

# **Zur Validität von Evaluationsergebnissen**

## Qualitätsdimensionen der Lehrerbildung im Lichte zentraler Testgütekriterien

Philipp Pohlenz, Elisabeth Grindel, Anne Köpke

**Abstract:** In der Praxis der Lehrevaluation werden subjektive Urteile und Qualitätseinschätzungen zu Lehrveranstaltungen oder Studienprogrammen erhoben, um die Sicht der „betroffenen“ Teilnehmer und Teilnehmerinnen zu beschreiben und diese Sicht zum Ausgangspunkt für die Entwicklung von etwaigen Verbesserungsstrategien zu machen. Dabei muss aber berücksichtigt werden, dass subjektive Urteile Quellen potenzieller Verzerrungen der erhobenen Daten enthalten, woraus sich mögliche Bedrohungen ihrer Validität ergeben. Diese beziehen sich auf sachfremde Einflüsse bspw. von Persönlichkeitsmerkmalen der beurteilenden Personen auf ihr Urteil.

Ein weiterer zentraler Validitätsaspekt von Qualitätsmessungen besteht in der Frage, ob die als relevant unterstellten Beurteilungsdimensionen (bspw. zur „allgemeinen Organisationsqualität“) den Qualitätseinschätzungen der beurteilenden Personen tatsächlich zu Grunde liegen. Lassen sich solche Qualitätsdimensionen empirisch bestätigen, so können die entsprechend validierten Evaluationsergebnisse zu einer rational basierten Beurteilung und Weiterentwicklung des Beurteilungsgegenstandes – hier: der zweiten Phase der Lehrerbildung – beitragen.

### **1. Einleitung**

Die Relevanz wissenschaftlicher Messungen wird insbesondere vor dem Hintergrund zentraler Testgütekriterien wie der Validität im Sinne einer gültigen Zuordenbarkeit der Messungen zu den entsprechenden Eigenschaften des zu messenden Gegenstandes beurteilt [BORTZ U. DÖRING 2002; KROMREY 1998].<sup>1</sup> Validität als Kriterium der Güte von Messungen

---

<sup>1</sup> Neben der Validität gelten Objektivität und Reliabilität von Messungen als weitere zentrale Testgütekriterien [z.B. STEYER U. EID 1993]. Im Kontext des vorliegenden Beitrags ist jedoch die Frage der Validität am vordringlichsten. Diese wird daher exemplarisch für Testgütekriterien von Messungen behandelt.

ist insbesondere auch in der Evaluationsforschung und der Lehrevaluation, die sich mit der Beurteilung der Qualität von Lehre und Studium an Hochschulen befasst, von zentraler Bedeutung.<sup>2</sup> In der Praxis der Lehrevaluation stützt sich die Beurteilung der Studienqualität vielfach auf diesbezügliche studentische Einschätzungen. Gleichzeitig entzieht sich die Qualität von Hochschulleistungen jedoch einer direkten Messung. Es gibt vielmehr zahlreiche unterschiedliche, sich teils sogar widersprechende Qualitätskonzeptionen [WISSENSCHAFTSRAT 1996] sowie erwünschte bzw. unerwünschte Effekte, die mit einem Studium verbunden sind [ENGEL U. KREKELER 2001; WISSENSCHAFTSRAT 1996]. So lässt sich bspw. Studienerfolg kaum eindeutig konzeptualisieren und selbst wenn man das erfolgreiche Abschließen eines Studienganges als augenscheinlichstes Erfolgskriterium zu Grunde legt, lässt sich dieses nicht eindeutig dem Einwirken des Studienganges und seiner unterschiedlichen Programmelemente in Studium und Lehre zuordnen [vgl. z.B. HAGER 2000]. Vielmehr bringen Studierende unterschiedliche Lerndispositionen mit, verbinden verschiedene Absichten und Erwartungen mit dem Studium, die ihrerseits zu unterschiedlichen Lernbiographien führen usw. Studentische Qualitätsurteile – und sofern man die zweite Phase der Lehrerbildung als einen Teil des entsprechenden Studienprogramms auffasst, ist die Situation hier identisch – bergen die Gefahr, dass die Subjektivität dieser Urteile, die sich in individuellen Beurteilungsperspektiven und einem potenziell urteilswirksamen Einfluss sachfremder Urteilskategorien ausdrückt, die Ergebnisse der Erhebungen in ihrer Validität beeinträchtigt.<sup>3</sup> Diese Gefahr ist (u.a. auch aus Gerechtigkeitserwägungen) insbesondere dann zu beachten, wenn die Evaluationsergebnisse für Zwecke benutzt werden sollen, die eine Steuerung der Mittelallokation zum Ziel haben [z.B. ENGEL U. KREKELER 2001]. Aber auch in einem allgemeinen Sinne und unabhängig von der Steuerungswirksamkeit von Evaluationen erscheint die Beantwortung der Frage nach dem Vorliegen bzw. der Verletzung von

---

<sup>2</sup> Für die Evaluationsforschung allgemein sind vom „Joint Committee“ Standards für Programmevaluationen entwickelt worden, die Fragen der Testgütekriterien von Messungen und der daran anschließenden Informationsnutzung explizit berühren [SANDERS 1999]. Für den Bereich der Evaluation von Hochschulleistungen vgl. z.B. LEVACIC U. GLATTER [2001].

<sup>3</sup> Zur Klärung der Frage nach der Validität von studentischen Qualitätsurteilen, die sich auf konkrete Lehrpräsentationen in Lehrveranstaltungen beziehen liegt bereits ein umfänglicher Literaturkörper vor. Dabei werden durchaus unterschiedliche Ergebnisse berichtet, die einerseits das Vorliegen von Validität studentischer Urteile bestätigen bzw. diese im Gegenteil bedroht sehen [vgl. z.B. RINDERMANN 2001, SPIEL 2001; KROMREY 2001, 2001a; ESSER 1997]. Bezogen auf die Ebene von ganzen Studienprogrammen, die verstärkt in den Fokus der Evaluation von Lehre und Studium sowie der Akkreditierung von Studiengängen rückt und die sich dabei ebenfalls auf studentische (Global-) Urteile stützt, ist diese Diskussion noch nicht in gleichem Maße geführt worden.

relevanten Validitätsaspekten zentral zu sein. So ist in der Verfügbarkeit einer validen Datenbasis über die Qualität von Lehre und Studium eine unabdingbare Voraussetzung für eine evidenzbasierte Beurteilung von Studienprogrammen sowie für die Entwicklung etwaiger Verbesserungsstrategien zu sehen [GINSBURG & RHETT 2003]. Evidenzbasierte Steuerung von (öffentlichen) Einrichtungen wird in einer zunehmenden Zahl von Kontexten und Politikfeldern und so auch im Bereich der Hochschulleistungen zu einem paradigmatischen Handlungsrahmen [MÜLLER-BÖLING 2000; NUTLEY ET AL. 2003]. Die Untermauerung gesellschaftlicher und politischer Entscheidungsprozesse mit validen Daten und Forschungsergebnissen wird daher immer dringlichere Forderung. Diese bereit zu stellen ist zentrale Aufgabe der Evaluationsforschung.

