompetenz erachten.

Die Aufgabe entspricht ebenfalls den von der Kultusministerkompetenz festgelegten Standards für die Lehrerbildung, welche unter anderem die Erwartung an die Lehrkraft formulieren, das Lehren und Lernen sinnvoll zu planen und zu organisieren (KMK, 2004). Als „Fachleute für das Lehren und Lernen“ (KMK, 2004, S. 4) sollen Lehrkräfte ihre Kenntnisse über methodische Konzepte, dem Umgang mit der Klasse sowie allgemeiner Lernprozesse für eine sorgfältige und zunächst fachunabhängige Konzeption von Unterricht einsetzen. Dies legt nahe, dass planerische und organisatorische Kompetenzen ebenso mit dem pädagogischen Handlungsrepertoire verknüpft sind, um als Resultat gelingenden Unterricht zu gestalten.

Zu diesem pädagogischen Kompetenzbereich zählen unter anderem methodische Kenntnisse der Unterrichtsplanung, der effizienten Klassenführung und der Strukturierung von Lerngelegenheiten (Bauer, 1998; Baumert & Kunter, 2011; König & Blömeke, 2009; Voss, Kunina-Habenicht, Hoehne & Kunter, 2015). Aus diesem Grunde erfasst die MMI-Situation ebenso Kompetenzen aus dem pädagogisch-erzieherischem Handlungsspektrum der Lehr-Lern-Gestaltung. Die Aufgabenstellung zielt insbesondere auf die Verwendung unterschiedlicher Methoden, Sozialformen und differenzierter Lernangebote ab, welche in der Literatur als wichtige Bestandteile des pädagogischen Kompetenzbereichs genannt werden (Baumert & Kunter, 2011; Göppel, 2013; König & Blömeke, 2009; Preuß, 2008; Voss & Kunter, 2011). Ebenso sollen erzieherische Handlungskompetenzen in Form von Klassenführung und Störungsprävention durch die Situation sichtbar werden, da Lehrkräfte vor allem im Bereich des Classroom Managements durch die Schaffung von Regeln und sozialen Strukturierungen aktive Lernzeit für die Schüler garantieren (Bauer, 1998; Göppel, 2013; König & Blömeke, 2009; König & Seifert, 2012; Voss et al., 2015).

Um diese planerischen und grundlegend pädagogischen Kompetenzen erfassen zu können, werden die Testpersonen in dieser MMI-Aufgabe gebeten, einen jahrgangsübergreifenden Ausflug in den Zoo zu gestalten. Als Besonderheit der Ausflugsplanung sieht die Aufgabenstellung vor, dass die Testperson die Rolle der Lehrkraft der zweiten Klasse einnimmt und gemeinsam mit der vierten

Klasse einen Zooausflug unternehmen möchte. Als Rahmeninformationen sind die jeweiligen Unterrichtsthematiken der beiden Klassen genannt, sowie der Hinweis darauf, dass die vierte Klasse bereits Referate zu ihrem Thema vorbereitet hat. Es erfolgt außerdem eine Nennung der Anfangs- und Endzeit des Ausflugs sowie die Bereitstellung einer Kopie des Zoo-Plans als Hilfsmittel. Die Aufgabe ist so gestellt, dass die Testperson stichpunktartig die wichtigsten Gesichtspunkte zum Inhalt, den Methoden und der zeitlichen Organisation für die Gestaltung des Ausflugs erläutern soll. Mit der Wahl eines jahrgangsübergreifenden Ausflugs soll die Testperson dazu angeregt werden, neben einer allgemein methodischen Unterrichtsplanung auch Maßnahmen zur Differenzierung zu erwähnen. Ein exkursionsartiger Ausflug spricht außerdem in stärkerem Ausmaß planerische und organisatorische Kompetenzen an, denn so findet er als räumliche Auslagerung vom Klassenraum statt und bedarf einer Vielzahl an vorherigen Planungsschritten.

Die Testperson hat zunächst zwei Minuten Zeit, sich die Aufgabenstellung durchzulesen und erste Überlegungen anzustellen. Anschließend wird sie gebeten, ihre Ideen innerhalb von acht Minuten monologisch zu präsentieren. Aufgrund der Komplexität der Aufgabe und der begrenzten Bearbeitungszeit von acht Minuten ist es möglich, dass Testpersonen ihrer Nervosität geschuldet nicht auf alle relevanten Gesichtspunkte eingehen können. Aus diesem Anlass wurden insgesamt neun Hinweisfragen erstellt, welche der Testleiter bei ausbleibender Bearbeitung von grundlegenden Rubriken wie der organisatorischen Planung, dem pädagogischen Umgang oder der methodischen Gestaltung gezielt äußern soll. Die komplette Aufgabenbeschreibung sowie die Auflistung aller Hinweisfragen sind im Anhang A und B zu finden.

Zur Auswertung der MMI-Aufgabe *Ausflug in den Zoo* erfolgt eine Einschätzung über einen Beurteilungsbogen, welcher über die Dimensionen *Strukturierung von Lerngelegenheiten*, *pädagogisch-erzieherisches Grundverständnis* und *Planungskompetenz* die Leistungen der Testperson bewertet. Jede der insgesamt drei Dimensionen wird über zwei Items auf einer fünfstufigen graphischen Antwortskala erhoben, so dass der Beurteilungsbogen aus insgesamt sechs Items besteht. Die Dimension *Strukturierung von Lerngelegenheiten* erfasst über zwei Items zum einen Methoden und Integration in eine Unterrichtseinheit und zum anderen die innere Differenzierung. Die Gruppenorganisation durch soziale Strukturierung sowie die allgemeine Klassenführung wird in der Dimension *pädagogisch-erzieherisches Grundverständnis* erfasst. Die Dimension der *Planungskompetenz* berücksichtigt die Zeit- und Ablaufplanung sowie die übergeordnete Ausflugsorganisation. Die Skalen sind so aufgebaut, dass schwache Leistungen mit niedrigen Werten und starke Leistungen mit hohen Werten bepunktet werden. Der niedrigste Wert liegt bei 1 und der höchste Wert liegt bei 5. Es kann ein Gesamtskalenwert errechnet werden, indem die Rohwerte der jeweils sechs Items aufaddiert werden. Der maximal erreichbare Wert der Gesamtskala liegt folglich bei 30. Die

Anzahl der Hinweisfragen wird zwar im Beurteilungsbogen vermerkt, geht aber zunächst nicht in die Bewertung mit ein. Der vollständige Beurteilungsbogen der Situation findet sich im Anhang C.

Berufswahl

Die MMI-Situation *Berufswahl* erfasst die motivationalen Orientierungen, welche im Zuge der Berufswahlentscheidung von Bedeutung sind. Es ist grundsätzlich anzunehmen, dass sich Lehramtsstudierende aus unterschiedlichen Motiven heraus für den Beruf des Lehrers entscheiden. Ob sich dabei nur jedoch tatsächlich geeignete Kandidaten des Lehrberufs für das Studium entscheiden, bleibt eine umstrittene Frage (Affolter et al., 2015). Aufgrund einer in Deutschland fehlenden offiziellen Eignungsabklärung zu Beginn des Lehrberufs, ist der Zugang zu einem Lehramtsstudium grundsätzlich jeder Person mit entsprechendem Schulabschluss möglich (Mayr, 2012; Rothland & Terhart, 2011). Dennoch erweist sich die Handlungskompetenz einer Lehrperson zugleich von motivationalen Orientierungen und Überzeugungen abhängig und wird in der Fachliteratur als eine der vier wesentlichen Komponenten der professionellen Lehrerkompetenz angesehen (Baumert & Kunter, 2006).

Berufswahlmotive bestimmen also nicht nur die Entscheidung für den Lehrberuf, sondern üben darüber hinaus einen nicht zu unterschätzenden Einfluss auf die Entwicklung der professionellen Lehrerrolle, die grundlegende Unterrichtsgestaltung sowie dem Engagement im Berufsalltag aus (Thomson, Turner & Nietfeld, 2012). Die Berufswahlmotivation kann etwa durch das primäre Vorherrschen von unrealistischen und übersteigerten Vorstellungen des Berufsalltags dazu führen, dass idealistische Erwartungen schnell in Überlastungen umschlagen (Schaarschmidt & Kieschke, 2007). Die Bedeutung von Berufswahlmotiven lässt sich auch im Sinne der Gesundheitsprävention nicht leugnen (Rothland, 2013).

Die Struktur von Berufswahlmotiven im Lehrberuf ist in der Regel von unterschiedlich ausgeprägten intrinsischen und extrinsischen Werten geformt. Die häufigsten Gründe für die Wahl des Lehramtsberufs lassen sich intrinsischen Motivlagen zuordnen, hierzu zählen Freude und das Interesse an der Arbeit mit Kindern und die mögliche Einflussnahme auf die kindliche Entwicklung (Richardson & Watt, 2006; Watt & Richardson, 2007). Intrinsisch motivierte Personen verbleiben länger im Beruf und berichten von stärker ausgeprägten Erfolgen beim Unterrichten (Mayr, 2009). Extrinsische Motive fokussieren die allgemeinen Vorzüge des Berufs wie Berufssicherheit, geregeltes Einkommen, Urlaubsmöglichkeiten und Zeit für die Familie (Richardson & Watt, 2006; Watt & Richardson, 2007). Berufswahlmotive beziehen darüber hinaus auch Überzeugungen der eigenen Fähigkeiten und Einschätzungen des Anspruchs der beruflichen Tätigkeit mit ein (Richardson &

Watt, 2006). Die Berufswahl besteht also nicht nur aus einer vereinzelt Motivausprägung, sondern lässt sich eher als Bündel verschiedener Überzeugungen und Orientierungen verstehen (Affolter et al., 2015). Häufig werden sie in übergeordnete Motivtypen kategorisiert (z.B. Affolter et al., 2015; Thomson et al., 2012).

Die MMI-Situation *Berufswahl* fordert die Testpersonen auf, sich anhand von vier verschiedenen Fragen zu ihrer Berufswahl zu äußern. Sie beziehen sich auf das Interesse für den Lehrberuf, die möglichen Schwierigkeiten und der geplante Umgang mit diesen, eignungsrelevante Eigenschaften sowie möglichen beruflichen Alternativen. Die Testpersonen haben zwei Minuten Zeit, sich zu den vier Fragen erste Überlegungen anzustellen und werden anschließend gebeten, diese innerhalb von acht Minuten darzulegen. Damit der Prozess so standardisiert wie möglich abläuft, werden die Fragen nacheinander durch Testleiter gestellt. Die Bearbeitungszeit beginnt demnach mit der Stellung der ersten Frage, welche die Testpersonen ohne weitere Interaktion mit dem Testleiter beantworten sollen. Sofern sich die Testperson nicht mehr zu einer Frage äußert, leitet der Testleiter durch Vorlesen der darauffolgenden Frage den nächsten Abschnitt ein, bis die letzte Frage gestellt und beantwortet wurde. Der originale Aufgabentext zur Aufgabe *Berufswahl* befindet sich im Anhang D.

Zur Auswertung der MMI-Aufgabe wird ein Beurteilungsbogen herangezogen, mit Hilfe dessen die Äußerungen der Testperson pro gestellte Frage und hinsichtlich des gesamten Gesprächs bewertet werden. Dies erfolgt für die ersten drei gestellten Fragen auf jeweils zwei Items, hinsichtlich der letzten Frage zu den beruflichen Alternativen nur auf einem Item und in Bezug auf das gesamte Gespräch über drei globale Items. Der Beurteilungsbogen besteht somit aus zehn Items, welche in fünf Dimensionen untergliedert sind. Die Auswertung der fragenbezogenen Äußerungen werden grundlegend in Bezug auf die Erfüllung der erwarteten Standards auf einer fünfstufigen Antwortskala vorgenommen, mit dem Skalenpunkt 0 als niedrigsten Wert („liegt weit ab vom erwarteten Standard (Erfragtes wird nicht genannt)“) und dem Skalenpunkt 4 als höchsten Wert („liegt über dem erwarteten Standard (Erfragtes wird genannt und erläutert und durch Belege begründet)“).

In Anlehnung an die von Richardson und Watt (2006) vorgenommene Unterteilung in intrinsische und extrinsische Motive der Berufswahl, erfassen zwei Items zur Frage nach dem Interesse beziehungsweise Reiz am Lehrberuf die Nennung von intrinsischen oder extrinsischen Motiven. Die Frage nach den Herausforderungen und Schwierigkeiten und Umgang mit diesen wird über zwei Items zur Ausführlichkeit und faktischen Begründung der Nennungen eingeschätzt. Bei der Einschätzung der günstigen und ungünstigen Eigenschaften für den Lehrberuf fließt ebenfalls in die Wertung mit ein, ob die Testpersonen die Bedeutung der Eigenschaften in Bezug auf den Lehr-

beruf darlegen. Äußerungen zu beruflichen Alternativen werden in Abhängigkeit ihres Umfangs und des Realitätsbezuges beurteilt. Der Beurteilungsbogen zur *Berufswahl* nimmt außerdem über drei Items zur Qualität der argumentativen Darlegung, der Realitätsnähe und der intellektuellen Durchdringung eine übergeordnete Einschätzung des gesamten Gespräches vor. Der vollständige Beurteilungsbogen mit allen Items und den spezifischen Skalenabstufungen befindet sich im Anhang E.

Bildbeschreibung

Die MMI-Aufgabe *Bildbeschreibung* legt das Augenmerk auf die Instruktionsfähigkeiten der Lehrkraft und fordert die Testpersonen heraus, ein komplexes geometrisches Bild so zu beschreiben, dass eine andere Person dieses problemlos nachzeichnen kann. Trotz des vorherrschenden didaktischen Ideals, Lernprozesse der Schüler so oft wie möglich durch selbstgesteuerte Lerngelegenheiten anzuregen, muss das lehrergeleitete Erklären und Instruieren dennoch als ein wichtiger Kompetenzbereich jeder Lehrkraft angesehen werden (Aeschbacher, 2009). Durch das Zurückgreifen auf die eigene Fachkompetenz zielen Lehrkräfte in ihrem Unterricht darauf ab, grundlegende Inhalte und komplexe Themen so darzustellen, dass sie von den Lernenden verstanden und verinnerlicht werden können. Für viele Praktiker ist diese Aufbereitung der Lerninhalte, so dass sie aus den Perspektiven der Schüler korrekt nachvollzogen werden können, eine entscheidende Kernaufgabe des Lehrens (Weiß et al., 2014).

In der Forschung wurden Instruktionen und Erklärungen im unterrichtlichen Kontext dahingehend untersucht, welche Faktoren ihre Effektivität beeinflussen. Wittwer und Renkl (2008) sichten eine Vielzahl von Forschungsergebnissen zu Instruktionen im Lehrkontext und identifizierten eine Reihe von Eigenschaften, die für eine bessere Qualität von Erklärungen sorgen können. Sie betonen beispielsweise die Notwendigkeit, Instruktionen mit dem Vorwissen der Lernenden zu verknüpfen und empfehlen, beim Erklären wichtige Konzepte und Prinzipien miteinzubeziehen (vgl. Lipowsky, 2007). Instruktionen sollen demnach weder zu leicht ausfallen, da sonst die kognitive Kapazität der Lernenden für das Verarbeiten der wenig anspruchsvollen Informationen belegt wird, noch zu Überforderungen seitens der Lernenden führen. Durch Vereinfachungen und Abstraktionen können komplexe Sachverhalte verständlich auf relevante Themen reduziert werden (Aeschbacher, 2009; Lipowsky, 2007). Beispiele zu lebensweltlichen Bezügen und gelegentliche Wiederholungen befähigen die Schüler dazu, neue Informationen erfolgreich in Konzepte und Vorwissen einzubetten (Wittwer & Renkl, 2008).

Eine Lehrkraft muss in ihren Erklärungen stets das Verständnis und die Ausgangslage der Schüler berücksichtigen. Dies gelingt nur durch das Hineinversetzen in die Denkweisen der Lernen-

den und die darauf aufbauende Fähigkeit, Erklärungen auf vielfältige Art an die unterschiedlichsten Verständnisprozesse anpassen zu können (Weiß et al., 2014). Auf dieser Grundlage kann der Unterricht durch Elemente des Instruierens und Erklärens bereichert werden, sofern diese „interessant, prägnant, klar, verständlich, gut gegliedert und vernetzt präsentiert“ sind (Lipowsky, 2007, S. 28). Genau diese Aspekte einer gelingenden und klaren Instruktion erfasst die MMI-Aufgabe *Bildbeschreibung*. Die Testpersonen erhalten auf einem DIN A4 Blatt die Abbildung eines geometrischen Bildes und werden gebeten, ein Audio-Tutorial für Schüler der Oberstufe zu erstellen. Hierbei soll der Nutzer des Audio-Tutorials dazu angeleitet werden, das Bild in seiner Komplexität nachzuzeichnen. In der Aufgabenbeschreibung wird bereits das Ziel des Audio-Tutorials erwähnt, nämlich komplizierte Sachverhalte so zu erklären, dass sie jeder verstehen kann. Zur Vorbereitung auf die Aufgabe erhalten die Testpersonen zwei Minuten Zeit, die Bearbeitung der Aufgabe nimmt im Anschluss acht Minuten Zeit in Anspruch. Da nicht zu erwarten ist, dass in dieser kurzen Zeit ein solch vernetztes Bild vollständig beschrieben werden kann, erhalten die Testpersonen zuvor einen entsprechenden Hinweis. Das genaue Bild sowie die Aufgabenstellung findet sich in Anhang F.

Die Auswertung dieser MMI-Aufgabe erfolgt über einen Beurteilungsbogen bestehend aus drei Dimensionen mit jeweils zwei Items, so dass der gesamte Bogen die Leistung der Testpersonen insgesamt über sechs Items beurteilt. Die Beurteilung wird über ein fünfstufiges graphisches Skalenformat vorgenommen, indem schlechte Leistungen mit niedrigen Werten und gute Leistungen mit hohen Werten einhergehen. Das Minimum der Skala stellt der Wert 1, das Maximum der Wert 5 dar. Unter der Dimension *Kommunikative Perspektive* wird über ein Item die adressatenbezogene Sprechweise erfasst, sowie durch ein zweites Item die Vermittlung eines Orientierungsrahmens beurteilt. Die Klarheit der Darstellung und die fachliche Präzision der Begriffe werden durch jeweils ein Item in der Dimension *Begriffliche Eindeutigkeit* berücksichtigt. Die dritte Dimension *Ordnung und Folgerichtigkeit* beurteilt die Leistungen hinsichtlich ihrer strukturierten Vorgehensweise und einer hierarchischen Gliederung. Für die Aufgabe lässt sich ein Gesamtskalenwert errechnen, indem die einzelnen Werte der sechs Items addiert werden. Insgesamt ist ein maximaler Gesamtskalenwert von 30 zu erreichen. Der vollständige Beurteilungsbogen ist im Anhang G nachzusehen.