Der vorliegende Beitrag beschäftigt sich in diesem Zusammenhang mit zwei zentralen Validitätsaspekten. Zum einen wird der genannte Aspekt des Einflusses sachfremder Beurteilungskategorien und –komponenten auf das Zustandekommen des Urteils der Lehramtskandidatinnen untersucht. Art und Ausmaß des Einflusses potenziell verzerrender Wirkungsgrößen sollen identifiziert und in ihrem jeweiligen Einfluss beschrieben werden. Auf diese Weise ist eine Standortbestimmung der Evaluationsergebnisse insgesamt möglich. Zum anderen werden die der „Potsdamer Lehramtskandidaten-Studie 2004/05“ [vgl. SCHUBARTH ET AL. in diesem Band] zu Grunde gelegten Qualitätsdimensionen der zweiten Phase der Lehrerbildung dahingehend untersucht, ob sie durch die beurteilenden Personen in der erwarteten Weise aufgefasst und zur Grundlage der abgegebenen Qualitätseinschätzungen gemacht wurden. Es stellt sich die Frage, ob die Urteile der Lehramtskandidatinnen<sup>4</sup> die theoretisch abgeleiteten Urteilskomponenten widerspiegeln. Entsprechende Befunde ermöglichen zu analysieren, inwieweit die erbetenen Einschätzungen zu verschiedenen Qualitätsaspekten mit den jeweiligen Zieldimensionen, die ihnen theoretisch zugeordnet wurden, in Verbindung gebracht werden können. Der zuletzt genannte Aspekt berührt insbesondere die Frage nach der Konstruktvalidität [FALTER 1977], während der erste sich dem (verzerrenden) Einfluss von Bias-Variablen [SPIEL 2001; ENGEL U. POHLENZ 2001] widmet. Neben diesen beiden relevanten Validitätsaspekten ließen sich noch vielerlei weitere nennen, die im Rahmen des vorliegenden Beitrags jedoch nicht behandelt werden können. Dieser hat vielmehr das Ziel, auf die Bedeutung von Testgütekriterien von Messungen bei der Evaluation von Lehre und Studium insgesamt hinzuweisen, Methoden zu

---

<sup>4</sup> Aus Gründen der sprachlichen Vereinfachung werden Gruppenbezeichnungen nur entweder in der männlichen oder der weiblichen Form genannt. Grundsätzlich sind Personen beiderlei Geschlechts gemeint. Im Falle der befragten brandenburgischen Lehramtskandidaten und –kandidatinnen wird hauptsächlich die weibliche Form benutzt, weil Frauen in dieser Gruppe eindeutig die Mehrheit stellen.

ihrer Überprüfung vorzustellen und die Evaluationsergebnisse der Studie zur Einschätzung der Ausbildungsqualität durch brandenburgische Lehramtskandidatinnen exemplarisch vor dem Hintergrund der ausgewählten Validitätsaspekte zu diskutieren.

## **2. Validitätsbedrohung durch subjektiv verzerrte Urteile?**

Studentische Befragungsdaten können relevante Einsichten in die Studienrealität liefern und Verbesserungsstrategien anstoßen, sofern sie „rational in identifizierbaren Eigenschaften des zu beurteilenden Gegenstandes (...) begründet sind“ [ENGEL U. POHLENZ 2001]. Problematisch sind die Urteile allerdings dann, wenn sich in ihnen nicht nur die Qualität bspw. des Ausbildungsprogramms widerspiegelt, sondern auch sachfremde Urteilkategorien, wie persönliche Interessen und Erwartungen, der elterliche Bildungshintergrund, o.ä. in das Urteil eingehen. Derartige Einflussgrößen („Bias-Variablen“) können Qualitätseinschätzungen urteilender Personen beeinflussen, lassen sich aber nicht aus den Merkmalen des beurteilten Gegenstandes ableiten. Sie sind entsprechend als (potenzielle) Fehlerquellen bei der entsprechenden Messung anzusehen [a.a.O.].

### **2.1 Erklärung des Urteils zur Ausbildungsqualität**

Zur Identifikation der verschiedenen Einflussgrößen des Urteils der Lehramtskandidatinnen über die Ausbildungsqualität wurden multiple lineare Regressionen gerechnet. Im Folgenden werden exemplarisch einige der erklärungskräftigen Prädiktoren der studentischen Globalurteile zur Ausbildungsqualität an der *Ausbildungsschule* diskutiert. Dabei ist die Frage zu klären, inwieweit sich diese Globalurteile auf sachbezogene Urteilkategorien zurückführen lassen bzw. inwieweit sie durch sachfremde Determinanten verzerrt sind und entsprechend von Validitätsbedrohungen der Daten auszugehen ist. Insbesondere werden dabei vier theoretisch fundierte Qualitätsdimensionen berücksichtigt, die die Basis für die „Potsdamer Lehramtskandidatinnen-Studie“ (vgl. SCHUBARTH ET AL. in diesem Band) bildeten. Bei diesen handelt es sich um die *Inputqualität*, die *Ergebnisqualität*, die *Prozessqualität* und die *Kontextqualität*. Input- und Kontextqualität beschreiben Eingangsvoraussetzungen der Lehramtskandidatinnen und Rahmenbedingungen der zweiten Phase der Lehrerbildung, die Prozessqualität bezieht sich auf Parameter des Ausbildungsprozesses und die diesbezüglichen Qualitätseinschätzungen, während die Ergebnisqualität die Frage der Urteile zu den Ergebnissen der Lernprozesse beleuchtet [vgl. a.a.O.].

### 2.1.1 Ausbildungsqualität an der Ausbildungsschule

Das Urteil der angehenden Lehrerinnen zur Ausbildungsqualität an ihrer jeweiligen Ausbildungsschule wird hinsichtlich der Eingangsvoraussetzungen und der Rahmenbedingungen (Input- und Kontextqualität) offensichtlich nur in geringem Maße durch sachfremde Beurteilungskategorien beeinflusst. So wirken bspw. persönliche Hintergrundmerkmale wie das Alter, der elterliche Bildungshintergrund oder das Geschlecht nicht als erklärungskräftige Prädiktoren des Qualitätsurteils. Allerdings spielen in diesem Zusammenhang Fragen nach der Motivation der Berufswahl eine entsprechende Rolle: Die Ausbildungsschule wird besser bewertet, wenn die Motivation, den Lehrerberuf zu ergreifen, darin begründet war, dass ein entsprechender Berufswunsch seit langem gehegt wurde. Sie wird kritischer bewertet, wenn mit der Aufnahme des Lehrerstudiums die Erwartung verbunden war, „Schule verbessern“ zu können (vgl. Tabelle 1). Offensichtlich wird diese Erwartung durch die gemachten Erfahrungen nicht bestätigt.

Erwartungen, die sich an ein Studium oder eine Ausbildung richten, sind durchaus als subjektive Urteilkategorie anzusehen. Dass Dinge, die die Erwartungen der diese Dinge beurteilenden Person erfüllen, in der Beurteilung besser wegkommen als solche, für die dieses nicht zutrifft, ist einleuchtend. Problematisch ist dabei, dass in dem Maß, in dem eine Sache Erwartungen erfüllt, sich auch Präferenzen und Eigenschaften des Betrachters widerspiegeln und nicht ausschließlich diejenigen der beurteilten Sache [ENGEL U. KREKELER 2001]. Bezogen auf den vorliegenden Fall würde dies bedeuten, dass die Qualitätseinschätzung der Ausbildung von dem Maß abhängt, zu dem die Kandidaten sich durch diese Ausbildung befähigt fühlen, „Schule zu verbessern“. Die Auffassungen darüber, was zu einer entsprechenden Befähigung nötig ist, sind jedoch ebenso individuell wie möglicherweise widersprüchlich. Für die Ausbildungsinstitutionen ist es in diesem Fall also nahezu unmöglich, auf die individuellen, subjektiv gefärbten Qualitätsanforderungen zu reagieren. Gleichzeitig muss betont werden, dass die genannten Prädiktoren<sup>5</sup> nur einen geringen Beitrag von rund 8 Prozent zur Varianzaufklärung leisten ( $R^2 = .08$ ).

Die folgenden Tabellen zu den multiplen linearen Regressionsanalysen informieren über die standardisierten Regressionskoeffizienten ( $\beta$ ) und die Prüfgröße  $t$ . Die Effekte der in Tabelle 1 gezeigten Prädiktoren gehen

---

<sup>5</sup> Ein weiterer erklärungskräftiger Prädiktor in diesem Modell war die Frage nach der Zugehörigkeit zu Gruppen verschiedener Berufsorientierungen: Die Ausbildungsschulen werden von Lehramtskandidatinnen die der Gruppe „Burn-Out-Typ“ zuzurechnen sind, kritischer bewertet als von den Referenzgruppen (vgl. Tabelle 1).

Variablentext	$\beta$	t
<i>Wie wichtig waren die folgenden Gründe bei der Entscheidung Lehrer zu werden?</i>		
Seit langem fester Berufswunsch	-.181**	-2,739
Wunsch, Schule zu verbessern	.160*	2,413
<i>Typ der beruflichen Orientierung</i>		
Burn-Out-Typ	.166*	2,519

\*  $p \leq .05$ ; \*\*  $p \leq .01$

*Tabelle 1: Prädiktoren des Globalurteils: Inputqualität*

in die erwartete Richtung.<sup>6</sup> Prädiktoren des Urteils zur Qualität der Ausbildungsschulen, die sich auf die Rahmenbedingungen der Ausbildung beziehen (Kontextqualität) reflektieren dagegen durchgängig qualitätsrelevante Merkmale der Ausbildung. Sowohl die Einschätzung der Betreuung der Lehramtskandidatinnen beim Einstieg in den Alltag der Ausbildungsschule, als auch der Transparenz von Leistungsanforderungen stellen sachbezogene Urteilkategorien dar, deren Steuerung durchaus im Handlungsspielraum der verantwortlichen Institutionen liegt. Gleiches gilt für die Möglichkeit der Lehramtskandidatinnen, im Unterricht von Kollegen zu hospitieren und für ihre Mitbestimmung. Die Zufriedenheit der beurteilenden Personen mit den zur Verfügung stehenden Materialien hingegen ist nur bedingt durch die beurteilten Ausbildungsschulen beeinflussbar, sondern hängt von der allgemeinen Infrastrukturausstattung ab. Hier liegt zwar offensichtlich eine sachbezogene Urteilkategorie vor, allerdings wird sie vor dem Hintergrund der Frage nach verzerrten Qualitätsurteilen über die Ausbildung als Fehlerquelle zu werten sein: Das Qualitätsurteil wird zwar durch sie beeinflusst, eine Möglichkeit des steuernden Eingriffs durch die beurteilte Institution im Sinne einer Qualitätssteigerung liegt jedoch nur sehr bedingt vor [ENGEL U. KREKELER 2001]. Tabelle 2 zeigt die Ergebnisse der Regression. Die Effekte gehen in die erwartete Richtung. Die berichteten Prädiktoren leisten mit insgesamt rund 58 Prozent aufgeklärter Varianz ( $R^2 = .58$ ) einen vergleichsweise starken Beitrag zur Erklärung des Globalurteils über die Ausbildungsqualität an den Ausbildungsschulen.

---

<sup>6</sup> Mit einer kritischen Bewertung der Ausbildungsschule ging ein hoher Skalenwert in der entsprechenden (abhängigen) Variablen einher, während die (unabhängigen) Variablen zur Beurteilung der Wichtigkeit von Studienmotivationen umgekehrt codiert waren.