4.2.2 Globaleinschätzungen

Bereits vor Ort und unmittelbar nach jeder Absolvierung der drei MMI-Aufgaben des Erhebungs-termins erfolgten globale Einschätzungen der Leistung jeder Testperson. Diese sind zum einen aufgabenunspezifisch und betreffen eher das allgemeine Auftreten und den Umgang mit der Gesamtsituation, zum anderen beziehen sie sich auf die Leistung in der spezifischen MMI-Aufgabe. Zum Schluss sollte der Testleiter zudem ein subjektives globales Urteil über den Gesamteindruck der

Testperson abgeben. Somit liegen für jede Testperson zwei vollständige Sätze an Globaleinschätzungen vor, da zu jedem Termin ein solcher Bogen durch den entsprechenden Testleiter ausgefüllt wurde.

Für den ersten Testtermin beziehen sich die aufgabenspezifischen Globaleinschätzungen dabei auf die Leistungen in den MMI-Aufgaben *Berufswahl*, *Bildbeschreibung* und *Ausflug in den Zoo*, für den zweiten Testtermin erfolgte die aufgabenspezifische Einschätzung entsprechend für die Aufgaben *Kollegengespräch*, *Forenbeitrag* und *Hausmeister*. Die nicht aufgabengebundenen Einschätzungen blieben beim ersten und zweiten Termin identisch. Da jedoch die Testpersonen zu den zwei Erhebungsterminen nicht immer durch dieselben Testleiter betreut wurden, können die zwei Globaleinschätzungssätze pro Testperson durchaus von zwei unterschiedlichen Testleitern angefertigt worden sein. Diese Arbeit fokussiert sich nur auf drei MMI-Aufgaben des ersten Erhebungszeitraums, so dass nur die Globaleinschätzungen des ersten Termins miteinbezogen werden.

Für die aufgabenspezifischen Einschätzungen sollten die Testleiter für jede Testperson auf einer fünfstufigen Skala für jede einzelne MMI-Aufgaben eine sofortige Bewertung ihrer Leistung anfertigen, diese reichte von „eher misslungen“ bis „sehr gut gelungen“. Es handelt sich bei den Skalen zur Einschätzung der MMI-Aufgaben um positiv verschobene Skalen, so dass der Mittelpunkt beim Skalenwert 3 bereits eine gute Leistung („gut gelungen“) widerspiegelt. Die Bearbeitung von jeweils drei Aufgaben pro Erhebungstermin wurden zusammengekommen hinsichtlich der allgemeinen Qualität der Argumentation (von „sprunghaft, einseitig“ bis „strukturiert, begründet“), des sprachlichen Ausdrucks (von „unklar, Mundart“ bis „prägnant, komplex“) sowie der Aufgeregtheit (von „angespannt, nervös“ bis „ruhig, entspannt“) der Testperson eingeschätzt.

Die aufgabenunspezifischen Bewertungen erfassen situations- und personenbezogenen Eigenschaften pro Erhebungstermin. Der Testleiter trifft hierfür ebenfalls auf einer fünfstufigen Skala Aussagen über die Umgänglichkeit/Zugewandtheit (von „distanziert, spröde“ bis „kontaktfreudig“), die Auffassungsgabe und das Aufgabenverständnis („von langsam, beschränkt“ bis „rasch, umfassend“), die Motiviertheit und das Interesse (von „träge, gelangweilt“ bis „eifrig, engagiert“) sowie über die Höflichkeit und das generelle äußere Auftreten (von „flapsig, nachlässig“ bis „gesittet, seriös“). Auf einer positiv verschobenen Skala hielten die Testleiter pro Termin schlussendlich ihren Gesamteindruck von der Leistung und dem Auftreten der Testperson fest, dieser reicht von „eher negativ“ bis „sehr positiv“, wobei „eher positiv“ bereits eine mittlere Bewertung darstellt. Die vollständigen Items der Globaleinschätzungen mit den entsprechenden Skalenabstufungen befinden sich im Anhang H.

4.2.3 Online-Fragebogen

Bevor die Testpersonen die einzelnen MMI-Aufgaben durchliefen, wurden sie gebeten, zunächst einen Online-Fragebogen auszufüllen. Dieser erfasst grundlegende demographische Informationen zum Alter, dem Geschlecht sowie Angaben zum Lehramtsstudium (Bachelor/Master, Fachsemester, Schulart und Unterrichtsfächer). Im Anschluss folgen Einschätzungen der eigenen Persönlichkeit sowie verschiedene Selbsteinschätzungen.

Die Persönlichkeitseinschätzung besteht aus einer Auflistung von neun Gegensatzpaaren unterschiedlicher Eigenschaften, wobei jedes Adjektiv des Gegensatzpaares einen Pol der Einschätzungen darstellt. Die Testpersonen sollen bewerten, welche der zwei Eigenschaften pro Wortpaar jeweils besser auf sie zutreffen. Zur Verfügung standen den Testpersonen dabei insgesamt sechs Skalenabstufungen pro Wortpaar, wovon sich jeweils eine Hälfte („völlig“, „viel eher“, „eher“) auf das jeweilige Eigenschaftswort als Pol der Skala bezieht. Jeweils drei der neun Items lassen sich zu einem der drei Persönlichkeitsfaktoren Extraversion, Verträglichkeit und emotionaler Stabilität zuordnen, welche dem Big-Five Modell entstammen (siehe Costa & McCrae, 1999). Die Items „gesprächig/schweigsam“, „kontaktfreudig/zurückhaltend“ und „anschlussbedürftig/eingelängert“ sollen Extraversion messen, die Items „gutmütig/reizbar“, „nachsichtig/barsch“ und „selbstlos/selbstsüchtig“ zählen zum Faktor Verträglichkeit und die drei Items „robust/verletzlich“, „selbstzufrieden/selbstmitleidig“ und „entspannt/überempfindlich“ messen die emotionale Stabilität.

Die Selbsteinschätzungen werden mit dem Titel „Wie sehen Sie sich selbst?“ eingeleitet und stellen eine Liste aus achtzehn Aussagen zu Einschätzungen des eigenen Erlebens und Handelns dar. Auf einer fünfstufigen Skala sollen die Testpersonen den Grad ihrer Zustimmung bezogen auf das Zutreffen von „stimme ganz und gar nicht zu“ bis „stimme voll und ganz zu“ angeben. Die achtzehn Items lassen sich sechs Fähigkeits- und Kompetenzbereichen zuordnen und beziehen sich auf die Berufswahl-, Erklärungs- und Planungskompetenz sowie das allgemeine Leistungsstreben, Stressbewältigung und soziale Kompetenz. Diese Fähigkeitsbereiche sind wiederum für die sechs MMI-Aufgaben jeweils von spezifischer Relevanz. Für die Aufgabe *Ausflug in den Zoo* spielen planerische Kompetenzen eine Rolle, für die Aufgabe *Bildbeschreibung* sind vor allem Kompetenzen des Erklärens von Bedeutung und die Aufgabe *Berufswahl* verlangt Berufswahlkompetenz. Die vollständige Version mit allen einzelnen Wortpaaren der Persönlichkeitseinschätzungen und den achtzehn Aussagen zu den Selbsteinschätzungen können im Anhang I eingesehen werden.

4.3 Stichprobe

Die Stichprobe für diese Arbeit besteht aus insgesamt $N = 50$ Lehramtsstudierenden der Universität Erfurt. Hiervon sind 47 Personen weiblichen Geschlechts und zwei Personen männlichen Geschlechts, bei einer Testperson fehlt die Angabe zum Geschlecht. Die Studierenden waren durchschnittlich 23.04 Jahre alt ($SD = 2.43$) mit einer Altersspanne von 19 bis 30. Bei den 50 Testpersonen handelt es sich um fast ausschließlich Studierende im Master ($n = 47$), nur drei Testpersonen haben angegeben, dass sie im Bachelor studieren. Von den drei Bachelorstudierenden befinden sich zwei Personen im ersten Fachsemester, und eine Person im siebten Fachsemester. Die Fachsemester der Masterstudierenden reichen vom ersten bis zum elften Semester. Insgesamt befinden sich 29 der 47 Masterstudierenden im siebten Fachsemester, gefolgt von fünfzehn Personen im ersten Fachsemester und jeweils einer Person im dritten, neunten und elften Fachsemester. Mehrheitlich haben sich die 47 Masterstudierenden mit 85.1 Prozent auf die Grundschule spezialisiert ($n = 40$), 10.6 Prozent studieren für das Lehramt der Regelschule ($n = 5$) und nur jeweils 2.1 Prozent gaben die Schulart Förderschule ($n = 1$) und berufsbildende Schulen ($n = 1$) an.

Globaleinschätzungen durch die Testleiter

Globaleinschätzungen des allgemeinen Auftretens und bezüglich der Bewältigung der drei MMI-Aufgaben durch die Testleiter liegen nur für einen Teil der Stichprobe von $N = 47$ vor. Betrachtet man zunächst die aufgabenspezifischen Einschätzungen der Testleiter für jede MMI-Aufgabe, so schätzten die Testleiter die Leistungen der Testpersonen in der ersten MMI-Aufgabe *Berufswahl* im Mittel am höchsten ein ($M = 3.51$, $SD = .98$), darauf folgt die mittlere Einschätzung der Bewältigung der *Zooaufgabe* ($M = 3.43$, $SD = 1.25$) und am niedrigsten fällt die mittlere Bewertung der Leistung in der Aufgabe *Bildbeschreibung* aus ($M = 2.83$, $SD = 1.17$). In Tabelle 3 sind zudem die minimalen und maximalen Bewertungen der jeweiligen Bewältigung der drei MMI-Aufgaben abgetragen, bei der *Bildbeschreibung* und dem *Ausflug in den Zoo* schöpfen die Bewertungen die Skalenspannweite voll aus, nur bei der *Berufswahl* wurde als niedrigster Wert insgesamt eine 2 anstelle einer 1 vergeben.

Tabelle 3

Globaleinschätzungen der Testleiter bezüglich der Bewältigung der drei MMI-Aufgaben

MMI-Aufgabe	<i>M</i>	<i>SD</i>	Minimum	Maximum
<i>Berufswahl</i>	3.51	0.98	2.00	5.00
<i>Bildbeschreibung</i>	2.83	1.17	1.00	5.00
<i>Ausflug in den Zoo</i>	3.43	1.25	1.00	5.00

Anmerkungen. Der theoretische Skalenwertebereich für die Einschätzungen der drei MMI-Aufgaben erstreckt sich von 1 bis 5.

Betrachtet man die allgemeinen Einschätzungen des Auftretens, so wurde generell die Höflichkeit der Testpersonen am höchsten ($M = 4.32$, $SD = .73$) und die Aufgeregtheit am niedrigsten eingeschätzt ($M = 3.36$, $SD = 1.24$). Die Umgänglichkeit, Auffassungsgabe und Motiviertheit sowie die Qualität der Argumentation und des sprachlichen Ausdrucks erhielten mittelhohe Einschätzungen von $M = 3.45$ bis $M = 3.89$. Der globale Eindruck von den Testpersonen liegt im Mittel bei $M = 3.51$ ($SD = .91$). Tabelle 4 zeigt alle Mittelwerte, Standardabweichungen und entsprechende Minimum- und Maximumwerte für jedes Einschätzungsitem.

Tabelle 4

Aufgabenunspezifische Globaleinschätzungen der Testleiter

Item	<i>M</i>	<i>SD</i>	Minimum	Maximum
Umgänglichkeit/Zugewandtheit	3.89	1.11	1.00	5.00
Auffassungsgabe/ Aufgabenverständnis	3.85	1.10	1.00	5.00
Motiviertheit/Interesse	3.77	1.20	1.00	5.00
Höflichkeit/Auftreten	4.32	0.73	2.00	5.00
Qualität der Argumentation	3.45	1.06	1.00	5.00
Qualität des sprachlichen Ausdrucks	3.62	0.92	1.00	5.00
Aufgeregtheit	3.36	1.23	1.00	5.00
Eindruck insgesamt	3.51	0.91	1.00	5.00

Anmerkungen. Der theoretische Skalenwertebereich für die Einschätzungen der drei MMI-Aufgaben erstreckt sich von 1 bis 5.

Persönlichkeits- und Selbsteinschätzungen der Testpersonen

Hinsichtlich der Persönlichkeitsmerkmale der Testpersonen ergibt sich folgendes Bild: Die Ausprägung von Extraversion beträgt in der vollständigen Stichprobe von $N = 50$ im Mittel 3.24 ($SD = 1.48$), die mittlere Ausprägung von Verträglichkeit $M = 2.48$ ($SD = 1.28$) und von emotionaler Stabilität $M = 3.16$ ($SD = 1.28$). Das Minimum des theoretischen Wertebereichs liegt bei 1.00 und das Maximum bei 7.00. Betrachtet man die Selbsteinschätzungen der Testpersonen, so wird das eigene Leistungsstreben mit einem Mittelwert von 3.86 ($SD = .65$) am höchsten eingestuft, wohingegen die eigene Planungskompetenz von den Testpersonen am niedrigsten eingeschätzt wird ($M = 3.40$, $SD = .44$). Kompetenzen des Erklärens und der Berufswahl, die soziale Kompetenz und die Stressbewältigung weisen im Mittel Einschätzungen von $M = 3.47$ bis $M = 3.74$ auf. In Tabelle 5 sind alle Selbsteinschätzungen übersichtlich aufgeführt.

Tabelle 5

Mittelwerte, Standardabweichungen sowie Minimum und Maximum für die Selbsteinschätzungen

Item	<i>M</i>	<i>SD</i>	Minimum	Maximum
Berufswahlkompetenz	3.48	.68	2.00	5.00
Erklärungskompetenz	3.59	.55	1.50	4.50
Planungskompetenz	3.40	.44	2.33	4.33
Leistungsstreben	3.86	.65	1.50	5.00
Soziale Kompetenz	3.74	.65	2.00	4.50
Stressbewältigung	3.47	.58	2.00	4.50

Anmerkungen. Der theoretische Skalenwertebereich für die Einschätzungen der drei MMI-Aufgaben erstreckt sich von 1 bis 5.

4.4 Auswertungsmethodik

Die Beurteilungen der drei MMI-Situationen *Berufswahl*, *Bildbeschreibung* und *Ausflug in den Zoo* erfolgte unter Beteiligung des Teaching Talent Centers und dem Lehrstuhl Pädagogisch-psychologische Diagnostik und Differentielle Psychologie mit einer Gesamtheit von zehn Beurteilern. Um alle Mitwirkenden mit dem Beurteilungsbogen der jeweiligen Situation vertraut zu machen und eine grundlegende Kalibrierung aller Beurteiler vorzunehmen, wurden zunächst drei zufällig ausgewählte Testpersonen aus dem Gesamtdatensatz ausgewertet und die Ergebnisse im Plenum besprochen. Erst nach Erreichen eines zufriedenstellenden Konsenses über die Handhabung der einzelnen Beurteilungsdimensionen wurde eine weitere Zufallsstichprobe von zehn Test-

personen durch alle zehn Beurteiler bewertet. Nach Berechnungen von Intraklassenkorrelation zur Analyse der Beurteilerübereinstimmungen erfolgten eventuelle Überarbeitungen der Beurteilungsbogen, um im Anschluss und als Grundlage dieser Arbeit die Leistungen von 50 Testpersonen in den drei MMI-Aufgaben durch ausgewählte Beurteiler einschätzen zu lassen. Die MMI-Aufgabe *Bildbeschreibung* wurde durch zwei Beurteiler, die Aufgabe *Berufswahl* durch zwei Beurteiler und die Aufgabe *Ausflug in den Zoo* durch drei Beurteiler anhand der aktuellsten Fassung des Beurteilungsbogens bearbeitet.

Zur Prüfung der Fragestellungen der vorliegenden Arbeit wurden die Daten anhand des statistischen Auswertungsprogrammes SPSS Version 21 von IBM und Excel 2016 von Microsoft Office analysiert. Um die Messgenauigkeit auf Basis der über die Beurteiler gemittelten Werte zu untersuchen, wurden über Trennschärfen- und Reliabilitätsanalysen entsprechende Kennwerte für jede MMI-Aufgabe ermittelt. Als Maß für die Reliabilität wird in dieser Arbeit die interne Konsistenz mittels Cronbachs Alpha wiedergegeben. Der Anspruch an gute Instrumente liegt hier mindestens bei einem Cronbachs Alpha von $\alpha_{\text{Cron}} = .80$ (Bortz & Döring, 2006; S. 199). Trennschärfen ergeben sich aus der korrigierten Item-Korrelation mit dem Gesamtwert, einer so genannten Item-Skala-Korrelation. Werte zwischen $r_{it} = .30$ und $r_{it} = .50$ gelten als mittelmäßig, Werte $r_{it} > .50$ als hoch (Bortz & Döring, 2006; S. 220).

Um für jede MMI-Aufgabe die Itemschwierigkeiten der Beurteilungsbögen zu berechnen, wurde der Schwierigkeitsindex bei mehrstufige Antworten p_m verwendet (Fisseni, 2004, S. 34). Die Items in den Beurteilungsbögen der MMI-Aufgaben *Ausflug in den Zoo* und *Bildbeschreibung* liegen in einem fünfstufigen Skalenformat mit dem Wertebereich von 1 bis 5 vor. Zur Bestimmung des Schwierigkeitsindex mussten diese Skalen zunächst in einen Wertebereich von 0 bis 4 umcodiert werden, damit die erreichten Wertpunkte pro Item zur maximalen Anzahl erreichbarer Wertpunkte ins Verhältnis gesetzt werden konnten. Das Skalenformat des Beurteilungsbogens der MMI-Aufgabe *Berufswahl* weist bereits passende Skalenwerte von 0 bis 4 auf. Von einer idealen Itemschwierigkeit spricht man bei Werten von $p_m = .50$. Aus praktischer Sicht können jedoch sowohl Items mit $p_m = .50$ sowie $.50 > p_m < .50$ vorhanden sein, um eine Differenzierungsvielfalt zu ermöglichen (Fisseni, 2004).

Zur Bestimmung der Reliabilität der Einschätzungen unterschiedlicher Beurteiler wurden Beurteilerübereinstimmungen über Intraklassenkorrelationen berechnet. Die Berechnung von Intraklassenkorrelationen ist hier zulässig, da die Verhaltensbeurteilungen anhand von Ratingskalen vorgenommen wurden und somit davon auszugehen ist, dass intervallskalierte Daten vorliegen. Die Intraklassenkorrelation ist dabei ein Maß, das die Stärke des Zusammenhangs der Urteile über dieselben Objekte von zwei oder mehreren Beurteilern angibt (Wirtz & Caspar, 2004). Die Aussagen

des *ICC*-Koeffizienten stehen dabei sowohl für die Reliabilität als auch die Objektivität der Messungen und machen Angaben für einen einzelnen Beurteiler (Stemmler & Margraf-Stiksrud, 2015). Von den drei möglichen *ICC*-Koeffizienten wird für die vorliegenden Daten der Intraklassenkorrelationskoeffizient *ICC2* errechnet, da innerhalb jeder MMI-Aufgabe alle 50 Beobachtungsobjekte, also Testpersonen, von denselben Beurteilern beurteilt worden sind. Diese Beurteiler sollen dabei eine zufällige Auswahl aus der Grundgesamtheit aller potentiellen Beurteiler darstellen. Aus diesen Gründen kann der *ICC2* als Reliabilitätsmaß genutzt werden (Stemmler & Margraf-Stiksrud, 2015; Wirtz & Caspar, 2004).