Variablentext	$\beta$	t
<i>Wie zufrieden sind Sie mit den allgemeinen Rahmenbedingungen?</i>		
Betreuung beim Einstieg	-.338**	-5,158
Transparenz der Leistungsanforderungen	-.186**	-3,113
Materialien zur Unterrichtsvorbereitung	-.152**	-2,713
Möglichkeiten zu Hospitationen	-.148**	-2,691
Möglichkeiten der Mitbestimmung	-.137*	-2,252

\*  $p \leq .05$ ; \*\*  $p \leq .01$

*Tabelle 2: Prädiktoren des Globalurteils: Kontextqualität*

Gleiches gilt für die Gruppe von Prädiktoren, die den Ausbildungsprozess in den Ausbildungsschulen beschreiben (Prozessqualität). Diese erklären gemeinsam 76 Prozent der Varianz im Globalurteil über die Qualität der Ausbildungsschulen ( $R^2 = .76$ ), die entsprechenden Effekte gehen in die erwartete Richtung (vgl. Tabelle 3 und Fußnote 6). Erklärungskräftige Prädiktoren beziehen sich hier insbesondere auf die Beziehungsqualität zwischen Lehramtskandidatinnen und deren Ausbilder. Erwartungsgemäß wird die Ausbildungsschule immer dann besser bewertet, wenn dieses Verhältnis als positiv beschrieben wird. Die entsprechenden erklärungsstarken Kriterien sind hier die Offenheit der Schule und der Ausbilder für Kritik und neue Ideen, Interesse und Informiertheit der Ausbilder sowie eine faire und transparente Leistungsbeurteilung.

Variablentext	$\beta$	t
Persönliche Belastung durch den Anpassungsdruck in der Ausbildungsschule.	.270**	5,041
In meiner Ausbildungsschule besteht Offenheit gegenüber Neuen.	-.380**	-7,268
Der Ausbildungslehrer ist offen für Kritik.	-.162**	-3,380
Die Schulleitung steht meiner Ausbildung desinteressiert gegenüber.	.145**	2,924
Der Ausbildungsleiter kontrolliert mich mehr als er mich unterstützt.	.119*	2,529
Der Hauptseminarleiter beurteilt meine Leistungen willkürlich.	-.118**	-2,876
Das Fachseminar hat eine klar erkennbare Struktur.	-.101*	-2,431
Der Ausbildungslehrer ist über die Ausbildung im Studienseminar informiert.	-.088*	-2,067

\*  $p \leq .05$ ; \*\*  $p \leq .01$

*Tabelle 3: Prädiktoren des Globalurteils: Prozessqualität*

Darüber hinaus wirkt sich positiv auf das Globalurteil aus, wenn keine Belastungen durch schulischen Anpassungsdruck empfunden wurden. Die genannten Beurteilungskriterien sind zwar durchweg offensichtlich subjektive Urteilkategorien. Gleichwohl reflektieren sie rational ableitbare Eigenschaften bzw. Qualitätskriterien der Ausbildung. Die genannten Kriterien sind zudem durchaus im Einflussspielraum der beurteilten Institutionen angesiedelt. Insofern können studentische Angaben zu ihren Qualitätseinschätzungen in diesem Zusammenhang einerseits als Ausgangspunkt für Verbesserungsstrategien dienen und andererseits weitere Analysen zur Frage der Determinanten des studentischen Qualitätsurteils inspirieren. Diese könnten darin bestehen zu untersuchen, von welchen Faktoren die Einschätzung des Verhältnisses zu den Ausbildern abhängt (bspw. von soziodemographischen Variablen wie dem Alter oder dem Geschlecht).

Zur Bewertung der Ergebnisse des Ausbildungsprozesses wurden den Lehramtskandidatinnen Fragen zu ihrer Einschätzung der durch die Ausbildung (in erster und zweiter Phase) erreichten beruflichen Qualifikation vorgelegt. Tabelle 4 zeigt Prädiktoren des Globalurteils über die Qualität der Ausbildungsschule, die sich auf die von den Befragten wahrgenommene Wichtigkeit von verschiedenen Ausbildungsinhalten und die Ergebnisse der Ausbildung im Bereich der Lehrerqualifikationen „Unterrichten“ und „Innovieren“ beziehen (Ergebnisqualität). Demnach wird die Ausbildungsschule besser beurteilt, wenn die Ausbildungsinhalte „Klassenlehrertätigkeit“ sowie „Lehr- und Lernmethoden“ als wichtig erachtet wurden und wenn die Lehramtskandidatinnen die Ausbildung hinsichtlich der mit den genannten Lehrerqualifikationen verbundenen Kompetenzen als erfolgreich einschätzten. Die in Tabelle 4 aufgeführten Prädiktoren erklären gemeinsam rund ein Fünftel der Varianz im Globalurteil über die Qualität der Ausbildungsschulen ( $R^2 = .21$ ), die Effekte gehen in die erwartete Richtung.

Die genannten Prädiktoren aus dem Bereich der Ergebnisqualität spielen im Kontext der hier berichteten Analysen insofern eine Sonderrolle, als sie sich auf eine abschließende Beurteilung des durch die Ausbildung erreichten Kompetenzerwerbs beziehen. Da der Ausbildungsprozess zum Zeitpunkt der Befragung jedoch noch nicht abgeschlossen war, ist die Frage zu stellen, inwieweit entsprechende Urteile rational sein können. So leisten sie auch nur einen vergleichsweise geringen Beitrag zur Varianzaufklärung (s.o.).



Variablentext	$\beta$	t
„Wie wichtig sind Ihnen die folgenden Ausbildungsinhalte?“		
Klassenlehrertätigkeit	-.170**	-2,760
Lehr- und Lernmethoden	-.157*	-2,556
„Inwiefern wurden die folgenden Handlungskompetenzen im Hinblick auf die Lehrerqualifikation ‚Unterrichten‘ vermittelt?“		
... spontan entstehende Diskussionen von Lernenden fruchtbar zu machen.	-.227**	-3,422
... Lernerfolgskontrollen regelmäßig und methodisch vielfältig vorzubereiten, durchzuführen und auszuwerten.	.213**	3,210
„Inwiefern wurden die folgenden Handlungskompetenzen im Hinblick auf die Lehrerqualifikation ‚Innovieren‘ vermittelt?“		
... meine persönliche Lehrerrolle zu entwickeln und zu verwirklichen.	-.165**	-2,640
... durch Engagement im Schulleben zur Profilierung der Schule beizutragen.	-.170*	-2,556
... ein persönliches Fortbildungs- und Zusatzausbildungsprogramm zusammenzustellen.	.195**	3,015