Neben Aussagen über einen einzelnen Beurteiler lassen sich anhand des *ICC(2,2)* Angaben zum mittleren Urteil berechnen. Er spielt eine Rolle, wenn die vorgenommenen Einschätzungen über alle Beurteiler gemittelt werden und diese Durchschnittswerte die Entscheidungsgrundlage darstellen (Stemmler & Margraf-Stiksrud, 2015; Wirtz & Caspar, 2004). Grundsätzlich spricht man ab einem Wert von *ICC* = 0.70 von einer akzeptablen Reliabilität (Wirtz & Caspar, 2004).

Zur Beantwortung der Fragestellungen, welche sich auf die Validität des Verfahrens beziehen, werden Faktorenanalysen nach der Hauptkomponentenanalyse mit Varimax-Rotation durchgeführt. Die Verifizierung der Stichprobeneignung für eine Faktorenanalyse erfolgt mithilfe des Kaiser-Meyer-Olkin-Kriteriums und über den Bartlett-Test auf Sphärizität. Das Maß der Stichprobeneignung sollte bezüglich des KMO-Kriteriums mindestens .5, jedoch wünschenswert einen Wert von mindestens .8 aufweisen. Der Bartlett-Test auf Sphärizität sollte signifikant ausfallen, um die Voraussetzungen für eine Faktorenanalyse zu erfüllen (Backhaus, Erichson, Plinke & Weiber, 2011, S. 342). Ab einer Ladungshöhe von .5 wird eine Variable einem Faktor zugeordnet

Zusammenhänge der über die MMI-Aufgaben erfassten Konstrukte mit den Global- und Selbsteinschätzungen werden mittels Korrelationen und multiplen Regressionen untersucht. Da die interessierenden Konstrukte intervallskaliert sind, wird eine Produkt-Moment-Korrelation verwendet. Die ermittelten Korrelationen sind je nach Höhe nach Bortz und Döring (2006) mit $r = .10$ als niedrig, $r = .30$ als mittel und mit $r = .50$ hoch bewertet. Multiple Regressionen werden errechnet, um multivariate Zusammenhänge zu überprüfen. Durch die Kombination von Prädiktorvariablen soll eine größtmögliche Varianzaufklärung in der Kriteriumsvariable erreicht werden. Für die Berechnungen erfolgt die Prüfung auf Signifikanz sowohl für das gesamte Modell als auch für die einzelnen Regressionskoeffizienten. Es wird zudem die korrigierte Varianzaufklärung der jeweiligen Regressionsanalysen angegeben.

Um Multikollinearität auszuschließen, wird zudem der Varianz-Inflations-Faktor (*VIF*) herangezogen, welcher die Grenze von *VIF* = 5 unterschreiten sollte (Urban & Mayerl, 2011; S. 232).

Um eine Alpha-Fehler-Kumulierung zu vermeiden, wird für jeden Signifikanztest das Signifikanzniveau nach der B-Y-Methode korrigiert (Narum, 2006). Um den testtheoretischen Vorgaben dieser Arbeit gerecht zu werden, müssen berichtete Ergebnisse Werte von mindestens $p < .01$ aufweisen, so dass sie als gesicherte Ergebnisse betrachtet werden können.

5 Ergebnisse

Im Folgenden werden die vorliegenden Daten der MMI-Aufgaben hinsichtlich ihrer Reliabilität und Validität untersucht. Hierfür werden bi- und multivariate Zusammenhänge zwischen den Leistungen in den MMI-Aufgaben und den Globaleinschätzungen durch die Testleiter sowie den Persönlichkeits- und Selbsteinschätzungen der Testpersonen überprüft.

5.1 Item- und Reliabilitätsanalysen

MMI-Aufgabe Berufswahl

Die Gesamtleistung der Testpersonen in der MMI-Aufgabe *Berufswahl* errechnet sich als Summe der einzelnen Punkte pro Einschätzungsitem und liegt in der vorliegenden Stichprobe von $N = 50$ bei einem Mittelwert von $M = 17.71$ ($SD = 3.98$). Grundlage für die Berechnung sind die über beide Beurteiler gemittelten Einschätzungen. Die niedrigste Leistung in dieser Aufgabe liegt bei einem Gesamtscore von 9.50, die höchste Leistung beläuft sich auf 27.50 (siehe Tabelle 6).

Tabelle 6

Gesamtleistung der Testpersonen in der MMI-Aufgabe Berufswahl

	<i>M</i>	<i>SD</i>	Minimum	Maximum
Gesamtscore <i>Berufswahl</i>	17.71	3.98	9.50	27.50

Anmerkungen. Basis der Werte ist sind die gemittelten Einschätzungen der zwei Beurteiler. Der theoretische Wertebereich der Gesamtleistung liegt von 0 bis 40.

In Tabelle 7 sind die mittleren Urteile der zwei Beurteiler über alle 50 Testpersonen für jedes der zehn Items der MMI-Aufgabe *Berufswahl* abgetragen, sowie der Mittelwert beider Beurteiler pro Item, die Standardabweichung und die Itemschwierigkeiten. Die Mittelwerte der Beurteiler für die zehn Items unterscheiden sich teilweise recht deutlich, so weist Beurteiler 1 bezüglich des Items Beruf_01 einen um nahezu einen Punkt geringeren Mittelwert als Beurteiler 2 bei diesem Item auf. Bei den Items Beruf_03 und Beruf_05 finden sich ebenfalls Unterschiede in den Mittelwerten pro Item, diese fallen jedoch geringer aus. Die über alle Beurteiler und Testpersonen errechneten Mittelwerte der zehn Items variieren von $M = 0.05$ (Beruf_02) bis $M = 2.65$ (Beruf_02). Die Itemschwierigkeiten der zehn Items sind in Tabelle 7 unter p_m abgetragen. Das Item Beruf_02 stellt mit einem Schwierigkeitsindex von $p_m = .01$ das schwierigste Item dar, nahezu keine Testperson erreichte hier Punkte. Das leichteste Item hingegen ist Beruf_01 mit einem Schwierigkeitsindex von $p_m = .66$. Insgesamt weisen sechs der zehn Items (Beruf_02, Beruf_04, Beruf_06, Beruf_07, Beruf_09, Beruf_19) Schwierigkeitsindizes unterhalb von .50 auf und lassen sich somit als eher schwierige Items klassifi-

zieren. Die Items Beruf_01, Beruf_03 und Beruf_08 sind aufgrund ihrer Schwierigkeitsindizes oberhalb von .50 eher leichte Items. Das Item Beruf_06 weist als einziges Item eine nahezu mittlere Schwierigkeit von $p_m = .52$ auf.

Tabelle 7

Gemittelte Urteile der zwei Beurteiler über alle Testpersonen in Bezug auf die zehn Items der MMI-Aufgabe Berufswahl sowie Schwierigkeitsindizes für jedes Item

Item	Beurteiler		<i>M</i>	<i>SD</i>	p_m
	1	2			
Beruf_01	2.16	3.14	2.65	.47	.66
Beruf_02	0.08	0.02	0.05	.25	.01
Beruf_03	2.14	2.60	2.37	.67	.59
Beruf_04	1.30	1.16	1.23	.92	.31
Beruf_05	1.80	2.35	2.08	.73	.52
Beruf_06	1.59	1.80	1.68	.79	.42
Beruf_07	1.96	1.70	1.81	.68	.45
Beruf_08	2.54	2.56	2.55	.59	.64
Beruf_09	1.62	1.66	1.64	.72	.41
Beruf_10	1.72	1.58	1.65	.56	.41

Anmerkungen. Der theoretische Wertebereich für jedes Item erstreckt sich von 0 bis 4.

Um die Beurteilerübereinstimmung der zwei eingesetzten Beurteiler bei der MMI-Aufgabe *Berufswahl* zu untersuchen, werden zunächst die Mittelwerte der Beurteiler über alle Items und Testpersonen betrachtet sowie die vergebenen Minimal- und Maximalwerte der Beurteiler. In Tabelle 8 sind diese Ergebnisse aufgeführt.

Tabelle 8

Mittelwerte, Standardabweichungen sowie Minimum und Maximum der Beurteiler

Beurteiler Berufswahl	M	SD	Minimum	Maximum
1	1.69	0.93	0.00	4.00
2	1.86	1.17	0.00	4.00

Anmerkungen. Grundlage sind die über alle Items und Testpersonen gemittelten Urteile. Der theoretische Wertebereich der zu vergebenden Punkte erstreckt sich von 0 bis 4.

Die Mittelwerte der zwei Beurteiler über alle Items und Testpersonen unterscheiden sich nur geringfügig und liegen bei $M = 1.69$ und $M = 1.86$. Beide Beurteiler weisen eine recht geringe Standardabweichung von $SD = 0.93$ und $SD = 1.17$ auf. Anhand der vergebenen Minimal- und Maximalwerte der Beurteiler lässt sich jedoch feststellen, dass der theoretische Wertebereich der Beurteilungsskala vollständig ausgenutzt wurde, so vergaben beide Beurteiler als niedrigsten Wert 0 und als höchsten Wert 4.

Um die Beurteilerübereinstimmung für jedes Item zu bestimmen, wurden Intraklassenkorrelationen berechnet. Dabei werden die zwei Koeffizienten $ICC2$ und $ICC(2,2)$ herangezogen, ersterer trifft Aussagen über die Reliabilität eines Beurteilers, letzterer steht für die Reliabilität des über beide Beurteiler gemittelten Urteils. Tabelle 9 zeigt die beiden Koeffizienten für die Beurteilungen aller zehn Items unter Berücksichtigung der zwei Beurteiler.

Bezüglich des Kennwertes $ICC2$ liegen neun der zehn Items deutlich unterhalb der Grenze von $ICC2 = .70$, so dass anhand eines Beurteilers keine zuverlässigen und genauen Urteile für die zehn Items erfolgen ($ICC2 = .10$ bis $ICC2 = .62$). Das Item Beruf_08 weist mit einem $ICC2$ von .69 einen knapp zufriedenstellenden Reliabilitätswert des Urteils auf. Betrachtet man den $ICC(2,2)$, so weisen lediglich die Items Beruf_04 und Beruf_08 eine hohe Beurteilerübereinstimmung auf, vorausgesetzt, man mittelt die Urteile beider Beurteiler. Alle anderen Items bis auf Item Beruf_01 weisen $ICC(2,2)$ Werte auf, welche um den Wert von .60 liegen. Die geringste Beurteilerübereinstimmung zeigt sich insgesamt bezüglich beider ICC -Kennwerte für das Item Beruf_01, so beläuft sich der $ICC2$ nur auf .10 und der $ICC(2,2)$ auf .17.

Tabelle 9

Intraklassenkorrelationen für alle zehn Items der MMI-Aufgabe Berufswahl unter Berücksichtigung der zwei Beurteiler

Item	ICC2	ICC(2,2)
Beruf_01	.10	.17
Beruf_02	.44	.61
Beruf_03	.40	.57
Beruf_04	.62	.77
Beruf_05	.44	.61
Beruf_06	.46	.63
Beruf_07	.45	.61
Beruf_08	.69	.81
Beruf_09	.44	.61
Beruf_10	.43	.60

Zur Bestimmung der Reliabilität der Gesamtskala mittels Cronbachs Alpha gingen alle zehn Items in eine Reliabilitätsanalyse ein. Datengrundlage stellten die über beide Beurteiler gemittelten Urteile dar. Insgesamt ließ sich für die Gesamtskala mit allen zehn Items ein bereits zufriedenstellendes Cronbachs Alpha von .81 ermitteln. Die mittlere Inter-Item-Korrelation der zehn Items beträgt .30. Tabelle 10 zeigt für jedes der zehn Items den Skalenmittelwert, die Skalenvarianz sowie das Cronbachs Alpha, wenn das entsprechende Item aus der Skala ausgeschlossen wird. Unter r_{it} ist die Trennschärfe des Items abgetragen, welche sich aus der korrigierten Item-Skala-Korrelation ergibt.

Die Items Beruf_01, Beruf_03, Beruf_05, Beruf_08, Beruf_09 und Beruf_10 weisen insgesamt akzeptable bis sehr zufriedenstellende Werte auf. Die höchste Trennschärfe zeigt das Item Beruf_10 ($r_{it} = .85$). Die Items Beruf_02, Beruf_04, Beruf_06 sowie Beruf_07 zeigen unzufriedenstellende Trennschärfen (von $r_{it} = -.03$ bis $r_{it} = .33$), weshalb eine erneute Reliabilitätsanalyse unter Ausschluss dieser vier Items vorgenommen wurde. Tabelle 11 zeigt die Ergebnisse der Analyse unter Berücksichtigung von nur sechs Items. Insgesamt lässt sich anhand von sechs Items für diese reduzierte Skala ein Cronbachs Alpha von .88 feststellen. Die Trennschärfen der sechs Items liegen allesamt oberhalb von $r_{it} = .50$ und können somit als zufriedenstellend erachtet werden. Kein Item führt durch Ausschluss zu einer erhöhten Skalenreliabilität.

Tabelle 10

Reliabilitätsanalyse der zehn Items der MMI-Aufgabe Berufswahl

Skalenmittelwert:	17.71			
Skalenvarianz:	15.83			
Cronbachs Alpha:	.81			
Item	M, ohne	Skalenvarianz, ohne	r_{it}	$\alpha_{Cron}, \text{ ohne}$
Beruf_01	15.06	13.70	.56	.78
Beruf_02	17.66	15.84	-.03	.82
Beruf_03	15.34	12.64	.58	.78
Beruf_04	16.48	12.85	.33	.82
Beruf_05	15.63	12.71	.50	.79
Beruf_06	16.03	13.54	.29	.81
Beruf_07	15.90	13.76	.32	.81
Beruf_08	15.16	12.33	.76	.76
Beruf_09	16.07	11.48	.79	.75
Beruf_10	16.06	12.23	.85	.75

Tabelle 11

Reliabilitätsanalyse unter Ausschluss der Items Beruf_02, Beruf_04, Beruf_06 und Beruf_07 der MMI-Aufgabe Berufswahl

Skalenmittelwert:	12.94			
Skalenvarianz:	8.83			
Cronbachs Alpha:	.88			
Item	M, ohne	Skalenvarianz, ohne	r_{it}	$\alpha_{Cron}, \text{ ohne}$
Beruf_01	10.29	7.22	.56	.88
Beruf_03	10.57	6.33	.61	.87
Beruf_05	10.86	6.23	.57	.88
Beruf_08	10.39	6.22	.77	.84
Beruf_09	11.30	5.58	.80	.83
Beruf_10	11.29	6.14	.86	.83

MMI-Aufgabe Bildbeschreibung

Tabelle 12 zeigt die Gesamtleistung der Testpersonen in der MMI-Aufgabe *Bildbeschreibung*. Sie wird auf Grundlage der über beide Beurteiler gemittelten Urteile als Summe aller sechs Itempunkte errechnet und liegt im Durchschnitt bei einem Gesamtscore von $M = 16.06$ ($SD = 3.22$). Die niedrigste Gesamtleistung der Testpersonen beläuft sich auf einen Wert von 8.00 und die höchste Gesamtleistung liegt bei 27.00.

Tabelle 12

Gesamtleistung der Testpersonen in der MMI-Aufgabe Bildbeschreibung

	<i>M</i>	<i>SD</i>	Minimum	Maximum
Gesamtscore <i>Bildbeschreibung</i>	16.06	3.22	8.00	27.00

Anmerkungen. Basis der Werte ist sind die gemittelten Einschätzungen der zwei Beurteiler. Der theoretische Wertebereich der Gesamtleistung liegt von 6 bis 30.

In Tabelle 13 die über die Testpersonen gemittelten Urteile der zwei Beurteiler für jedes der sechs Items abgetragen. In den Spalten daneben ist der Mittelwert über beide Beurteiler für alle Items aufgeführt, sowie die Standardabweichung und der Itemschwierigkeitsindex p_m . Die Mittelwerte der zwei Beurteiler pro Item liegen relativ nah beieinander und unterscheiden sich in der Regel nicht stärker als um 0.30. Die Mittelwerte der sechs Items, welche über beide Beurteiler ermittelt wurden, variieren von $M = 1.15$ (Bild_02) bis $M = 3.11$ (Bild_04). Die Itemschwierigkeiten liegen für fünf der sechs Items im mittleren Bereich um $p_m = .50$, das Item Bild_02 sticht als extrem schwieriges Item mit einem Schwierigkeitsindex von $p_m = .04$ deutlich heraus.

Betrachtet man die zwei Beurteiler der MMI-Aufgabe *Bildbeschreibung*, so liegen jeweiligen mittleren Einschätzungen über alle Items und Testpersonen mit $M = 2.76$ und $M = 2.59$ nicht weit auseinander. Ebenfalls weisen die beiden Mittelwerte eine ähnliche Standardabweichung auf, so dass beide Beurteiler gleichermaßen in ihren Einschätzungen differenzierten. In Tabelle 14 sind ebenfalls die vergebenen Minimal- und Maximalpunktwerte der beiden Beurteiler aufgeführt. Da pro Item Werte von 1 bis 5 vergeben werden konnten, nutzen beide Beurteiler mit einem Minimalwert von 1 und einem Maximalwert von 5 die Punkteskala vollständig aus.

Tabelle 13

Gemittelte Urteile der zwei Beurteiler über alle Testpersonen in Bezug auf die sechs Items der MMI-Aufgabe Bildbeschreibung sowie Schwierigkeitsindizes für jedes Item

Item	Beurteiler		<i>M</i>	<i>SD</i>	<i>p_m</i>
	1	2			
Bild_01	3.18	2.84	3.01	.91	.50
Bild_02	1.18	1.12	1.15	.52	.04
Bild_03	3.04	2.68	2.86	.76	.47
Bild_04	3.26	2.96	3.11	.72	.53
Bild_05	3.10	3.02	3.06	.79	.52
Bild_06	2.82	2.92	2.87	.61	.47

Anmerkungen. Der theoretische Wertebereich für jedes Item erstreckt sich von 1 bis 5.

Tabelle 14

Mittelwerte, Standardabweichungen sowie Minimum und Maximum der Beurteiler

Beurteiler Bildbeschreibung	<i>M</i>	<i>SD</i>	Minimum	Maximum
1	2.76	1.04	1.00	5.00
2	2.59	1.06	1.00	5.00

Anmerkungen. Grundlage sind die über alle Items und Testpersonen gemittelten Urteile. Der theoretische Wertebereich der zu vergebenden Punkte erstreckt sich von 1 bis 5.