\*  $p \leq .05$ ; \*\*  $p \leq .01$

*Tabelle 4:* Prädiktoren des Globalurteils: Ergebnisqualität

## 2.2 Zusammenfassung der Ergebnisse

Abschließend ist zu bemerken, dass die Qualitätsurteile der Lehramtskandidatinnen in nur geringem Maß durch Variablen beeinflusst werden, denen attestiert werden könnte, dass sie einen verzerrenden Einfluss auf die Einschätzungen ausüben. Lediglich Variablen zu individuellen Studien- bzw. Ausbildungserwartungen zeitigen einen entsprechenden Effekt. Zudem sind diese Variablen aus der Gruppe der Faktoren, die die Eingangsvoraussetzungen der Lehramtskandidatinnen beschreiben (Inputqualität) von nur sehr geringer Erklärungskraft für das Globalurteil zur Qualität der Ausbildungsschulen. Verzerrende Einflüsse, die von den persönlichen Dispositionen der befragten Personen ausgehen, konnten dagegen nicht als prädiktiv für das zu erklärende Qualitätsurteil identifiziert werden. Solche Variablen stehen typischerweise in dem Verdacht, einen subjektiv verzerrenden Einfluss auf Qualitätsurteile von Personen, gemäß ihrer individuellen Dispositionen, die sich bspw. aus ihrem Geschlecht, ihrem Alter, dem Einkommen oder dem elterlichen Bildungshintergrund ableiten, auszuüben. Offensichtlich lässt sich das Urteil der Lehramtskandidatinnen weitgehend aus Bewertungskategorien ableiten, die rational in den Eigenschaften des zu beurteilenden Gegenstandes be-

gründet sind. Variablen zur Beschreibung der Rahmenbedingungen der Ausbildung sowie des Ausbildungsprozesses selber liefern die erklärendenkräftigsten Prädiktoren. Die Eingangsvoraussetzungen und die Ergebnisse des Ausbildungsprozesses sind dagegen offensichtlich von geringerer Bedeutung für die globale Beurteilung der Ausbildungsqualität an den Ausbildungsschulen.

Die vier genannten, durch den Untersuchungsansatz der hier beschriebenen Studie theoretisch hergeleiteten Urteilsdimensionen werden im Folgenden daraufhin analysiert, ob sie durch die Urteile der befragten Personen empirisch widerspiegelt werden. In dieser Frage ist ein weiterer zentraler Validitätsaspekt zu sehen, der sich auf die Konstruktvalidität der Daten bezieht.

### **3. Übereinstimmung der Konstrukte mit den Qualitätsurteilen?**

Zur Untersuchung der Konstruktvalidität der erhobenen Daten hinsichtlich der vier angestrebten Zieldimensionen wurde eine konfirmatorische Faktorenanalyse (KFA) gerechnet [HOPKO 2003; PROOJEN & KLOOT 2001]. KFA ermöglichen es, den Zusammenhang von beobachtbaren Indikatorvariablen und den ihnen zu Grunde liegenden latenten, also nicht direkt beobachtbaren Faktoren zu analysieren [z.B. BYRNE 2001; KLINE 1998]. Bezogen auf den vorliegenden Fall lässt sich mittels KFA ermitteln, ob und inwieweit die theoretisch formulierten Qualitätsdimensionen der zweiten Phase der Lehrerbildung durch die sie repräsentierenden Variablen beschrieben werden. Eine entsprechende Übereinstimmung würde signalisieren, dass die theoretischen Dimensionen durch die empirisch erhobenen Qualitätsurteile der befragten Personen bestätigt werden, mithin Konstruktvalidität dieser Dimensionen vorliegt. Umgekehrt könnte ein fehlender Zusammenhang der Dimensionen mit den ihnen jeweils zugeordneten Indikatorvariablen andeuten, dass die Interpretation der erhobenen Daten, die vier dem Urteil der Lehramtskandidatinnen zu Grunde liegende Qualitätsdimensionen unterstellt, überdacht werden muss. Diesbezügliche Kenntnisse sind deshalb relevant, weil sich eine etwaige Formulierung von Entwicklungsstrategien des beurteilten Gegenstandes, vermutlich in erster Linie an die als relevant unterstellten Qualitätskonzepte anlehnen würde (z.B. „Verbesserung der Kontextqualität durch eine Veränderung/Verbesserung der Organisationsleistungen“). Um solche Aktivitäten rational bzw. auf empirischer Evidenz zu basieren, bedarf es jedoch einer empirischen Validierung der Qualitätskonzeptionen.

### 3.1 Zieldimensionen der Qualitätskonzeption

Für die konfirmatorische Faktorenanalyse wurden aus dem umfangreichen Variablen­set des eingesetzten Erhebungsinstruments diejenigen als Indikatorvariablen ausgewählt, die die Inhalte der zu überprüfenden Zieldimensionen aus vorliegender Sicht am deutlichsten repräsentieren (vgl. Tabelle 1).<sup>7</sup> Dabei wurden insbesondere die Ergebnisse der bereits an anderer Stelle vorgelegten deskriptiv-statistischen Auswertungen der Daten berücksichtigt [vgl. SCHUBARTH ET AL. 2005]. So wurden im Fall der Indikatorvariablen, die die *Inputqualität* repräsentieren solche ausgewählt, die insbesondere die intrinsische Motivation der Kandidatinnen bei der Wahl ihres beruflichen Lebensweges zum Ausdruck bringen. Diese wurden als besonders charakteristisch für die Eingangsbedingungen der angehenden Lehrerinnen identifiziert (a.a.O). Zur Abbildung der *Kontextqualität* wurden Variablen ausgewählt, die insbesondere Organisations- und Betreuungsaspekte der Ausbildung reflektieren. Die (studentische) Einschätzung der Organisationsqualität und der Organisationsleistung verantwortlicher Institutionen und Personen konnte an anderer Stelle bereits als relevant für die Beurteilung von Studien- und Rahmenbedingungen identifiziert werden [ENGEL U. POHLENZ 2001; POHLENZ U. TINSNER 2004]. Die *Prozessqualität* wird unterteilt in den Aspekt der Beziehungsqualität (zwischen Lehramtskandidatinnen und Ausbildern) und den der fachlich-methodischen Qualität. Die Beziehungsqualität wird im vorliegenden KFA-Modell durch eine Frage zur hierarchischen Organisation („Unterordnung“; vgl. Tab. 1) und die fachlich-methodische Qualität durch die diesbezügliche Qualität von Prozessen und verantwortlichen Personen repräsentiert. Die *Ergebnisqualität* schließlich wird durch relevante Outcome-Aspekte der Ausbildung, die sich auf die Selbsteinschätzung der Lehramtskandidatinnen hinsichtlich der erreichten Kompetenzen in den zentralen Feldern der Lehrerqualifikationen beziehen, widerspiegelt.

---

<sup>7</sup> Bei den ausgewählten Variablen handelt es sich um fünfstufige Ratingskalen (z.B. „völlig unwichtig“ (1) → „sehr wichtig“ (5); „sehr unzufrieden“ (1) → „sehr zufrieden“ (5), etc.).

Dimension	Var.-Name <sup>8</sup>	Variablentext
Inputqualität	<i>Wie wichtig waren die folgenden Gründe bei der Entscheidung Lehrer zu werden?</i>	
	1d	Seit langem fester Berufswunsch.
	1g	Freude am Umgang mit Kindern und Jugendlichen.
	1h	Anderen helfen.
	1j	Freude an Wissensvermittlung.
Ergebnisqualität	<i>Einschätzung der eigenen Kompetenzen hinsichtlich der Lehrerqualifikationen Unterrichten, Erziehen, Beraten, Beurteilen, Innovieren und Organisieren</i>	
	9_3i	<i>Ich fühle mich kompetent, ...handlungsorientierte Lernformen anzuwenden.</i>
	10_3	...den Unterricht so zu organisieren, dass die Lernenden soziale Kompetenzen erlernen können.
	11_3g	...Techniken der Gesprächsführung anzuwenden, um einfühlsam und sachgerecht zu beraten.
	12_3b	...den Lernenden ein motivierendes Feedback über ihre Lernfortschritte zu geben.
	13_b	...den eigenen und fremden Unterricht kritisch zu analysieren und zu verbessern.
	14_i	...die allgemein- und schulrechtlichen Bestimmungen und Normen situationsgerecht anzuwenden.
Prozessqualität	<i>Inwiefern treffen die folgenden Aussagen auf die Ausbildung im Hauptseminar zu?</i>	
	3b	Das Hauptseminar ist methodisch vielfältig gestaltet.
	3d	Der Hauptseminarleiter zeichnet sich durch ein hohes fachliches Niveau aus.
	3i	Der Hauptseminarleiter will nur, dass ich mich ihm unterordne.
Kontextqualität	<i>Wie zufrieden sind Sie mit den allgemeinen Rahmenbedingungen (in der Ausbildungsschule)</i>	
	2_2o	Betreuung beim Einstieg (in die Ausbildungsschule).
	2_2p	Möglichkeiten zu Hospitationen.
	2_1l	Zeitliche Verfügbarkeit der Ausbildungslehrer.

*Tabelle 5: In die KFA einbezogene Variablen*

### *3.1.1 Hypothesen*

Im Sinne der Überprüfung, ob für die vier beschriebenen Dimensionen Konstruktvalidität vorliegt, wurden mit dem Messmodell, das der nachfolgend berichteten KFA zugrunde liegt, drei zu überprüfende Annahmen verknüpft:

<sup>8</sup> Vgl. Pfadmodell der KFA in Abbildung 1.

- i) Die für die Faktoren (Dimensionen) spezifizierten Indikatorvariablen korrelieren hoch mit ihrem und nur mit ihrem jeweiligen Faktor (Diskriminanz).
- ii) Die Indikatorvariablen eines Faktors korrelieren untereinander hoch (Konvergenz)
- iii) Im Sinne einer Messung distinkter Konstrukte durch die vier Faktoren, weisen diese untereinander bestenfalls geringe Korrelationen auf.

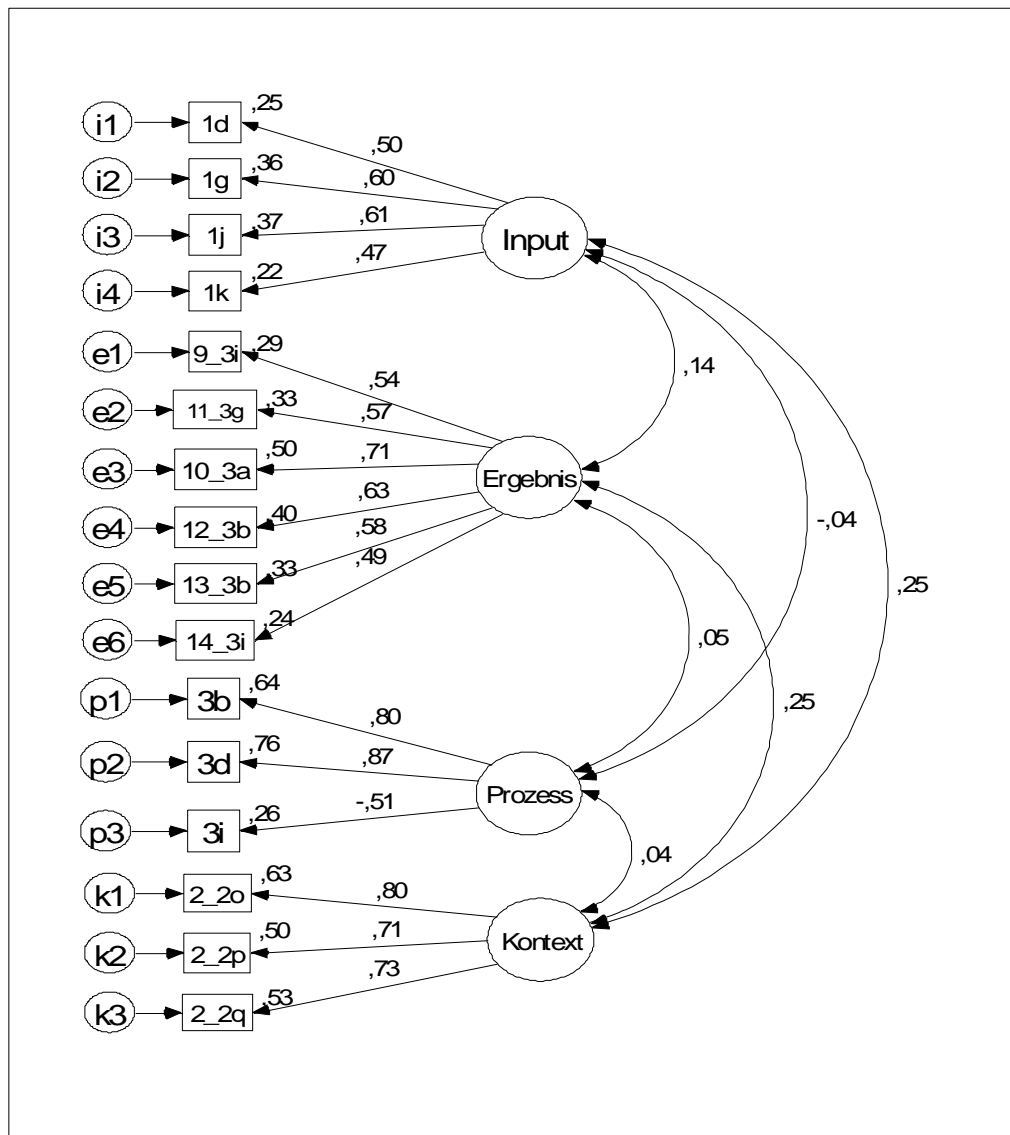
In den drei formulierten Annahmen spiegeln sich einschlägige Kriterien zur Beurteilung der Konstruktvalidität wider [z.B. SCHNELL ET AL. 1999: 155 ff]. Diskriminanz oder „diskriminante Validität“ liegt vor, wenn sich für ein Konstrukt nachweisen lässt, dass es andere Sachverhalte misst, als andere Konstrukte, wenn also durch das Konstrukt spezifische Sachverhalte erfasst werden, die nicht gleichzeitig auch durch andere Konstrukte erfasst werden könnten. Konvergenz oder „konvergente Validität“ zeigt dagegen an, ob verschiedene Operationalisierungen eines Konstrukts miteinander in Zusammenhang stehen [a.a.O]. Durch die dritte der formulierten Annahmen wird darüber hinaus das Verhältnis der latenten Faktoren untereinander beschrieben, wobei davon ausgegangen wird, dass diese im Sinne voneinander unabhängiger Dimensionen nur in geringem Maße in einem korrelativen Zusammenhang stehen.