Zur Berechnung der Beurteilerübereinstimmung wurden auch bei der MMI-Aufgabe *Bildbeschreibung* die beiden Intraklassenkorrelationskoeffizienten *ICC2* und *ICC(2,2)* herangezogen, so dass die Urteilsgenauigkeit für einen Beurteiler (*ICC2*) und für das durchschnittliche Urteil beider Beurteiler (*ICC(2,2)*) bestimmt wurde. In Tabelle 15 sind beide *ICC*-Kennwerte für die sechs Items der *Bildbeschreibung* unter Berücksichtigung der beiden Beurteiler abgetragen.

Tabelle 15

Intraklassenkorrelationen für alle sechs Items der MMI-Aufgabe Bildbeschreibung unter Berücksichtigung der zwei Beurteiler

Item	ICC2	ICC(2,2)
Bild_01	.60	.75
Bild_02	.83	.91
Bild_03	.62	.77
Bild_04	.63	.77
Bild_05	.80	.89
Bild_06	.63	.77

Betrachtet man zunächst die Urteilsgenauigkeit eines Beurteilers anhand des ICC2-Koeffizienten, so erreichen die Items Bild_02 ($ICC2 = .83$) und Bild_05 ($ICC2 = .80$) sehr zufriedenstellende Werte oberhalb von .70. Die restlichen vier Items (Bild_01, Bild_03, Bild_04 und Bild_06) liegen mit Werten von $ICC2 = .60$ bis $ICC2 = .63$ unterhalb der akzeptablen Grenze von .70 und erreichen somit anhand nur eines Beurteilers keine reliablen Ergebnisse. Die Intraklassenkorrelationen für das gemittelte Urteil der beiden Beurteiler ($ICC(2,2)$) fällt für alle sechs Items sehr zufriedenstellend aus, die Kennwerte liegen zwischen $ICC(2,2) = .75$ und $ICC(2,2) = .91$.

Untersucht man die Reliabilität unter Zuhilfenahme des Kennwertes Cronbachs Alpha zur Bestimmung der internen Konsistenz anhand aller sechs Items der MMI-Aufgabe *Bildbeschreibung*, ergibt sich ein Cronbachs Alpha von .83 und eine mittlere Inter-Item-Korrelation von .44. In Tabelle 16 sind die Trennschärfen der Items abgetragen sowie Angaben über den Mittelwert, die Skalenvarianz und Cronbachs Alpha, wenn das jeweilige Item aus der Reliabilitätsanalyse ausgeschlossen wird. Die Items Bild_01, Bild_03 und Bild_05 erreichen zufriedenstellende Trennschärfen von $r_{it} = .72$ bis $r_{it} = .76$, die andere Hälfte der Items weist allerdings nur passable bis nicht zufriedenstellende Trennschärfen auf ($r_{it} = .41$ bis $r_{it} = .52$). Der Ausschluss keines Items führt zu einer deutlich relevanten Erhöhung der Skalenreliabilität.

Tabelle 16

Reliabilitätsanalyse der sechs Items der MMI-Aufgabe Bildbeschreibung

Skalenmittelwert:	16.06			
Skalenvarianz:	10.83			
Cronbachs Alpha:	.83			
Item	M, ohne	Skalenvarianz, ohne	r_{it}	$\alpha_{Cron}, \text{ ohne}$
Bild_01	13.05	6.27	.72	.78
Bild_02	14.91	8.86	.41	.84
Bild_03	13.20	6.86	.75	.77
Bild_04	12.95	7.76	.52	.82
Bild_05	13.00	6.66	.76	.77
Bild_06	13.19	8.32	.48	.83

MMI-Aufgabe Ausflug in den Zoo

Die Leistungen in der MMI-Aufgabe *Ausflug in den Zoo* wurden im Gegensatz zu den anderen beiden MMI-Aufgaben von drei Beurteilern eingeschätzt. Die Gesamtleistung einer Testperson errechnet sich jedoch auch hier aus der Summe der einzelnen Itempunkte, nur dass in diesem Fall über die Urteile von drei Beurteilern gemittelt wurde. Insgesamt erreichten die 50 Testpersonen in der MMI-Aufgabe *Ausflug in den Zoo* im Durchschnitt einen Wert von $M = 18.89$ ($SD = 2.94$). Die niedrigste Gesamtleistung liegt dabei nahezu doppelt so hoch als der minimal zu erreichende Wert von 6, nämlich bei 11.67. Die höchste Leistung einer Testperson liegt bei 24.33. Insgesamt schnitten die Testpersonen in dieser MMI-Aufgabe recht positiv ab. Da den Testpersonen während der Bearbeitung der Aufgabe insgesamt neun Hinweisfragen gestellt werden konnten, um auf relevante, bisher nicht ausgeführte Aspekte aufmerksam zu machen, ergibt sich für jede Testperson ebenfalls ein Summenwert der gesamten benötigten Hinweise. Der Mittelwert, die Standardabweichung sowie die minimal und maximal benötigte Anzahl von Hinweisen befinden sich neben den Angaben zur Gesamtleistung der Testpersonen in Tabelle 17. Im Durchschnitt wurden den Testpersonen 2.50 Hinweisfragen gestellt ($SD = 1.66$). Es wurden insgesamt bis zu maximal acht Hinweisfragen pro Testperson gestellt, ebenfalls gab es Fälle, in denen keine Hinweisfrage zum Einsatz kam.

Tabelle 17

Gesamtleistung und benötigte Hinweise der Testpersonen in der MMI-Aufgabe Ausflug in den Zoo

	<i>M</i>	<i>SD</i>	Minimum	Maximum
Gesamtscore <i>Ausflug in den Zoo</i>	18.89	2.94	11.67	24.33
Hinweise	2.50	1.66	0.00	8.00

Anmerkungen. Basis der Werte ist sind die gemittelten Einschätzungen der drei Beurteiler. Der theoretische Wertebereich der Gesamtleistung erstreckt sich von 6 bis 30. Insgesamt konnten 9 Hinweisfragen gestellt werden.

In die Itemanalysen der MMI-Aufgabe *Ausflug in den Zoo* gingen die über alle Testpersonen gemittelten Urteile von drei Beurteilern pro Item ein. Tabelle 18 zeigt die Mittelwerte der drei Beurteiler für jedes der sechs Items, sowie über alle Testpersonen und Beurteiler gemittelte Werte pro Item, die entsprechende Standardabweichung und den Schwierigkeitsindex. Die Mittelwerte der Items über alle drei Beurteiler variieren von $M = 3.81$ bis $M = 2.54$. Von den sechs Items erweisen sich die drei Items Zoo_01, Zoo_05 und Zoo_06 als psychometrisch leicht, so liegen die Schwierigkeitsindizes oberhalb von .50 ($p_m = .60$ bis $p_m = .70$). Das Item Zoo_02 liegt genau im mittleren Schwierigkeitsbereich mit einem Schwierigkeitsindex von $p_m = .51$. Die zwei Items Zoo_03 und Zoo_04 stellen eher schwierige Items dar, so belaufen sich ihre Schwierigkeitsindizes auf $p_m = .38$ und $p_m = .40$.

Tabelle 18

Gemittelte Urteile der drei Beurteiler über alle Testpersonen in Bezug auf die sechs Items der MMI-Aufgabe Ausflug in den Zoo sowie Schwierigkeitsindizes für jedes Item

Item	Beurteiler			<i>M</i>	<i>SD</i>	p_m
	1	2	3			
Zoo_01	3.96	2.58	3.96	3.50	0.53	.62
Zoo_02	3.32	2.46	3.34	3.04	0.66	.51
Zoo_03	2.56	2.08	3.20	2.61	1.08	.40
Zoo_04	2.48	2.18	2.96	2.54	1.10	.38
Zoo_05	3.74	3.06	3.38	3.39	0.81	.60
Zoo_06	4.08	3.32	4.02	3.81	0.96	.70

Anmerkungen. Der theoretische Wertebereich für jedes Item erstreckt sich von 1 bis 5.

Betrachtet man die in Tabelle 19 aufgeführten Mittelwerte der drei Beurteiler über alle Testpersonen für jedes Item, so fällt auf, dass Beurteiler 2 für jedes Item teilweise deutlich niedrigere Mittelwerte zeigt. Dies schlägt sich auch in den Beurteilmittelwerten über alle Testpersonen und Items (siehe Tabelle 18). Zwar nutzen alle drei Beurteiler alle Skalenpunkte von 1 bis 5, doch so unterscheiden sich ihre mittleren Urteile voneinander. Beurteiler 3 weist den höchsten Mittelwert von 3.48 ($SD = 1.12$) auf, Beurteiler 1 liegt mit einem mittlerem Urteil von $M = 3.36$ ($SD = 1.20$) nur geringfügig unterhalb Beurteiler 1, das mittlere Urteil von Beurteiler 2 hingegen weicht mit $M = 2.61$ ($SD = 1.18$) deutlich von den zwei anderen Beurteilern ab.

Tabelle 19

Mittelwerte, Standardabweichungen sowie Minimum und Maximum der Beurteiler

Beurteiler Ausflug in den Zoo	M	SD	Minimum	Maximum
1	3.36	1.20	1.00	5.00
2	2.61	1.18	1.00	5.00
3	3.48	1.12	1.00	5.00

Anmerkungen. Grundlage sind die über alle Items und Testpersonen gemittelten Urteile. Der theoretische Wertebereich der zu vergebenden Punkte erstreckt sich von 1 bis 5.

Vergleicht man die drei mittleren Urteile der Beurteiler varianzanalytisch, so ergibt sich bei der Berechnung einer ANOVA ein signifikanter Effekt von $F(2,897) = 48.37$ ($p < .001$). Das Urteil von Beurteiler 2 scheint im Vergleich zu den Urteilen von Beurteiler 1 und 3 deutlich strenger auszufallen und kann in diesem Vergleich von einer Strengetendenz des Beurteilers 2 gesprochen werden. In Abbildung 4 sind die durchschnittlichen Einschätzungen der drei Beurteiler aufgeführt.

Zur Bestimmung der Beurteilerübereinstimmungen der drei Beurteiler wurde die Intraklassenkorrelation herangezogen und ebenfalls der Kennwert $ICC2$ zur Bestimmung der Urteilsgenauigkeit eines Beurteilers sowie der $ICC(2,3)$ zur Bestimmung der Urteilsgenauigkeit des über die drei Beurteiler gemittelten Urteils errechnet. Tabelle 20 zeigt die Kennwerte für alle sechs Items unter Berücksichtigung der drei Beurteiler.

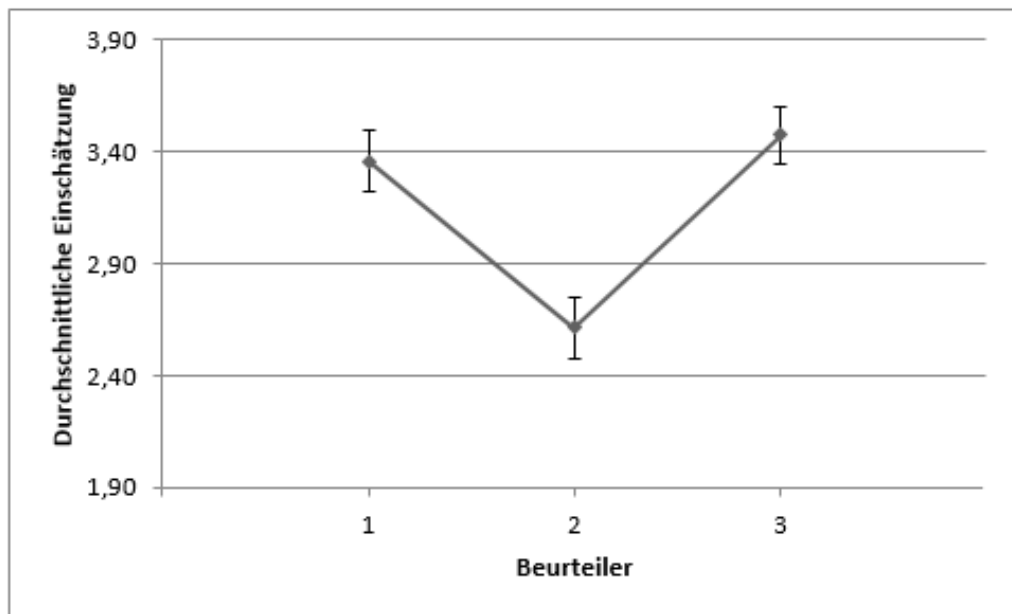


Abbildung 4. Mittelwerte der drei Beurteiler über alle Items und Testpersonen. Die Fehlerbalken repräsentieren 95 %-Konfidenzintervalle.

Tabelle 20

Intraklassenkorrelationen für alle sechs Items der MMI-Aufgabe Ausflug in den Zoo unter Berücksichtigung der drei Beurteiler

Item	ICC2	ICC(2,3)
Zoo_01	.14	.32
Zoo_02	.25	.50
Zoo_03	.39	.66
Zoo_04	.69	.87
Zoo_05	.45	.71
Zoo_06	.60	.82

Der ICC2-Kennwert fällt für die ersten beiden Items Zoo_01 und Zoo_02 mit $ICC2 = .14$ und $ICC2 = .25$ besonders niedrig aus, so dass hier keine Reliabilität bezüglich des Urteils eines einzelnen Beurteilers gegeben ist. Die drei Items Zoo_03, Zoo_05 und Zoo_06 variieren in ihren Kennwerten im nicht akzeptablen Bereich von $ICC2 = .39$ bis $ICC2 = .60$, das Item Zoo_04 liegt nur marginal unter der Grenze von $.70$ ($ICC2 = .69$). Mittelt man die Urteile der drei Beurteiler und betrachtet die Genauigkeit dieses mittleren Urteils, so führen nur die Hälfte der Items (Zoo_04, Zoo_05, Zoo_06) zu

reliablen Urteilen von $ICC(2,3) = .71$ bis $ICC(2,3) = .87$. Die ersten drei Items Zoo_01, Zoo_02 und Zoo_03 sind auch unter Berücksichtigung des gemittelten Urteils unterhalb der akzeptablen Grenze von .70.

Um auch die Beurteilerübereinstimmung von nur zwei Beurteilern analysieren zu können, wurden für die drei Beurteiler jeweils die drei Möglichkeiten der Kombination von zwei Beurteilern errechnet. Tabelle 21 zeigt die Intraklassenkorrelationen für die jeweiligen Kombinationsmöglichkeiten der drei Beurteiler. Keine der drei möglichen Kombinationen aus zwei Beurteilern erreicht weder im Kennwert $ICC2$ noch im Kennwert $ICC(2,2)$ für mehr als drei Items zufriedenstellende Werte größer oder gleich .70. Somit ist keine der Kombinationen aus zwei Beurteilern der gesamtheitlichen Berücksichtigung dreier Beurteiler vorzuziehen. Die Reliabilität der Urteile von zwei und drei Beurteilern fällt ähnlich aus.

Tabelle 21

Intraklassenkorrelationen für alle sechs Items der MMI-Aufgabe Ausflug in den Zoo unter Berücksichtigung jeweils einer Kombination aus zwei der drei Beurteiler

Item	Beurteiler 1 und 3		Beurteiler 1 und 2		Beurteiler 2 und 3	
	$ICC2$	$ICC(2,2)$	$ICC2$	$ICC(2,2)$	$ICC2$	$ICC(2,2)$
Zoo_01	.44	.61	.11	.20	.07	.13
Zoo_02	.43	.60	.23	.37	.18	.30
Zoo_03	.58	.73	.16	.27	.41	.58
Zoo_04	.67	.80	.78	.88	.62	.76
Zoo_05	.32	.49	.52	.68	.48	.65
Zoo_06	.70	.82	.54	.71	.58	.74

Bezieht man die über die drei Beurteiler gemittelten Einschätzungen in eine Reliabilitätsanalyse der Gesamtskala bestehend aus sechs Items ein, so ergibt sich auch für diese Gesamtskala keine zufriedenstellende interne Konsistenz, Cronbachs Alpha beläuft sich auf den nicht akzeptablen Wert von .55. Die mittlere Inter-Item-Korrelation beträgt .22. In Tabelle 22 ist der Skalenmittelwert, die Skalenvarianz sowie Cronbachs Alpha abgetragen, wenn das entsprechende Item entfernt wird. Der Kennwert r_{it} gibt die Trennschärfe der Items als korrigierte Item-Skala-Korrelation an.

Tabelle 22

Reliabilitätsanalyse der sechs Items der MMI-Aufgabe Ausflug in den Zoo

Skalenmittelwert:	18.89			
Skalenvarianz:	8.63			
Cronbachs Alpha:	.55			
Item	M, ohne	Skalenvarianz, ohne	r_{it}	α_{Cron} , ohne
Zoo_01	15.39	7.01	.48	.47
Zoo_02	15.85	6.98	.35	.49
Zoo_03	16.28	6.51	.17	.58
Zoo_04	16.35	6.89	.09	.62
Zoo_05	15.50	5.78	.56	.38
Zoo_06	15.09	6.03	.36	.47

Die Trennschärfen aller sechs Items sind insgesamt mit Werten von $r_{it} = .09$ bis $r_{it} = .56$ unterdurchschnittlich und nicht zufriedenstellend. Das Item Zoo_04 weist sogar einen Wert nahe Null auf ($r_{it} = .09$). Da der Ausschluss der zwei Items mit den geringsten Trennschärfen (Zoo_03 und Zoo_04) auch zu einer relevanten Verbesserung von Cronbachs Alpha führt, wurde die Reliabilitätsanalyse für die um zwei Items reduzierte Skala wiederholt. Die Ergebnisse der erneuten Reliabilitätsanalyse sind in Tabelle 23 abgetragen. Der Ausschluss der zwei Items führt zu einer Erhöhung von Cronbachs Alpha auf .64, die mittlere Inter-Item-Korrelation beläuft sich auf .34. Trotz der positiven Veränderung der internen Konsistenz erreicht Cronbachs Alpha noch keinen zufriedenstellenden Wert. Auch die Trennschärfen verbleiben für die vier berücksichtigten Items in nicht zufriedenstellenden Bereichen. Der Ausschluss keines weiteren Items führt zu einer deutlichen Verbesserung der internen Konsistenz.