### 3.2 Ergebnisse

Das berichtete KFA-Modell wurde nach der Maximum Likelihood Methode geschätzt. Diese ist vergleichsweise unempfindlich gegen die Verletzung von Modellannahmen, die sich auf die multivariate Normalverteilung der einbezogenen Variablen beziehen [Bühner 2004: 201].<sup>9</sup> Abbildung 1 zeigt das Pfadmodell für die ermittelten Faktorstrukturen und Faktorladungen sowie die Maßzahlen für den Modell-Fit an.

---

<sup>9</sup> Die Verteilungen der in das berechnete Faktorenmodell einbezogenen Variablen weisen zwar nicht durchgängig Normalverteilung auf, so dass die zentrale Normalverteilungsannahme verletzt ist. Schiefe und „Exzess“ (Kurtosis) der Verteilungen der einzelnen Variablen bleiben aber innerhalb empfohlener Grenzwerte (West, Curran u. Finch 1995; zit. n. Bühner 2004: 232).



$$\chi^2 = 102,67, df = 98, p = .354; CFI = .995, NFI = .904;$$

$$RMSEA = .013 (LO 90/HI 90 = .000/.034)^{10}$$

Abbildung 1: Pfadmodell der KFA (standardisierte Schätzer)

Der exakte Modell-Fit bestätigt mit einem nichtsignifikanten  $\chi^2$ -Wert ( $p = .354$ ) von 102,67 bei 98 Freiheitsgraden die Güte des Modells. Dieser Befund wird zusätzlich durch die hohen (CFI, NFI) bzw. niedrigen

<sup>10</sup> Die Berechnungen wurden mit dem AMOS-Programmpaket durchgeführt. Dieses bietet zur Behandlung fehlender Werte eine Rechenprozedur an, die ohne den fall- bzw. listenweisen Ausschluss fehlender Werte auskommt (die Anteile fehlender Werte in den in die Analyse einbezogenen Variablen betragen zwischen 0,3 % und 4,3 % mit Ausnahme, dass jeweils ca. 10 % der Befragten die Fragen nach der Betreuung beim Einstieg, nach den Möglichkeiten zu Hospitationen und nach der zeitlichen Verfügbarkeit der Lehrer nicht beantworteten). Allerdings werden in diesem Verfahren einige der für die Beurteilung von linearen Strukturgleichungsmodellen üblichen Fit-Indizes (z.B. GFI, SRMR) nicht ausgegeben [z.B. BYRNE 2001].

(RMSEA<sup>11</sup>) Werte der Indizes für den approximativen Modell-Fit bestätigt. Die in das Modell einbezogenen Variablen legen damit ein tatsächliches Vorliegen der unterstellten 4-Faktorenstruktur bzw. der vier beschriebenen, den Bewertungen zugrunde liegenden Dimensionen nahe.<sup>12</sup>

Die Variablen laden durchweg hoch auf den jeweiligen Faktoren, die standardisierten partiellen Regressionsgewichte betragen im Fall des Faktors Inputqualität zwischen .47 und .61. Besonders deutlich fallen die Zusammenhänge der Indikatorvariablen mit den Konstrukten Kontextqualität (.71 bis .80) und Prozessqualität aus. Hier betragen die Gewichte zwischen |.51| und .87. Die Indikatorvariablen zur Beschreibung der Ergebnisqualität laden ebenfalls mit .49 bis .71 hoch. Die erste Annahme, die eine Korrelation der Indikatorvariablen mit ihrem jeweiligen Faktor und nur mit diesem Faktor unterstellt wird durch das Modell bestätigt. Somit ist auch eines der Kriterien für das Vorliegen von Konstruktvalidität, namentlich die Diskriminanz erfüllt. Für die quadrierten multiplen Korrelationen der jeweiligen Indikatorvariablen eines Faktors untereinander werden ebenfalls teils sehr hohe Werte (bis zu .76), aber auch eine Reihe von vergleichsweise kleinen bis mittleren Werten (.22 bis .37) ermittelt. Die quadrierten multiplen Korrelationen geben die gemeinsame Varianz der Indikatorvariablen, die sich auf den Zusammenhang mit ihrem jeweiligen Faktor zurückführen lässt, an. Sie beschreiben mithin das Maß, zu dem die Indikatorvariablen eines Faktors auf die jeweils anderen zurückgeführt werden können. Die mehrheitlich hohen Werte für die quadrierte multiple Korrelation zeigen an, dass die vier Faktoren durchaus in konsistenter Weise durch die ihnen zugeordneten Indikatorvariablen erklärt werden und dementsprechend reliable Messungen der Konstrukte durch diese Variablen vorliegen. Die durchweg hohen Werte für die Zusammenhangsmaße der Indikatorvariablen zeigen ein zufrieden stellendes Maß an Konvergenz an. Auch die zweite der oben beschriebenen Hypothesen wird damit durch das Modell bestätigt.<sup>13</sup>

---

<sup>11</sup> Das 90-prozentige Konfidenzintervall für den RMSEA (LO 90/ HI 90) schließt den Wert .000 ein und bleibt unter dem von HU U. BENTLER [1999; zit. n. BÜHNER 2004: 243] empfohlenen Wert von .060, was auf einen zufrieden stellenden Modell-Fit hinweist.