Tabelle 23

Reliabilitätsanalyse unter Ausschluss der Items Zoo_03 und Zoo_04 der MMI-Aufgabe Ausflug in den Zoo

Skalenmittelwert:	13.74			
Skalenvarianz:	4.41			
Cronbachs Alpha:	.64			
Item	M, ohne	Skalenvarianz, ohne	r_{it}	α_{Cron} , ohne
Zoo_01	10.24	3.14	.52	.54
Zoo_02	10.70	3.07	.39	.59
Zoo_05	10.35	2.51	.48	.52
Zoo_06	9.93	2.40	.37	.64

5.1 Validitätsanalysen

Die Untersuchung der Konstruktvalidität der drei MMI-Aufgaben erfolgt über eine Korrelationsanalyse. Hierzu wurden die auf dem gemittelten Urteil der Beobachter basierenden und über alle Items aufsummierten Gesamtleistungsscores in den MMI-Aufgaben *Berufswahl*, *Bildbeschreibung* und *Ausflug in den Zoo* mit den jeweiligen globalen Einschätzungen der Bewältigung der MMI-Aufgaben durch die Testleiter korreliert. Die hierbei resultierende Korrelationsmatrix entspricht in etwa einer so genannten Multitrait-Multimethod Matrix und lässt Rückschlüsse auf die konvergente und diskriminante Validität zu (Campbell & Fiske, 1959). Jedoch werden die erfassten Konstrukte über zwei Methoden erfasst, nämlich dem Urteil der Beurteiler anhand des Beurteilungsbogens und einer globalen Einschätzung der Leistung durch den Testleiter, welche sich beide auf dieselben Leistungen beziehen, nämlich in den drei MMI-Aufgaben. Tabelle 24 zeigt die Korrelationsmatrix.

Im Sinne der konvergenten Validität werden die Korrelationen der Konstrukte über alle Methoden betrachtet. Der MMI-Gesamtscore der Aufgabe *Berufswahl* korreliert mit $r = .46$ ($p < .01$) signifikant mit der Globaleinschätzung der Bewältigung der *Berufswahl*-Aufgabe durch den Testleiter. Eine etwas höhere Korrelation zeigt sich zwischen dem MMI-Gesamtscore der Aufgabe *Bildbeschreibung* und der Globaleinschätzung der Bewältigung der *Bildbeschreibung*-Aufgabe durch den Testleiter mit $r = .60$, diese Korrelation ist ebenfalls mit $p < .01$ signifikant. Die höchste signifikante Korrelation weist der Gesamtscore der MMI-Aufgabe *Ausflug in den Zoo* mit der entsprechenden Globaleinschätzung der Bewältigung der Aufgabe durch den Testleiter auf und beläuft sich auf $r = .67$ ($p < .01$).

Tabelle 24

Korrelationsmatrix der Gesamtscores der einzelnen MMI-Aufgaben und den Globaleinschätzungen der Bewältigung der einzelnen MMI-Aufgaben

	Globaleinschätzungen			MMI-Gesamtscore		
	Beruf	Bild	Zoo	Beruf	Bild	Zoo
Globaleinschätzungen						
Beruf	–					
Bild	.35*	–				
Zoo	.32*	.23	–			
MMI-Gesamtscore						
Beruf	.46**	.34*	.31*	–		
Bild	.15	.60**	.31*	.33*	–	
Zoo	.24	.08	.67**	.34*	.23	–

Anmerkungen. * Die Korrelation ist auf einem Niveau von $p \leq .05$ signifikant; ** Die Korrelation ist auf einem Niveau von $p \leq .01$ signifikant.

Verschiedene Konstrukte sollten im Sinne der diskriminanten Validität über eine Methode niedrig korrelieren. Die Korrelationen zwischen dem MMI-Gesamtscore der *Berufswahl*, der *Bildbeschreibung* und dem *Ausflug in den Zoo* fallen mit Werten von $r = .33$ ($p < .05$) und $r = .34$ mittelmäßig aus. Die Korrelation zwischen dem Gesamtscore der *Bildbeschreibung* und der Zoo-Aufgabe fällt mit $r = .23$ etwas niedriger aus. Die Korrelationen der Globaleinschätzungen fallen ebenso für die Zusammenhänge verschiedener MMI-Aufgaben mittelmäßig positiv aus ($r = .23$ bis $r = .35$). Die Korrelationen zwischen verschiedenen Konstrukten und unterschiedlichen Methoden sollte insgesamt niedrig ausfallen. Betrachtet man die Zusammenhänge unterschiedlicher MMI-Gesamtscores und Globaleinschätzungen der Bewältigung der MMI-Aufgaben, so ergeben sich Korrelationskoeffizienten von niedriger bis mittelmäßiger Höhe von $r = .08$ bis $r = .34$. Um einer Alpha-Fehler-Kumulierung vorzubeugen, sollte laut der B-Y-Methode ein Signifikanzniveau von mindestens $p < .01$ angestrebt werden. Das Erreichen des Signifikanzniveaus von $p < .05$ reicht in diesem Falle nicht aus, um ein gesichertes Ergebnis zu liefern.

Es wird deutlich, dass die Gesamtscores der verschiedenen MMI-Aufgaben nicht ganz unabhängig voneinander sind und somit eventuell nicht nur drei aufgabenspezifische Konstrukte gemessen werden. Neben der Korrelationsmatrix der Gesamtscores und den Globaleinschätzungen wurde daher zudem die Faktorenstruktur der Gesamtheit aller Beurteilungssitems der drei MMI-Aufgaben genauer betrachtet. Zuvor wurde der vorliegende Datensatz auf eine Eignung für eine

Faktorenanalyse untersucht. Betrachtet man die Korrelationsmatrix aller Items der drei Beurteilungsbögen, so liegen häufig geringe Korrelationswerte vor, was nur begrenzt für die faktorenanalytische Eignung der Daten spricht. Das Kaiser-Meyer-Olkin-Kriterium, ein Maß der Stichprobeneignung, weist einen Wert von .602 auf, befindet die Daten nach Konvention also als mittelmäßig geeignet für eine faktorenanalytische Untersuchung. Der Bartlett-Test auf Sphärizität fällt hingegen signifikant aus ($\chi^2(231) = 507.895, p < .0001$). Aufgrund der zwar mittelmäßigen aber dennoch gegebenen Eignung der Stichprobe für eine Faktorenanalyse, wurden die Daten einer Hauptkomponentenanalyse mit Varimax-Rotation unterzogen. Nach dem Kaiser-Guttman-Kriterium werden jedoch sechs Faktoren extrahiert (vgl. Tabelle 25). Die sechs extrahierten Faktoren klären insgesamt 67.10 Prozent der Varianz auf, wobei der erste Faktor mit 26.20 Prozent deutlich am varianzstärksten ist, die Anteile der erklärten Varianz der restlichen sechs Faktoren beläuft sich auf 11.64 bis 5.24 Prozent. Die Kommunalitäten der einzelnen Items reichen von $h^2 = .48$ bis $h^2 = .86$.

Anstelle von drei aufgabenspezifischen Faktoren ergeben sich insgesamt sechs Faktoren. Auf Faktor 1 laden ausschließlich Items der MMI-Aufgabe *Berufswahl* hoch. Auf Faktor 2 lassen sich ausschließlich hohe Ladungen von Items der MMI-Aufgabe *Bildbeschreibung* erkennen. Auf den dritten Faktor laden nur die Items Zoo_01, Zoo_05 und Zoo_06 hoch. Aufgrund der jeweils aufgabenspezifischen Itemladungen auf die drei Faktoren können diese als weitestgehend aufgabenspezifische Faktoren gelten, somit kann Faktor 1 als spezifisch für die Aufgabe *Berufswahl*, Faktor 2 als spezifisch für die Aufgabe *Bildbeschreibung* und Faktor 3 als spezifisch für die Aufgabe *Ausflug in den Zoo* angesehen werden.

Die drei letzten Faktoren weisen weniger aufgabengebundene Ladungen auf. Auf Faktor 4 laden nur die Items Beruf_02 und Zoo_04 hoch, wobei es sich bei dem Item Beruf_02 um eine negative Ladung handelt. Die Items Zoo_02 und Beruf_04 weisen geringere Ladungen von .62 und .49 bezüglich des vierten Faktors auf. Faktor 5 zeigt hohe Ladungen der Items beruf_06, Beruf_07 und Bild_02 auf, wobei die Ladung von Bild_02 negativ ausfällt. Auf den sechsten Faktor lädt nur das Item Zoo_03 mit einer Faktorladung von .79 hoch. Die Ladung von Bild_06 kann mit .46 ebenfalls als relevant betrachtet werden. Anhand der Itemladungen auf Faktor 4 lassen sich keine inhaltlich eindeutigen Interpretationen für diesen Faktor anstellen. Auf Faktor 5 laden die Items Beruf_06, Beruf_07 und negativ das Item Bild_02 hoch. Diese Items stehen für lehrberufsrelevante persönliche Schwächen, Alternativen zum Lehrberuf und Vermittlung eines Orientierungsrahmens bei der *Bildbeschreibung*. Diese Items erfassen grundsätzlich Aspekte der Performanzen, welche möglicherweise ein hohes Maß an Selbstreflexion erfordern sowie die Fähigkeit, eine Metaperspektive einzunehmen. Faktor 6 weist insgesamt nur niedrige Faktorladungen auf, bezieht man die zwei Items mit den höchsten Ladungen ein, Zoo_03, also die Gruppenorganisation durch soziale Strukturierung

und Bild_06, also die hierarchische Gliederung der Darstellungen in der *Bildbeschreibung*, so laden auf diesen Faktor vor allem Items, die für eine Strukturierungs- und Gliederungsfähigkeit stehen.

Tabelle 25

Faktorenanalyse über alle Items der drei MMI-Aufgaben

Item	Rotierte Faktorladungen						Kommunalität h^2
	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	
Beruf_01	.69						.55
Beruf_02				-.77			.62
Beruf_03	.66			.31			.60
Beruf_04	.46			.49			.56
Beruf_05	.66					-.36	.69
Beruf_06	.37			-.35	.54		.64
Beruf_07					.69		.68
Beruf_08	.82						.76
Beruf_09	.87						.79
Beruf_10	.87						.86
Bild_01		.86					.80
Bild_02	.34	.34			-.58		.61
Bild_03		.80					.77
Bild_04		.70					.61
Bild_05		.87					.80
Bild_06		.52	-.38		-.31	.46	.75
Zoo_01			.63			.33	.66
Zoo_02			.37	.41	.39	.33	.58
Zoo_03						.79	.73
Zoo_04				.62			.48
Zoo_05			.64			.30	.63
Zoo_06			.77				.63
Anteil an Varianz- aufklärung	26.20	11.64	9.95	7.56	6.51	5.24	

Anmerkungen. Faktorladungen < .30 werden nicht aufgeführt. Der Anteil an Varianzaufklärung ist in Prozent angegeben.

Zur Untersuchung der Kriteriumsvalidität wurden bi- und multivariate Zusammenhänge der anhand der Beurteiler eingeschätzten Leistungen in den MMI-Aufgaben mit den Globaleinschätzungen der Testleiter sowie den Persönlichkeits- und Selbsteinschätzungen der Testpersonen vorgenommen. Die Selbsteinschätzungen der Testpersonen umfassen Items, welche sich zu drei aufgabenspezifisch relevanten Fähigkeits- und Kompetenzbereiche zusammenfassen lassen. So liegen Selbsteinschätzungen zur Berufswahlkompetenz vor, welche im Zusammenhang mit der Gesamtleistung in der MMI-Aufgabe *Berufswahl* stehen sollten. Selbsteinschätzungen der Erklärungskompetenz sollten Zusammenhänge mit der Gesamtleistung in der MMI-Aufgabe *Bildbeschreibung* aufweisen. Die Einschätzung der eigenen Planungskompetenz durch die Testpersonen sollte mit der Gesamtleistung in der MMI-Aufgabe *Ausflug in den Zoo* zusammenhängen. Hierfür wurden die jeweiligen MMI-Gesamtscores mit den Selbsteinschätzungen der drei Kompetenzbereiche korreliert. Tabelle 26 zeigt die Korrelationsmatrix.

Tabelle 26

Korrelationen der Selbsteinschätzungen mit den Gesamtscores in den MMI-Aufgaben

	Selbsteinschätzungen		
	Berufswahlkompetenz	Erklärungskompetenz	Planungskompetenz
MMI-Gesamtscore			
Berufswahl	-.02	-.05	-.07
Bildbeschreibung	-.07	-.18	-.05
Ausflug in den Zoo	.37**	.14	.30*

Anmerkungen. * Die Korrelation ist auf einem Niveau von $p \leq .05$ signifikant; ** Die Korrelation ist auf einem Niveau von $p \leq .01$ signifikant.

Die Gesamtleistung in der MMI-Aufgabe *Berufswahl* steht in keinem Zusammenhang mit der Selbsteinschätzung der Berufswahlkompetenz ($r = -.02$). Es zeigen sich zudem auch keine Zusammenhänge mit den Selbsteinschätzungen der Erklärungs- und Planungskompetenz ($r = -.05$ und $r = -.07$). Die Gesamtleistung in der MMI-Aufgabe *Bildbeschreibung* zeigt sogar einen minimalen negativen Zusammenhang mit der Selbsteinschätzung der Erklärungskompetenz ($r = -.18$). Die Gesamtleistung in dieser MMI-Aufgabe zeigt ebenfalls keine Zusammenhänge mit den Selbsteinschätzungen der nicht relevanten Berufs- und Planungskompetenz ($r = -.07$ und $r = -.05$). Die Gesamtleistung in der MMI-Aufgabe *Ausflug in den Zoo* weist eine mittlere signifikante Korrelation von $r = .37$ ($p < .01$) mit der Selbsteinschätzung der Berufswahlkompetenz auf. Es lässt sich ebenfalls ein mäßig positiver Zusammenhang von $r = .30$ ($p < .05$) mit den Selbsteinschätzungen der Planungskompetenz feststel-

len. Aufgrund der Vermeidung von Alpha-Fehler-Kumulierungen reicht hier der Wert von $p < .05$ jedoch nicht aus, um von einem gesicherten signifikanten Ergebnis sprechen zu können.

In den Beurteilungen der Leistungen bezüglich der MMI-Aufgabe *Berufswahl* und *Bildbeschreibung* durch die Beurteiler wurden Dimensionen berücksichtigt, welche vor allem Aspekte der sprachlichen Qualität wie den Ausdruck und die Struktur der sprachlichen Äußerungen einschätzten (Items Beruf_08, Beruf_10 sowie Bild_03 und Bild_05). In den Globaleinschätzungen durch die Testleiter wurde ebenfalls die Qualität des Ausdrucks und der Argumentation über alle drei MMI-Aufgaben bewertet (Item Qualität der Argumentation und Item Qualität des sprachlichen Ausdrucks). Die sprachlichen Dimensionen der MMI-Aufgaben *Berufswahl* und *Bildbeschreibung* wurden anhand bivariater Korrelationsanalysen auf Zusammenhänge mit den Globaleinschätzungen der sprachlichen Qualität durch die Testleiter untersucht. Tabelle 27 zeigt die entsprechenden Korrelationskoeffizienten.

Tabelle 27

Korrelationen der sprachlichen Beurteilungsdimensionen der MMI-Aufgaben Berufswahl und Bildbeschreibung mit der Globaleinschätzung der sprachlichen Qualität durch die Testleiter

Globaleinschätzung sprachlicher Qualität	
Dimensionen der MMI-Aufgaben	
Sprachlich_Berufswahl	.34*
Sprachlich_Bildbeschreibung	.45**

Anmerkungen. * Die Korrelation ist auf einem Niveau von $p \leq .05$ signifikant; ** Die Korrelation ist auf einem Niveau von $p \leq .01$ signifikant.

Die sprachlichen Beurteilungsdimensionen der MMI-Aufgabe *Bildbeschreibung* korrelieren signifikant mit der Globaleinschätzung der sprachlichen Qualität durch die Testleiter und dies relativ hoch ($r = .45$, $p < .01$). Die sprachlichen Beurteilungsdimensionen der MMI-Aufgabe *Berufswahl* korrelieren moderat mit den Globaleinschätzungen der sprachlichen Qualität ($r = .34$, $p < .05$). Aufgrund der Vermeidung von Alpha-Fehler-Kumulierungen reicht hier der Wert von $p < .05$ jedoch nicht aus, um von einem gesicherten signifikanten Ergebnis sprechen zu können.

Da es sich bei der MMI-Aufgabe *Bildbeschreibung* und *Ausflug in den Zoo* um eher komplexe Aufgabenstellungen handelt (die Ausflugsplanung an sich sowie das zu beschreibende Bild sind sehr umfassend), benötigen die Testpersonen eine gewisse Auffassungsgabe, um in der kurzen Bearbeitungszeit entsprechende Leistungen zu erbringen. Die MMI-Aufgabe *Berufswahl* legt insbesondere Wert auf die sprachliche Darstellung der Inhalte, da es sich um eine reine Narration berufsbezogener Inhalte handelt. Auch in der MMI-Aufgabe *Bildbeschreibung* spielen Aspekte des

sprachlichen Ausdrucks eine Rolle. Der jeweilige Gesamtscore in den drei MMI-Aufgaben wird auf seinen Zusammenhang mit den Globaleinschätzungen der Testleiter bezüglich der Auffassungsgabe und der sprachlichen Qualität untersucht. In Tabelle 28 sind die Ergebnisse der Korrelationsanalyse abgetragen.

Tabelle 28

Korrelationen zwischen den Gesamtscores der MMI-Aufgaben und den Globaleinschätzungen der Auffassungsgabe und der sprachlichen Qualität durch die Testleiter

	Globaleinschätzungen	
	Auffassungsgabe	Sprachliche Qualität
MMI-Gesamtscore		
Berufswahl	.33*	.24
Bildbeschreibung	.45**	.37*
Ausflug in den Zoo	.33*	.43*

Anmerkungen. * Die Korrelation ist auf einem Niveau von $p \leq .05$ signifikant; ** Die Korrelation ist auf einem Niveau von $p \leq .01$ signifikant.

Betrachtet man die Gesamtleistung der MMI-Aufgabe *Berufswahl* und nicht nur die sprachspezifische Beurteilungsdimension (siehe Tabelle 28), so lässt sich nahezu kein nennenswerter positiver Zusammenhang zur Globaleinschätzung der sprachlichen Qualität durch die Testleiter feststellen ($r = .24$). Die Gesamtleistung in der MMI-Aufgabe *Bildbeschreibung* hingegen zeigt eine höhere, mittlere Korrelation von $r = .37$ ($p < .05$) mit der Globaleinschätzung der sprachlichen Qualität durch die Testleiter. Die höchste Korrelation mit der Globaleinschätzung der sprachlichen Qualität zeigt die Gesamtleistung in der MMI-Aufgabe *Ausflug in den Zoo* mit $r = .43$ ($p < .05$). Um einer Alpha-Fehler-Kumulierung vorzubeugen, gelten diese Ergebnisse als nicht signifikant, da die Korrelationen nur auf einem Niveau von .05 signifikant werden. Die Gesamtleistungen in der MMI-Aufgabe *Berufswahl* und *Ausflug in den Zoo* korrelieren gleich hoch mit der Globaleinschätzung der Auffassungsgabe ($r = .33$, $p < .05$). Der stärkste positive Zusammenhang mit der Globaleinschätzung der Auffassungsgabe lässt sich mit $r = .45$ in Bezug auf die Gesamtleistung in der MMI-Aufgabe *Bildbeschreibung* finden, dieser wird sogar auf einem Niveau von $p < .01$ signifikant.

Der globale Eindruck, den die Testleiter von den Testpersonen nach Bewältigung der drei MMI-Aufgaben hatten, steht in einem stark positiven Zusammenhang mit der Gesamtleistung aller drei MMI-Aufgaben (errechnet durch die Aufsummierung der Aufgabengesamtscores), so beträgt die Korrelation $r = .57$ und ist signifikant ($p < .01$) (siehe Tabelle 29). Es lässt sich kein Zusammen-

hang zwischen der aufsummierten Gesamtleistung in den drei MMI-Aufgaben und der von den Testleitern eingeschätzten Aufgeregtheit der Testpersonen feststellen ($r = .09$).

Tabelle 29

Zusammenhänge zwischen der Gesamtleistung in allen drei MMI-Aufgaben und den Globaleinschätzungen des Gesamteindrucks und der Aufgeregtheit

	Globaleinschätzungen	
	Gesamteindruck	Aufgeregtheit
Gesamtleistung in allen drei MMI-Aufgaben	.51**	.09

Anmerkungen. ** Die Korrelation ist auf einem Niveau von $p \leq .01$ signifikant. Hohe Werte in der Globaleinschätzung der Aufgeregtheit stehen für Gelassenheit, niedrige Werte für Aufgeregtheit

Zur Überprüfung des multivariaten Zusammenhangs zwischen den drei erfassten Persönlichkeitsfaktoren Extraversion, Verträglichkeit und emotionale Stabilität sowie den aufgabenrelevanten Selbsteinschätzungen und der Gesamtleistung in den drei MMI-Aufgaben (aufsummierte Gesamtscores der drei MMI-Aufgaben), wurde eine lineare Regression des Kriteriums der MMI-Gesamtleistung auf die Prädiktoren Extraversion, Verträglichkeit, emotionale Stabilität, Berufswahlkompetenz, Erklärungskompetenz und Planungskompetenz durchgeführt. In Tabelle 30 sind die Ergebnisse abgetragen. In diesem Modell der Regressionsanalyse wird mit den sechs genannten Prädiktoren lediglich eine Varianzaufklärung von 8 Prozent erzielt ($F(6,43) = .690$; $p = .658$). Keiner der Prädiktoren hat einen signifikanten Einfluss auf die MMI-Gesamtleistung.

Tabelle 30

Ergebnisse der Regressionsanalyse für die MMI-Gesamtleistung.

Prädiktoren	β	T	p	VIF
SE_Berufswahlkompetenz	0.218	5.761	.282	1.886
SE_Erklärungskompetenz	-0.266	1.089	.182	1.812
SE_Planungskompetenz	0.169	-1.356	.382	1.731
Extraversion	-0.022	0.882	.891	1.157
Emotionale Stabilität	-0.180	-0.137	.244	1.101
Verträglichkeit	-0.174	-1.180	.265	1.122
R²			0.08	

6 Diskussion

Die vorliegende Arbeit widmet sich der Überprüfung der psychometrischen Qualität von Multiplen Mini-Interviews im Einsatzkontext der Erfassung von lehrberufsrelevanten Kompetenzen bei Lehramtsstudierenden. Hierzu wurden insgesamt sechs MMI-Aufgaben in einem zweigeteilten Erhebungsdurchlauf anhand von Lehramtsstudierenden überwiegend im Masterstudiengang durchgeführt und erprobt. Die Auswertungen und Analysen der vorliegenden Arbeit berücksichtigen ausschließlich den ersten Erhebungsteil bestehend aus drei MMI-Aufgaben zu Kompetenzbereichen der Berufswahl, der Erklärungs- und Instruktionskompetenz sowie planerischen und organisatorischen Kompetenzfacetten. Im Folgenden soll auf Grundlage der statistischen Ergebnisse systematisch auf die Untersuchungshypothesen eingegangen werden und die Befunde zur Reliabilität, Objektivität, Validität und Ökonomie kritisch erläutert und interpretiert werden.

Reliabilität

Ein sehr wichtiges Qualitätskriterium stellt die Reliabilität dar und trifft Aussagen darüber, wie zuverlässig und genau ein Verfahren misst. Die drei berücksichtigten MMI-Aufgaben *Berufswahl*, *Bildbeschreibung* und *Ausflug in den Zoo* treffen die Leistungsbeurteilung über theoretisch konzipierte Beurteilungsbögen, welche anhand einer gewissen Anzahl von Items das Verhalten unter Berücksichtigung verschiedener Aspekte der Performanz bewerten. Summiert man alle erreichten Punkte über die Items auf, so ergibt sich der Gesamtleistungsscore für die entsprechende Aufgabe. Dieser gibt Auskunft darüber, wie gut oder schlecht die Testpersonen in dieser Aufgabe unter Berücksichtigung der vorher festgelegten Leistungsfacetten abgeschnitten haben. Bei allen drei Aufgaben lässt sich feststellen, dass die berücksichtigte Stichprobe im Durchschnitt in etwa mittlere Leistungen in Bezug auf die zu erreichende Gesamtpunktzahl erzielt hat. Keine der Aufgaben sticht durch überwiegend niedrige Gesamtleistungen heraus. Analysiert man die Beschaffenheit der MMI-Aufgaben auf Itemebene der Beurteilungsbögen, so lassen sich dennoch einige Schwächen herausarbeiten. Betrachtet man die Trennschärfen der Items, welche dafür stehen, inwiefern ein Item zwischen den Leistungen der Testpersonen differenzieren kann, so zeigen die drei MMI-Aufgaben stellenweise wenig trennscharfe Items.

Die höchsten Trennschärfen weisen die Items der *Bildbeschreibung*-Aufgabe auf, diese sind allesamt hoch. Die Items der *Berufswahl* und der Aufgabe *Ausflug in den Zoo* hingegen differenzieren nicht alle entsprechend zufriedenstellend zwischen den Leistungen. Die Schwierigkeiten der Items variieren ebenfalls innerhalb der Aufgaben. Sowohl die Aufgabe *Bildbeschreibung* als auch *Berufswahl* weisen jeweils ein Item auf, das von nahezu keiner der 50 Testpersonen gelöst wurde. Bei der *Berufswahl*-Aufgabe handelt es sich hierbei um das Item, das die extrinsischen Berufswahl-

motive erfasst. In der vorliegenden Stichprobe überwog jedoch die Nennung von intrinsischen Motiven, so dass der hohe Schwierigkeitsindex in diesem Fall kein Zeichen für eine ungünstige Itemkonstruktion ist, sondern eher durch die vorliegende Stichprobenverteilung bedingt wurde. Insgesamt muss hier die Stichprobenabhängigkeit der Itemkennwerte berücksichtigt werden.

Die Leistungen der Testpersonen wurden mit Hilfe des entsprechenden Beurteilungsbogens für die MMI-Aufgaben *Berufswahl* und *Bildbeschreibung* von jeweils zwei Beurteilern und für die MMI-Aufgabe *Ausflug in den Zoo* von drei Beurteilern eingeschätzt. Da die Beurteilerabhängigkeit der Messungen bei simulations- und verhaltensorientierten Verfahren ein generelles Problem darstellt, wird diesem häufig mit dem Einsatz von mehreren geschulten Beurteilern zur Auswertung begegnet. Die Absicht, durch eine Vielzahl von Beurteilern reliable Messergebnisse zu erreichen, steht jedoch im Gegensatz zu dem Anspruch an das Verfahren, zeit- und ressourcensparend anwendbar zu sein. Aus diesem Grund wurde in der vorliegenden Arbeit die Anzahl der auswertenden Beurteiler auf ein Maximum von drei beschränkt. In der Forschungsliteratur zum MMI im medizinischen Kontext wird angegeben, dass mindestens zwei Beurteiler zur Vorbeugung von Beurteilerverzerrungen in den Messungen eingesetzt werden sollten (Hecker et al., 2009; Hissbach et al., 2014).

Als Maß der Reliabilität der Messungen wurde die Intraklassenkorrelation bezogen auf die jeweiligen zwei oder drei Beurteiler für jede MMI-Aufgabe berechnet. Dabei wurden die zwei Koeffizienten $ICC2$ und $ICC(2,k)$ herangezogen; ersterer trifft Aussagen über die Reliabilität eines Beurteilers, letzterer steht für die Reliabilität des über die Beurteiler gemittelten Urteils. Die Reliabilität des Urteils von nur einem Beurteiler fällt für die meisten Items aller drei MMI-Aufgaben nicht zufriedenstellend aus und macht deutlich, dass keine der drei MMI-Aufgaben mit der Zuhilfenahme von nur einem einzigen Beurteiler in der Lage ist, bezüglich aller Beurteilungsisems reliable Messergebnisse zu produzieren. Mittelt man jedoch die Urteile über die eingesetzten Beurteiler, so lassen sich nur anhand der MMI-Aufgabe *Bildbeschreibung* in allen Items reliable Einschätzungen der Leistungen erreichen. Das gemittelte Urteil aus zwei Beurteilern erweist sich für die MMI-Aufgabe *Berufswahl* nur in Bezug auf zwei der zehn Items als reliable Auswertung. Die restlichen Items können den notwendigen Kennwert von .70, um von einer reliablen Messung sprechen zu können, teilweise nur sehr nicht erreichen.

Für das gemittelte Urteil der MMI-Aufgabe *Ausflug in den Zoo* lagen Einschätzungen von drei Beurteilern vor. Berücksichtigt man alle drei Beurteiler, so können über drei der sechs Items reliable Werte erzielt werden. Die drei möglichen Zweier-Kombinationen aus den drei Beurteilern zeigen ebenfalls nur in Bezug auf maximal drei Items eine reliable Messung, wobei diese für jede Kombination in unterschiedlichen Items erreicht wird. Vergleicht man die Mittelwerte der drei Beurteiler über alle Testpersonen und Items, so zeigt sich eine deutliche Strengetendenz des zweiten

Beurteilers im Gegensatz zu den anderen beiden Beurteilern. Hier könnte also eine systematische Beurteilerverzerrung in den Auswertungen der MMI-Aufgabe Zoo vorliegen. Der Einsatz von drei anstelle von zwei Beurteilern bei dieser MMI-Aufgabe liegt in dem krankheitsbedingten Ausfall des eigentlich vorgesehenen zweiten Beurteilers begründet, so dass unter großen Zeitdruck zwei Hilfskräfte für die Auswertungen gewonnen werden mussten. Die nicht akzeptablen Beurteilerübereinstimmungen für die MMI-Aufgabe *Ausflug in den Zoo* könnten also dem Umstand geschuldet sein, dass zwei relativ kurzfristig akkreditierte Beurteiler mit nur einer vergleichsweise geringen Beurteilerschulung zum Einsatz kamen. Die Ergebnisse sprechen in diesem Fall deutlich für die Notwendigkeit einer umfassenden Schulung der Beurteiler, bevor die Leistungen der Testpersonen anhand der Beurteilungsbögen eingeschätzt werden.

Ermittelt man die Reliabilität anhand der über die jeweils eingesetzten Beurteiler gemittelten Werte für die Gesamtskala der Beurteilungsbögen, so ergeben sich heterogene Ergebnisse. Die Gesamtskala bestehend aus zehn Items der MMI-Aufgabe *Berufswahl* weist die höchste interne Konsistenz, Cronbachs Alpha, auf. Mit $\alpha_{\text{Cron}} = .81$ erzielt die Gesamtskala bereits sehr zufriedenstellend reliable Ergebnisse. Unter Ausschluss der vier Items mit den geringsten Trennschärfen (Extrinsische Motive, Begegnung der Herausforderungen im Beruf, lehrberufsrelevante persönliche Schwächen und berufliche Alternativen) erreicht die Gesamtskala sogar einen Wert von $\alpha_{\text{Cron}} = .88$. Die MMI-Aufgabe *Bildbeschreibung* erreicht ebenfalls eine sehr zufriedenstellende Reliabilität von $\alpha_{\text{Cron}} = .83$. Lediglich die MMI-Aufgabe *Ausflug in den Zoo* lässt auf Skalenebene bezüglich der Reliabilität zu wünschen übrig. Hier führt auch kein Ausschluss bestimmter Items zu einer verbesserten Skalenreliabilität. In Bezug auf diese MMI-Aufgabe lässt sich jedoch feststellen, dass sie sehr heterogene Leistungsaspekte der Testpersonen erfasst, so weist der Beurteilungsbogen Items zu drei abgrenzbaren Dimensionen auf. Die mittlere Inter-Item-Korrelation beträgt lediglich .22 und stützt die Annahme der relativ unabhängigen Beurteilungsisems.

Insgesamt weisen die drei MMI-Aufgaben unterschiedlich reliable Messungen auf. Die MMI-Aufgabe *Bildbeschreibung* führt insgesamt zu den zuverlässigsten Messergebnissen, sofern die Aufgabe anhand von zwei Beurteilern ausgewertet wird. Die MMI-Aufgabe *Berufswahl* zeigt Schwächen in Bezug auf die Beurteilerübereinstimmung und bedarf möglicherweise einer Überarbeitung des Beurteilungsbogens, so dass dieser zu einheitlicheren Einschätzungen führt. Möglicherweise werden die einzelnen Beurteilungsisems von den Beurteilern unterschiedlich interpretiert oder die Einschätzungen werden von individuellen Urteilsmaßstäben beeinflusst. Die MMI-Aufgabe *Ausflug in den Zoo* weist insgesamt die niedrigste Reliabilität auf. Wie bereits erwähnt, sind in Bezug auf den Beurteilungsbogen besonders umfassende Beurteilerschulungen nötig. Ebenfalls zeigt der Be-

urteilungsbogen auf Itemebene noch Spielraum für Überarbeitungen, um die Beurteilungsmaßstäbe einheitlich verständlich und weniger abhängig von der Expertise des Beurteilers zu gestalten.

Objektivität

Da die Intraklassenkorrelation als Maß für die Beurteilerübereinstimmung verstanden werden kann, lassen sich aus den Kennwerten ebenso Aussagen über die Auswertungsobjektivität ableiten. Geringe Intraklassenkorrelationen sprechen dafür, dass die Beurteiler zu unterschiedlichen Einschätzungen der Leistungen kommen und die Ergebnisse der MMI-Aufgaben somit nicht völlig unabhängig von der auswertenden Person sein können. Bereits in den Vorläufen der eigentlichen Einschätzung anhand der Beurteilungsbögen wurden Kalibrierungsversuche der Urteile anhand der Bögen vorgenommen. An der Kalibrierung der Urteile haben insgesamt zehn Beurteiler teilgenommen und zeigten deutliche Probleme in der einheitlichen Interpretation der Beurteilungsdimensionen in Bezug auf die Beurteilungsbögen der drei MMI-Aufgaben. Erst nach sehr aufwändiger Besprechung und Abgleichung der verschiedenen Einschätzungen fühlten sich die jeweils ausgewählten Beurteiler für die vorliegende Stichprobe überhaupt annähernd sicher genug, mit den Beurteilungsbögen zu arbeiten. Die zwei Hilfskräfte, welche als Krankheitsersatz zur Beurteilung der Leistungen in der MMI-Aufgabe *Ausflug in den Zoo* eingesetzt wurden, waren an den Kalibrierungsbemühungen nicht beteiligt. Möglicherweise ist dies ein weiterer Grund für die wenig zufriedenstellenden Beurteilerübereinstimmungen der MMI-Aufgabe *Ausflug in den Zoo*.

Die Durchführungsobjektivität ist für alle drei MMI-Aufgaben grundsätzlich gegeben. Für jede der drei MMI-Aufgaben lag eine standardisierte Instruktion für den Testleiter vor und die Interaktion mit den Testpersonen wurde so gering wie möglich gehalten. Bei den MMI-Aufgaben *Berufswahl* und *Ausflug in den Zoo* wurden den Testpersonen während der Bearbeitungszeit Fragen durch den Testleiter gestellt. Diese waren alle vorher ausformuliert und festgelegt. Die Fragen der MMI-Aufgabe *Ausflug in den Zoo* hingegen waren größtenteils dem Ermessen des Testleiters unterlegen. So entschied der Testleiter, wann er den Hinweis auf ein bisher nicht genanntes Themengebiet einbrachte, und dies führte häufig zu unterschiedlichen Handhabungen der Hinweise durch die Testleiter. Hier sollten die Hinweisfragen ähnlich wie in der *Berufswahlaufgabe* bestenfalls bereits während der Vorbereitungszeit auf strukturierte Art den Testpersonen präsentiert werden, so dass diese während der Bearbeitungszeit ihre Ausführungen daran ausrichten können.

Validität

Die Validität eines Verfahrens trifft Aussagen darüber, ob das Verfahren auch tatsächlich das Konstrukt erfasst, was es zu messen vorgibt (Kemper et al., 2015). Somit ist dieses Testgütekriterium mit am bedeutendsten, denn so entscheidet das Maß der Validität vor allem darüber, inwiefern die Ergebnisse generalisierbar sind. In der Forschung zu den Multiplen Mini-Interviews im medizinischen Anwendungskontext gibt es bislang nur wenig empirische Befunde zur Validität des Verfahrens. Häufig wird von einer hohen Augenscheinvalidität berichtet, welche jedoch nur auf dem subjektiven Eindruck der Messabsicht beruht. Die MMIs in der Medizin zeigen teilweise eine hohe prognostische Validität, so lässt sich das Abschneiden im MMI positiv mit der Leistung in einem medizinischen Praxistest in Verbindung bringen oder aber mit der Zusage einer medizinischen Hochschule (Jerant et al., 2012; Pau et al., 2013). In der vorliegenden Arbeit sollen vor allem die bislang wenig untersuchten Validitätsarten der Konstrukt- und Kriteriumsvalidität untersucht werden. Hierfür wurden zum einen Globaleinschätzungen der Testleiter über die aufgabenspezifische und -unspezifische Leistung der Testpersonen sowie Selbsteinschätzungen der Testpersonen zu verschiedenen Kompetenz- und Fähigkeitsbereichen und der Persönlichkeit herangezogen.

Zur Bestimmung der Konstruktvalidität untersucht man die Übereinstimmung der Testscores mit einem nomologischen Netz anderer Scores, welche theoretisch verwandt sind. In den Untersuchungshypothesen wurde bereits konstatiert, dass sich ein systematischer Zusammenhang zwischen den durch die Beurteiler eingeschätzten Leistungen in den drei MMI-Aufgaben mit den globalen Einschätzungen der Bewältigung der drei MMI-Aufgaben durch die Testleiter zeigen sollte. In diesem Fall ließ sich die Validität anhand einer Multitrait-Multimethod ähnlichen Matrix über die konvergente und diskriminante Validität bestimmen. Es liegen jedoch nur zwei Methoden vor und die gewonnenen Daten beziehen sich beide auf die gleiche Leistung in den MMI-Aufgaben. Die konvergente Validität, also die Übereinstimmung eines gemessenen Konstrukts über verschiedene Methoden, in diesem Fall die Leistungsbeurteilung durch die Beurteiler und die Globaleinschätzung durch die Testleiter, fällt für die MMI-Aufgaben hoch aus. Die diskriminante Validität ist ebenfalls grundsätzlich gegeben. So korrelierten die Gesamtscores der MMI-Aufgaben untereinander nur gering und nicht signifikant, es ließen sich also nur geringe positive Zusammenhänge der Leistungen in den drei Aufgaben finden. Die durch die Beurteiler eingeschätzten Gesamtleistungen und die Globaleinschätzungen der Bewältigung durch den Testleiter verschiedener MMI-Aufgaben zeigten kaum nennenswerte Zusammenhänge.

Da die drei MMI-Aufgaben unterschiedliche Konstrukte messen, nämlich grundsätzlich Planungskompetenz, Instruktionsklarheit und Erklärungskompetenz sowie Berufswahlkompetenzen, sollten sich diese drei Aufgaben- beziehungsweise Kompetenzfaktoren ebenfalls in einer Faktoren-

analyse ergeben. Eine explorative Hauptkomponentenanalyse identifizierte jedoch insgesamt sechs Faktoren, welche die Varianz der Gesamtheit aller Items aus den drei Beurteilungsbögen erklärten. Unter den sechs Faktoren ließen sich die ersten drei dennoch als aufgabenspezifische Faktoren interpretieren, so luden auf den ersten Faktor nur Items der *Berufswahlaufgabe* hoch, auf den zweiten Faktor nur Items der *Bildbeschreibungsaufgabe* hoch und auf den dritten Faktor nur Items der *Zooaufgabe* hoch. Die faktorielle Validität ist demzufolge nur bedingt gegeben, als dass zum einen nicht alle Items einer MMI-Aufgabe vollständig auf den entsprechenden Aufgabenfaktor laden und zudem drei Faktoren existieren, welche die Varianz unterschiedlicher Aufgabenitems erklären.

Die Kriteriumsvalidität zeigt ein deutlich heterogenes Ergebnisbild. Zwei der drei MMI-Aufgaben berücksichtigen in der Leistungsbeurteilung nicht nur die inhaltlichen Gesichtspunkte der Performanz, sondern beziehen ebenfalls den sprachlichen Ausdruck und die Art der Strukturierung der verbalen Darstellungen mit ein. Hier bedarf es also neben den spezifisch fokussierten Kompetenzen des Instruieren, Erklärens und der Berufswahl ebenfalls sprachlicher Fähigkeiten für die Bewältigung der Aufgaben, welche innerhalb der Beurteilungsbögen auch erfasst wird. Nur für die sprachlichen Einschätzungsdimensionen der *Bildbeschreibung* lässt sich ein signifikanter und mittel positiver Zusammenhang zu den Globaleinschätzungen der sprachlichen Qualität durch die Testleiter feststellen. Je höher die sprachliche Qualität der Ausführungen durch die Testleiter global eingeschätzt wird, desto besser schneiden die Testpersonen in den sprachlichen Leistungsdimensionen der *Bildbeschreibungsaufgabe* ab. Dieser Zusammenhang zeigt sich jedoch nicht signifikant und etwas schwächer in Bezug auf den Gesamtscore in der MMI-Aufgabe *Bildbeschreibung*, so dass die Gesamtleistung nur einen schwachen Zusammenhang zu sprachlichen Qualitäten aufweist. Für die *Berufswahlaufgabe* zeigen sich lediglich schwache und nicht signifikante Zusammenhänge zur sprachlichen Qualität und dies trotz der Tatsache, dass der sprachliche Ausdruck und die Argumentation ausdrücklich in den Einschätzungen berücksichtigt werden.

Da es sich bei den Aufgaben der *Bildbeschreibung* und dem *Ausflug in den Zoo* um sehr komplexe Aufgaben handelt, wird hierfür möglicherweise eine hoch ausgeprägte Auffassungsgabe benötigt, um in der kurzen Zeit die Aufgabenanforderungen zu analysieren beziehungsweise das komplexe Bild aufgabenrelevant zu entschlüsseln. Es zeigt sich jedoch nur ein signifikant positiver Zusammenhang der Gesamtleistung in der *Bildbeschreibungsaufgabe* mit der durch die Testleiter beurteilte Auffassungsgabe. Im Gegensatz zu den beiden anderen MMI-Aufgaben, stellt die *Bildbeschreibung* die einzige Aufgabe dar, in welcher der Testleiter die Aufgabenbearbeitung nicht durch Fragen unterstützen kann. Somit ist es insbesondere in dieser Aufgabe von Bedeutung, die Aufgabe rasch und umfassend zu durchdringen.

Die von Jerant et al. (2012) gefundenen Verbindungen der MMI-Leistungen mit Persönlichkeitsfaktoren konnte hier bezüglich Extraversion, emotionaler Verträglichkeit und Stabilität nicht gefunden werden. Keiner der Persönlichkeitsfaktoren trug dazu bei, die Varianz der Gesamtleistung über alle MMI-Aufgaben aufzuklären. In der Forschung zur Anwendung des MMIs im medizinischen Fachbereich wird häufig betont, dass die multiplen und aufeinander folgenden Aufgaben den Einfluss von Nervosität mindern können. Denn bei jeder neuen Aufgabe kann durch vermehrte Anstrengung ein zuvor schlechtes Ergebnis ausgeglichen werden und nimmt somit den Leistungsdruck und schwächt Effekte der Aufgeregtheit ab (Bingham & Scharf, 2011; Eva et al., 2004c; Pau et al., 2013). Auch die Gesamtleistung in Bezug auf alle drei MMI-Aufgaben weist keinen Zusammenhang mit der durch die Testleiter eingeschätzte Aufgeregtheit der Testpersonen auf.

Überraschenderweise konnten keine Zusammenhänge der Gesamtscores in den drei MMI-Aufgaben mit den Selbsteinschätzungen der aufgabenrelevanten Kompetenzbereiche nachgewiesen werden. Es zeigen sich Korrelationen nahe 0 für den Gesamtscore in der *Berufswahl* und der Selbsteinschätzung der Berufswahlkompetenz und für den Gesamtscore der *Bildbeschreibungsaufgabe* und der Selbsteinschätzung der Erklärungskompetenz. Ausschließlich die Selbsteinschätzung der Planungskompetenz zeigt einen gering positiven, jedoch nicht signifikanten, Zusammenhang mit der Leistung in der MMI-Aufgabe *Ausflug in den Zoo*. Höhere Einschätzungen der Planungskompetenz gehen also mit höheren Leistungen in der MMI-Aufgabe *Ausflug in den Zoo* einher. Der einzig signifikant positive Zusammenhang zwischen Selbsteinschätzung und Leistung in einer MMI-Aufgabe besteht für die Einschätzung der Berufswahlkompetenz und der Leistung in der MMI-Aufgabe *Ausflug in den Zoo*. Hohe Einschätzungen der Berufswahlkompetenzen gehen mit hohen Leistungen in dieser MMI-Aufgabe einher.

Da außer bezüglich der Planungs- und Berufswahlkompetenz und der *Zooaufgabe* keinerlei Zusammenhänge zwischen Selbsteinschätzungen und den Gesamtleistungen in den MMI-Aufgaben nachgewiesen werden konnten und die Selbsteinschätzungen der aufgabenspezifischen Kompetenzen auch keine Varianz der MMI-Gesamtleistung aufklären, lassen sich in diesem Kontext keine eindeutigen Aussagen zur Validität treffen. Wie in Kapitel 2.3.3 bereits angedeutet, ist die Verlässlichkeit von Selbsteinschätzungen zur Messung von Fähigkeiten und Kompetenzen stark durch Erinnerungsverzerrungen und selbstwertdienlichen Tendenzen beeinträchtigt. Es bleibt zu bezweifeln, inwiefern Studierende in der Lage sind, realistische Vorstellungen ihrer eigenen Kompetenzen zu entwickeln. Zum einen können die hier berücksichtigten MMI-Aufgaben doch andere Kompetenzbereiche erfassen, als das Verfahren es vorgibt und die Studierenden bezüglich der persönlichen Ausprägungen somit eingeschätzt haben, zum anderen mögen die Selbsteinschätzungen keinerlei systematische Übereinstimmungen mit den tatsächlich vorhandenen Kompetenzausprägungen

aufweisen. Die globalen Eindrücke der Testleiter bezüglich der aufgabenspezifischen Leistungen hingegen stimmen positiv mit den anhand der Beurteilungsbögen ermittelten Gesamtscores überein.

Insgesamt zeichnet sich auch für die Anwendung der MMIs im Lehrberuflichen Kompetenzbereich kein homogenes Bild der Validität ab. Hierfür müssten die Messergebnisse der drei MMI-Aufgaben mit weiteren, bereits als valide eingeschätzten Messungen der gleichen Kompetenzkonstrukte verglichen werden. Bis auf die Selbsteinschätzungen der Testpersonen, bei welchen Verzerrungen nicht auszuschließen sind, wurden zur Validierung der MMI-Aufgaben ausschließlich Einschätzungen verschiedener Variablen berücksichtigt, welche sich allesamt auf das in den MMI-Aufgaben gezeigte Verhalten bezogen.

Ökonomie

Trotz der knapp bemessenen Zeit pro MMI-Aufgabe von zwei Minuten Vorbereitungs- und acht Minuten Bearbeitungszeit konnte eine große Varianz an Leistungen der Testpersonen produziert werden. Häufig nutzten die Testpersonen die acht Minuten für die Bearbeitung der Aufgabe nicht vollständig aus und beendeten ihre Ausführungen schon vor Ablauf der vorgegebenen Zeit. 30 Minuten zur Bearbeitung von drei Aufgaben erweisen sich für eine Testperson nicht allzu zeitintensiv, jedoch bedarf es bei der Erhebung von einer Vielzahl von Studierenden enormer zeitlicher, finanzieller, räumlicher und personaler Ressourcen, denn so muss für jede Testperson ein Testleiter und insgesamt genügend Räumlichkeiten zur Verfügung stehen. Neben der reinen Durchführungszeit muss zudem die Auswertungszeit und der Einsatz von mindestens zwei Beurteilern bedacht werden. Ergebnisse der Beurteilerübereinstimmungen wiesen bereits im Kontext der Reliabilität darauf hin, dass die Auswertung mit nur zwei oder drei Beurteilern zwar zu personalen Einsparungen führt, jedoch umfassende und zeitlich vorgelagerte Kalibrierungsprozesse und Schulungen bedarf, um eine hohe Messgenauigkeit zu garantieren. Die Entwicklung und Konzeption theoretisch fundierter MMI-Aufgaben mit entsprechend hochwertigen Beurteilungsbögen, welche in der Anwendung den hohen Qualitätsansprüchen der Testgüte genügen können, muss aus Perspektive dieser Arbeit definitiv als ein alle Maße übersteigend aufwändiger Prozess eingeschätzt werden.

Anhand der Ergebnisse dieser Arbeit lässt sich noch kein globales Fazit über die psychometrische Qualität von Multiplen Mini-Interviews im Anwendungskontext Lehrberuflicher Kompetenzdiagnostik ziehen. Die Testgüte des gesamten Verfahrens steht und fällt mit der theoretischen Konzeption und präzisen Entwicklung von inhaltlich eindeutigen Beurteilungsbögen zu den jeweiligen MMI-Aufgaben. So zeigten die drei MMI-Aufgaben, welche in dieser Arbeit berücksichtigt worden,

in Bezug auf die Reliabilität, Objektivität und Validität heterogene Ergebnisse, so dass Aussagen zu der Güte des MMIs nur aufgabenspezifisch erfolgen kann. Von einer globalen Bewertung des MMIs zur Diagnostik lehrberufsrelevanter Kompetenzen, wie es im medizinischen Kontext in der Regel schon erfolgt, scheint das Verfahren bislang durch die starke Aufgabenabhängigkeit der Qualität noch weit entfernt zu sein. Das Gelingen der Anwendung Multipler Mini-Interviews zur Diagnostik von lehrberuflichen Kompetenzen erweist sich ausgehend von den Ergebnissen dieser Arbeit noch stark von den einzelnen konzipierten MMI-Aufgaben abhängig. So ökonomisch die spätere Anwendung des MMIs im Vergleich zu klassischen Interviews auch ausfallen mag, darf der Konzeptionsaufwand zur Sicherstellung der psychometrischen Qualität des Verfahrens nicht außer Acht gelassen und an dem späteren Nutzen des Verfahrens abgewogen werden.

Weitere Forschungsbemühungen zur Konzeption und Anwendung von Multiplen Mini-Interviews bezüglich Diagnostik von Kompetenzen des Lehrberufs sollten bereits während der Aufgabenkonstruktion die psychometrischen Gütekriterien umfassend berücksichtigen. Erst dann kann der Schritt gewagt werden, anhand der Messungen des MMIs Kompetenzrückmeldungen für Lehramtsstudierende abzuleiten.

Literaturverzeichnis

- Abujatum, M., Arold, H., Knispel, K., Rudolf, S. & Schaarschmidt, U. (2007). Intervention durch Training und Beratung. In U. Schaarschmidt & U. Kieschke (Hrsg.), *Gerüstet für den Schulalltag: Psychologische Unterstützungsangebote für Lehrerinnen und Lehrer* (S. 117-156). Weinheim: Beltz.
- Aeschbacher, U. (2009). Eine Lanze für das Erklären. *Beiträge zur Lehrerbildung*, 27(3), 431.
- Affolter, B., Hollenstein, L. & Brühwiler, C. (2015). Unsere zukünftigen Lehrpersonen: Idealistisch, realistisch oder selbstbewusst pragmatisch. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 33(1), 69.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2011). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (13. Auflage). Berlin, Heidelberg: Springer.
- Bauer, J. (2009). Burnout bei schulischen Lehrkräften. *PiD - Psychotherapie im Dialog*, 10(3), 251-255.
- Bauer, K. (1998). Pädagogisches Handlungsrepertoire und professionelles Selbst von Lehrerinnen und Lehrern. *Zeitschrift für Pädagogik*, 44(3), 343.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 520.
- Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29-53). Münster: Waxmann.
- Biedermann, H., Brühwiler, C. & Krattenmacher, S. (2012). Lernangebote in der Lehrerausbildung und Überzeugungen zum Lehren und Lernen: Beziehungsanalysen bei angehenden Lehrpersonen. *Zeitschrift für Pädagogik*, 58(4), 475.
- Bieri Buschor, C. & Schuler Braunschweig, P. (2011). Check-point Assessment Centre für angehende Lehramtsstudierende: Empirische Befunde zur prognostischen Validität und zur Übereinstimmung von Selbst- und Fremdeinschätzung eignungsrelevanter Merkmale. *Zeitschrift für Pädagogik*, 57(5), 695.
- Bieri Buschor, C., Schuler Braunschweig, P. & Strinemann Wolf, B. (2006). Assessment-Centers als Aufnahmeverfahren für pädagogische Hochschulen? *Beiträge zur Lehrerbildung*, 24(1), 55.
- Bingham, R. & Scharf, P. (2011). Preparing students for multiple mini interviews. *The Advisor*, 19-21.

- Blömeke (2011). Zum Verhältnis von Fachwissen und unterrichtsbezogenen Überzeugungen bei Lehrkräften im internationalen Vergleich. *Stationen Empirischer Bildungsforschung*, 395-411.
- Blömeke, S., Olsen, R.V. & Suhl U. (2016). Relation of Student Achievement to the Quality of Their Teachers and Instructional Quality. In T. Nilsen & J.-E. Gustafsson (Hrsg.), *Teacher Quality, Instructional Quality and Student Outcomes* (S. 21-50). IEA, SpringerOpen.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler: Mit 87 Tabellen* (4., überarb. Aufl.). Heidelberg: Springer.
- Bromme, R. & Haag, L. (2008). Forschung zur Lehrerpersönlichkeit Handbuch der Schulforschung, 803-819.
- Campbell, D. & Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.
- Clauser, B. E. (2000). Recurrent issues and recent advances in scoring performance assessments. *Applied Psychological Measurement*, 24(4), 310-324.
- Combe, A. & Kolbe, F. (2008). Lehrerprofessionalität: Wissen, Können, Handeln. *Handbuch der Schulforschung*, 857-875.
- Costa, P. T., Jr. & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa: Psychological Assessment Resources.
- Dodson, M., Crotty, B., Prideaux, D., Carne, R., Ward, A. & de Leeuw, E. (2009). The multiple mini-interview: How long is long enough? *MEDICAL EDUCATION*, 43(2), 168-174.
- Eva, K., Reiter, H., Rosenfeld, J. & Norman, G. (2004). The relationship between inter-viewers' characteristics and ratings assigned during a multiple mini-interview. *Academic Medicine*, 79(6), 602-609.
- Eva, K., Reiter, H., Rosenfeld, J., Trinh, K., Wood, T. & Norman, G. (2012). Association between a medical school admission process using the multiple mini-interview and national licensing examination scores. *JAMA*, 308(21), 2233-2240.
- Eva, K., Rosenfeld, J., Reiter, H. & Norman, G. (2004c). An admissions OSCE: The multiple mini-interview. *Medical Education*, 38(3), 314-326.

- Fisseni, H. (2004). *Lehrbuch der psychologischen Diagnostik: Mit Hinweisen zur Intervention* (3., überarb. und erw. Aufl.). Göttingen: Hogrefe.
- Frey, A. (2004). Die Kompetenzstruktur von Studierenden des Lehrerberufs.: Eine inter-nationale Studie. *Zeitschrift für Pädagogik*, 50(6), 903-925.
- Frey, A. (2006). Methoden und Instrumente zur Diagnose beruflicher Kompetenzen von Lehrkräften: Eine erste Standortbestimmung zu bereits publizierten Instrumenten. *Zeitschrift für Pädagogik*, 51, Beiheft, 30-46.
- Frey, A. (2008). *Kompetenzstrukturen von Studierenden in der ersten und zweiten Phase der Lehrerbildung: Eine nationale und internationale Standortbestimmung*. 1. Aufl. Landau: Verlag Empirische Pädagogik.
- Frey, A. & Jung, C. (2011). Kompetenzmodelle und Standards in Lehrerbildung und Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 540-572). Münster: Waxmann.
- Gartmeier, M., Bauer, J., Fischer, M. R., Karsten, G. & Prenzel, M. (2011). Modellierung und Assessment professioneller Gesprächsführungskompetenz von Lehrpersonen im Lehrer-Elterngespräch. In O. Zlatkin-Troitschanskaia (Hrsg.), *Stationen Empirischer Bildungsforschung. Traditionslinien und Perspektiven* (S. 412-426). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Girmes, R. (2006): Lehrerprofessionalität in einer demokratischen Gesellschaft. Über Kompetenzen und Standards in einer erziehungswissenschaftlich fundierten Lehrerbildung. *Zeitschrift für Pädagogik*, 51, Beiheft, 14-29.
- Göppel, R. (2013). Was macht eine gute Lehrerin, einen guten Lehrer aus? – Was kennzeichnet gute LehrerInnenbildung? Und was sollen Lehramtsstudierende speziell im erziehungswissenschaftlichen Studium lernen? *Research on Steiner Education*, 3(2), 68-82.
- Hascher, T. (2004). Professionelle Entwicklung von Lehrpersonen. *Journal für Lehrerinnen- und Lehrerbildung*, 1, 44-50.
- Hecker, K., Donnon, T., Fuentealba, C., Hall, D., Illanes, O., Morck, D. & Muelling, C. (2009). Assessment of applicants to the veterinary curriculum using a multiple miniinterview method. *Journal of Veterinary Medical Education*, 36(2), 166-173.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Klett-Kallmeyer.

- Hissbach, J. C., Sehner, S., Harendza, S. & Hampe, W. (2014). Cutting costs of multiple mini-interviews – changes in reliability and efficiency of the Hamburg medical school admission test between two applications. *BMC medical education*, 14(1), 54.
- Höft, S. & Funke, U. (2006). Simulationsorientierte Verfahren in der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (2. Aufl., S. 145-188). Göttingen: Hogrefe Verlag.
- Jerant, A., Griffin, E., Rainwater, J., Henderson, M., Sousa, F., Bertakis, K., Fenton, J. & Franks, P. (2012). Does applicant personality influence multiple mini-interview performance and medical school acceptance offers? *Academic Medicine*, 87(9), 1250-1259.
- Jerusalem, M., Drössler, S., Kleine, D., Klein-Heßling, J., Mittag, W. & Röder, B. (2007). *Selbstwirksamkeit und Selbstbestimmung im Unterricht*. Berlin: Humboldt-Universität zu Berlin, Lehrstuhl für Pädagogische Psychologie und Gesundheitspsychologie.
- Kanning, U. P. (2009a). *Diagnostik sozialer Kompetenzen* (2., aktualisierte Auflage). Göttingen: Hogrefe.
- Keller-Schneider, M. (2011). Lehrer/in werden – eine Entwicklungsaufgabe: Kompetenzentwicklung in der Auseinandersetzung mit Wissen und Überzeugungen. *PADUA*, 6(4), 6-14.
- Kemper, C., Ziegler, M., Krumm, S., Heene, M. & Bühner, M. (2015). Testkonstruktion. In G. Stemmler & J. Margraf-Stiksrud (Hrsg.), *Lehrbuch psychologische Diagnostik* (1. Aufl.). *Psychologie-Lehrbuch* (S. 157-222). Bern: Verlag Hans Huber.
- Kleinmann, M. (1997). *Assessment-Center: Stand der Forschung - Konsequenzen für die Praxis*. Göttingen: Verl. für Angewandte Psychologie.
- Knorr, M. & Hissbach, J. (2014). Multiple mini-interviews: same concept, different approaches. *Medical Education*, 48(12), 1157-1175.
- Kolbe, F. & Combe, A. (2008). Lehrerbildung. In W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (S. 877-901). Wiesbaden: VS.
- König, J. & Blömeke, S. (2009). Pädagogisches Wissen von angehenden Lehrkräften: Erfassung und Struktur von Ergebnissen der fachübergreifenden Lehrerausbildung. *Zeitschrift für Erziehungswissenschaft*, 12(3), 499-527.
- König, J. & Seifert, A. (2012). *Lehramtsstudierende erwerben pädagogisches Professionswissen: Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerausbildung*. Münster: Waxmann.

- König, J., Kaiser, G. & Felbrich, A. (2012). Spiegelt sich pädagogisches Wissen in den Kompetenzselbsteinschätzungen angehender Lehrkräfte?: Zum Zusammenhang von Wissen und Überzeugungen am Ende der Lehrerausbildung. *Zeitschrift für Pädagogik*, 58(4), 476.
- Landert, C. (2007). *Die Berufszufriedenheit der Deutschschweizer Lehrerinnen und Lehrer. Bericht über die Onlinebefragung vom September/Oktober 2006*. Zürich. Zuletzt aufgerufen am 17.12.2017 unter https://www.l-gl.ch/fileadmin/files/documents/Positionspapiere/061209_Berufszufriedenheit_Lehrerinnen.pdf
- Lehrberg, R. & Lüth, E. (2012). Hamburger Mindeststandards. Eignungsberatung und Kompetenzentwicklung von Lehramtsstudierenden. In B. Weyand, M. Justus & M. Schratz (Hrsg.), *Auf unsere Lehrerinnen und Lehrer kommt es an. Geeignete Lehrer/innen gewinnen, (aus-)bilden und fördern* (S. 80-85). Essen: Edition Stifterverband.
- Linn, R. L., Baker, E. L. & Dunbar, S. B. (1991). Complex, performance-based assesement: Expectations and validation criteria. *Educational Researcher*, 20(8), 15-21.
- Lipowsky, F. (2006): Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. *Zeitschrift für Pädagogik*, 51, Beiheft, 47-70.
- Lipowsky, F. (2007). Was wissen wir über guten Unterricht?: Im Fokus: die fachliche Lernentwicklung. *Friedrich-Jahresheft*, 25, 26-29.
- Mayr, J. (2009). *Studien- und Berufswahlmotive von Lehrpersonen. Wie sie entstehen, wie sie sich verändern und was sie bewirken*. Kommentierte Folien zum Referat bei der 5. Tagung der Sektion „Empirische Bildungsforschung“ der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE). Landau. Zuletzt aufgerufen am 10.01.2018 unter https://www.researchgate.net/profile/Johannes_Mayr2/publication/277020758_Studien-und_Berufswahlmotive_von_Lehrpersonen_Wie_sie_entstehen_wie_sie_sich_verandern_und_was_sie_bewirken/links/555f772008ae8c0cab30b2ab.pdf
- Mayr, J. (2012). Ein Lehramtsstudium beginnen? Ein Lehramtsstudium beginnen lassen? Laufbahnberatung und Bewerberauswahl konstruktiv gestalten. In B. Weyand, M. Justus & M. Schratz (Hrsg.), *Auf unsere Lehrerinnen und Lehrer kommt es an. Geeignete Lehrer/innen gewinnen, (aus-)bilden und fördern* (S. 38-57). Essen: Edition Stifterverband.
- McAndrew, R. & Ellis, J. (2012). An evaluation of the multiple mini-interview as a selection tool for dental students. *BRITISH DENTAL JOURNAL*, 212(7), 331-335.

- Messner, H. (2007). Vom Wissen zum Handeln – vom Handeln zum Wissen. Zwei Seiten einer Medaille. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 25 (3), 364-376.
- Moser, K. (2004). Selbstbeurteilung. In H. Schuler (Hrsg.), *Beurteilung und Förderung beruflicher Leistung* (S. 83-100). Göttingen: Hogrefe.
- Mulder, R. & Gruber, H. (2011). Die Lehrperson im Lichte von Professions-, Kompetenz- und Expertiseforschung – die drei Seiten einer Medaille. *Stationen Empirischer Bildungsforschung*, 427-438.
- Nieskens, B. (2012). Einblicke in die Praxis: Trends und Verfahren der Eignungsabklärung, Beratung und Potenzialentwicklung in der Lehrerausbildung. In B. Weyand, M. Justus & M. Schratz, (Hrsg.), *Auf unsere Lehrerinnen und Lehrer kommt es an: Geeignete Lehrer/-innen gewinnen, (aus-)bilden und fördern* (S. 124-177). Essen: Edition Stifterverband.
- Oser, F. (2006). Zugänge ermöglichen, Zugänge verwehren: Entwurf eines Ausleseverfahren in der Lehrerbildung (ein Essay). *Beiträge zur Lehrerbildung*, 24(1), 30.
- Oser, F., Curcio, G. & Düggeli, A. (2007). Kompetenzmessung in der Lehrerbildung als Notwendigkeit.- Fragen und Zugänge. *Beiträge zur Lehrerbildung*, 25(1), 14-26.
- Palm, T. (2001). *Performance and authentic assessment, realistic and real life tasks: A conceptual analysis of the literature*. Department of educational measurement. Umeå universitet.
- Päßler, K., Hell, B. & Schuler, H. (2011). Grundlagen der Berufseignungsdiagnostik und ihre Anwendung auf den Lehrerberuf. *Zeitschrift für Pädagogik*, 57(5), 639.
- Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F. & Cleland, J. (2016). How effective are selection methods in medical education? A systematic review. *Medical Education*, 50, 36-60.
- Pau, A., Jeevaratnam, K., Chen, Y. S., Fall, A. A., Khoo, C. & Nadarajah, V. D. (2013). The multiple mini-interview (MMI) for student selection in health professions training – A systematic review. *Medical Teacher*, 35, 1027-1041.
- Preuß, E. (2008). Lehrerkompetenzen zur Gestaltung von Innerer Differenzierung. In W. Plöger (Hrsg.), *Was müssen Lehrerinnen und Lehrer können? Beiträge zur Kompetenzorientierung in der Lehrerbildung* (S. 209-224). Paderborn: Schöningh.
- Rauin, U. & Meier, U. (2007): Subjektive Einschätzungen des Kompetenzerwerbs in der Lehramtsausbildung. In M. Lüders & J. Wissinger (Hrsg.), *Forschung zur Lehrerbildung. Kompetenzentwicklung und Programmevaluation* (S. 103-136). Münster: Waxmann.

- Rees, E. L., Hawarden, A. W., Dent, G., Hays, R., Bates, J. & Hassell, A. B. (2016). Evidence regarding the utility of multiple mini-interview (MMI) for selection to undergraduate health programs: A BEME systematic review: BEME Guide No. 37. *Medical Teacher*, 38, 443-455.
- Reiter, H., Eva, K., Rosenfeld, J. & Norman, G. (2007). Multiple mini-interviews predict clerkship and licensing examination performance. *Medical Education*, 41(4), 378-384.
- Richardson, P. & Watt, H. (2006). Who Chooses Teaching and Why? Profiling Characteristics and Motivations Across Three Australian Universities. *Asia-Pacific Journal of Teacher Education*, 34(1), 27-56.
- Rosenfeld, J. M., Reiter, H. I., Trinh, K. & Eva, K. W. (2008). A cost efficiency comparison between the multiple mini-interview and traditional admissions interviews. *Advances in Health Sciences Education*, 13, 43-58.
- Rothland, M. & Terhart, E. (2011). Eignungsabklärung angehender Lehrerinnen und Lehrer. Einführung in den Thementeil. *Zeitschrift für Pädagogik*, 57(5), 635.
- Rothland, M. (2010). Soziale Kompetenz: angehende Lehrkräfte, Ärzte und Juristen im Vergleich: Empirische Befunde zur Kompetenzausprägung und Kompetenzentwicklung im Rahmen des Studiums. *Zeitschrift für Pädagogik*, 56(4), 603.
- Rothland, M. (2013). "Riskante" Berufswahlmotive und Überzeugungen von Lehramtsstudierenden. *Erziehung & Unterricht*, 163(1/2), 71.
- Schaarschmidt, U. & Kieschke, U. (2007). Einführung und Überblick. In U. Schaarschmidt & U. Kieschke, (Hrsg.), *Gerüstet für den Schulalltag: Psychologische Unterstützungsangebote für Lehrerinnen und Lehrer* (S. 17-44). Weinheim: Beltz.
- Schaarschmidt, U. (2012). Eignung für den Lehrerberuf frühzeitig erkennen und kontinuierlich fördern. Implikationen aus den Potsdamer Lehrerstudien. In B. Weyand, M. Justus & M. Schratz, (Hrsg.), *Auf unsere Lehrerinnen und Lehrer kommt es an: Geeignete Lehrer/-innen gewinnen, (aus-)bilden und fördern* (S. 58-77). Essen: Edition Stifterverband.
- Schaarschmidt, U., Arold, H., Kieschke, U. (2000). *Die Bewältigung psychischer Anforderungen durch Lehrkräfte. Information über ein Forschungsprojekt an der Universität Potsdam*. Manuskript. Zuletzt aufgerufen am 04.01.2018 unter <https://www.bug-nrw.de/fileadmin/web/Lehrergesundheit/Schaarschmidt-2000%20stress-bei-lehrern.pdf>

- Schratz, M. (2012). Auf den Lehrer/die Lehrerin kommt es an. In B. Weyand, M. Justus & M. Schratz, (Hrsg.), *Auf unsere Lehrerinnen und Lehrer kommt es an: Geeignete Lehrer/-innen gewinnen, (aus-)bilden und fördern* (S. 180-192). Essen: Edition Stifterverband.
- Schuler, H. (2004). *Beurteilung und Förderung beruflicher Leistung*. Göttingen: Hogrefe.
- Serres, M. & Peeters, M. (2012). Overcoming content specificity in admission interviews: the next generation? *American Journal of Pharmaceutical Education*, 76(10), 207.
- Sieland, B. & Tacke, M. (2000). *Ansätze zur Förderung der Gesundheit und Leistungsfähigkeit dienstälterer Lehrkräfte in Niedersachsen*. Unveröffentlichter Abschlussbericht. Universität Lüneburg. Zuletzt aufgerufen am 27.12.2017 unter <http://opus.uni-lueneburg.de/opus/volltexte/2001/17/pdf/sieland9.pdf>
- Sieland, B. (2001). *Ansätze zur Förderung der Gesundheit und Leistungsfähigkeit dienstälterer Lehrkräfte in Niedersachsen*. Universität Lüneburg.
- Sliwka, A. & Klopsch, B. (2012). Auf den Lehrer/die Lehrerin kommt es an. Lehrerbildung und Lehrerprofessionalität in internationaler Perspektive. In B. Weyand, M. Justus & M. Schratz (Hrsg.), *Auf unsere Lehrerinnen und Lehrer kommt es an. Geeignete Lehrer/-innen gewinnen, (aus-)bilden und fördern* (S. 14-35). Essen: Edition Stifterverband.
- Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK). (2014). *Standards für die Lehrerbildung: Bildungswissenschaften*. Beschluss der Kultusministerkonferenz vom 16. 12. 2004. i. d. F. vom 12.06.2014.
- Staub, F. C. & Stern, E. (01.06.2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains: Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology*, 94(2), 344.
- Stemmler, G. & Margraf-Stiksrud, J. (Hrsg.). (2015). *Lehrbuch psychologische Diagnostik* (1. Aufl.). *Psychologie-Lehrbuch*. Bern: Verlag Hans Huber.
- Stiggins, R., Arter, J., Chappuis, J. & Chappuis, S. (2004). Assess how? Designing assessments to do what you want. In J. Chappuis, R. Stiggins, S. Chappuis, & J. Arter (Hrsg.), *Classroom assessment for student learning. Doing it right-using it well* (S. 89-121). Portland: Assessment Training Institute.
- Stronge, J. H. & Hindman, J. L. (2003). Hiring the best teachers. *Educational Leadership*, 60(8), 48-52.

- 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.
- Thomson, M. M., Turner, J. E. & Nietfeld, J. L. (2012). A typological approach to investigate the teaching career decision: Motivations and beliefs about teaching of prospective teacher candidates. *Teaching and Teacher Education*, 28(3), 324-335.
- Tiller, D., O'Mara, D., Rothnie, I., Dunn, S., Lee, L. & Roberts, C. (2013). Internet-based multiple mini-interviews for candidate selection for graduate entry programmes. *Medical Education*, 47, 801-810.
- Tschannen-Moran, M., Hoy, A. W. & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of educational research*, 68(2), 202-248.
- Voss, T., Kunina-Habenicht, O., Hoehne, V. & Kunter, M. (2015). Stichwort Pädagogisches Wissen von Lehrkräften: Empirische Zugänge und Befunde. *Zeitschrift für Erziehungswissenschaft*, 18(2), 187-223.
- Watt, H. & Richardson, P. (2007). Motivational Factors Influencing Teaching as a Career Choice: Development and Validation of the FIT-Choice Scale. *The Journal of Experimental Education*, 75(3), 167-202.
- Weiß, S., Schramm, S. & Kiel, E. (2014). Was sollen Lehrerinnen und Lehrer können? Anforderungen an den Lehrer/innenberuf aus Sicht von Lehrkräften und Ausbildungspersonen. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 15(3).
- Weyand, B. (2012). „Drum prüfe wer sich ewig bindet“. Gelingende Passung zwischen Person und Beruf. In B. Weyand, M. Justus & M. Schratz, (Hrsg.), *Auf unsere Lehrerinnen und Lehrer kommt es an: Geeignete Lehrer/-innen gewinnen, (aus-)bilden und fördern* (S. 86-109). Essen: Edition Stifterverband.
- Wilbert, J. & Gerdes, H. (2007). Lehrerbild von Schülern und Lehrern: Eine empirische Studie zum Vergleich der Vorstellungen vom idealen und vom typischen Lehrer. *Psychologie in Erziehung und Unterricht*, 54(3), 222.
- Wiley, D. E. & Haertel, E. H. (1996). Extended assessment tasks: Purposes, definitions, scoring, and accuracy. In M. Kane & R. Mitchell (Hrsg.), *Implementing performance assessment: Promises, problems, and challenges* (S. 61-90). Mahwah, NJ: Lawrence Erlbaum Associates.

- Wirtz, M. A. & Caspar, F. (2004). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.
- Wittwer, J. & Renkl, A. (2008). Why Instructional Explanations Often Do Not Work: A Framework for Understanding the Effectiveness of Instructional Explanations. *Educational Psychologist*, 43(1), 49-64.

Abbildungsverzeichnis

Abbildung 1. Hierarchisches Strukturmodell von Handlungskompetenz nach Frey (2008). Darstellung des Modells ergänzt durch Zusammenfassungen aus Frey & Jung (2011).	7
Abbildung 2. Modell professioneller Handlungskompetenz nach Baumert & Kunter (2006).	8
Abbildung 3. Ablauf der MMI-Aufgaben zu beiden Erhebungsteilen.	47
Abbildung 4. Mittelwerte der drei Beurteiler über alle Items und Testpersonen. Die Fehlerbalken repräsentieren 95 %-Konfidenzintervalle.	74

Tabellenverzeichnis

Tabelle 1	28
Tabelle 2	48
Tabelle 3	58
Tabelle 4	58
Tabelle 5	59
Tabelle 6	63
Tabelle 7	64
Tabelle 8	65
Tabelle 9	66
Tabelle 10	67
Tabelle 11	67
Tabelle 12	68
Tabelle 13	68
Tabelle 14	69
Tabelle 15	70
Tabelle 16	71
Tabelle 17	72
Tabelle 18	72
Tabelle 19	73
Tabelle 20	74
Tabelle 21	75
Tabelle 22	76
Tabelle 23	77
Tabelle 24	78
Tabelle 25	80
Tabelle 26	81
Tabelle 27	82
Tabelle 28	83
Tabelle 29	84
Tabelle 30	84

Anhänge

Die Anhänge sind dieser Masterarbeit nicht beigefügt, um die Aufgaben und ihre Bewertung vor Bekanntwerden bei der Zielgruppe zu schützen. Wir bitten um Verständnis.

Fachkolleginnen und -kollegen, die die Materialien gerne einsehen würden, werden gebeten, sich an die Autorin oder den Herausgeber zu wenden.