<sup>12</sup> Gleichzeitig darf aus der Tatsache, dass das Modell „auf die erhobenen Daten passt“, nicht geschlossen werden, dass es das einzig gültige Modell zur Beurteilung der Beziehungsstrukturen zwischen den erhobenen Daten ist. Die einbezogenen Variablen wurden zwar aufgrund a priori getroffener Annahmen (s. 2.3) ausgewählt. Sie entstammen aber einer Fülle möglicher, die Zielkonstrukte u. U. ebenfalls abbildenden Kombinationen von Indikatorvariablen.

<sup>13</sup> Zu den nur in geringem Maße zur Varianzerklärung beitragenden Items (Werte um .20) könnten Modellmodifikationen vorgenommen werden, auf die jedoch angesichts des insgesamt zufrieden stellenden Modells im vorliegenden Beitrag verzichtet wird.

Hinsichtlich der Zusammenhänge zwischen den vier (latenten) Zieldimensionen kann eine schwache Korrelation zwischen den Faktoren *Inputqualität* und *Ergebnisqualität* (.14) sowie die moderaten Korrelationen der Faktoren *Kontextqualität* und *Inputqualität* (.25) sowie *Kontextqualität* und *Ergebnisqualität* (.25) ermittelt werden. Der Befund (weitgehend) unabhängiger latenter Variablen spricht generell dafür, dass durch die vier latenten Variablen tatsächlich vier inhaltlich distinkte Dimensionen gemessen werden. Durch diesen Befund wird auch die letzte der drei formulierten Annahmen gestützt, die als Zeichen der Unabhängigkeit der vier Konstrukte deren bestenfalls geringe Kovariation unterstellte.

Auffällig ist dabei aber auch, dass der Faktor *Prozessqualität* im Zusammenspiel der vier Dimensionen keinerlei Zusammenhang mit den anderen aufweist. Die Dimension Prozessqualität wird offensichtlich als eigenständiges Konstrukt beurteilt.

#### **4. Diskussion**

Die berichteten Regressionsmodelle legen den Schluss nahe, dass die erhobenen Daten weitgehend verzerrungsfrei sachbezogene Qualitätsurteile der angehenden Lehrerinnen abbilden. Durch die konfirmatorische Faktorenanalyse konnte darüber hinaus bestätigt werden, dass die dem Evaluationsmodell zu Grunde gelegten Beurteilungsdimensionen inhaltlich distinkte und jeweils für sich relevante Sachverhalte messen, dass mithin das Vorliegen von Konstruktvalidität hinsichtlich dieser Beurteilungsdimensionen angenommen werden kann. Das Kriterium valider Evaluationsergebnisse wird in der Diskussion um die Voraussetzungen für ihre Nützlichkeit in Reformprozessen immer wieder als grundlegend angeführt [GINSBURG & RHETT 2003; NUTLEY ET AL. 2003]. Empirisch validierte Evaluationsergebnisse können ein evidenzbasiertes Wissen über die Situation sowie zukünftig zu behebbende Defizite der jeweils evaluierten Gegenstände -bzw. wie im vorliegenden Fall: Studienprogramme- liefern und einen wichtigen Beitrag zu deren Weiterentwicklung liefern. Bezogen auf den vorliegenden Anwendungsfall ließe sich im Anschluss an die berichteten Analysen bspw. nach Optimierungspotenzialen innerhalb der verschiedenen Qualitätsdimensionen suchen. Darüber hinaus lässt sich die Frage klären, in welchen Beziehungen die verschiedenen Urteilskategorien untereinander stehen und langfristig analysieren, welche Stellschrauben jeweils zu einer Erhöhung des Gesamterfolges des Studienprogramms betätigt werden müssen.



## Literatur

BORTZ, JÜRGEN U. NICOLA DÖRING (2002): *Forschungsmethoden und Evaluation*. 3. Aufl. Berlin u.a.: Springer.

BÜHNER, MARKUS (2004): *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson.

BYRNE, BARBARA M. (2001): *Structural Equation Modeling with Amos. Basic Concepts, Applications, and Programming*. Multivariate Applications Book Series, New Jersey/London: Lawrence Earlbaum Ass. Publishers.

ENGEL, UWE U. GABY KREKELER (2001): Studienqualität. Über Studentische Bewertungen und Rankings von Studienfächern einer Universität. In: U. Engel (Hrsg.) *Hochschul-Ranking. Zur Qualitätsbewertung von Studium und Lehre*. 121-176. Frankfurt/Main: Campus

ENGEL, UWE U. PHILIPP POHLENZ (2001): Lehre und Studium im Spiegel studentischer Bewertungen: Über das Potsdamer Modell der Lehrevaluation. In: C. Spiel (Hrsg.) *Evaluation universitärer Lehre - zwischen Qualitätsmanagement und Selbstzweck*. 131-150. Münster u.a.: Waxmann

ESSER, JÜRGEN (1997): Zweifel an der Evaluation der Lehre. *WiSt*, 1. 45-49.

FALTER, JÜRGEN W. (1977): Zur Validierung theoretischer Konstrukte - wissenschaftstheoretische Aspekte des Validierungskonzepts. *Soziologie* 6(4): 349-369.

GINSBURG, ALAN & NANCY RHETT (2003): Building a Better Body of Evidence: New Opportunities to Strengthen Evaluation Utilization. *American Journal of Evaluation* 24(4): 489-498.

HAGER, WILLI (2000): Wirksamkeits- und Wirksamkeitsunterschiedshypothesen, Evaluationsparadigmen, Vergleichsgruppen und Kontrolle In: H. Brezinger (Hrsg.) *Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien*. 180-201. Bern u.a.: Huber

HOPKO, DEREK R. (2003): Confirmatory Factor Analysis of the Math-Anxiety Rating Scale-Revised. *Educational and Psychological Measurement* 63(2): 336-351.

KLINE, REX B. (1998): *Principles and Practice of Structural Equation Modeling*. Methodology in the Social Sciences, New York, London: Guilford Press.

KROMREY, HELMUT (1998): *Empirische Sozialforschung*. Opladen: Leske und Budrich (UTB).

KROMREY, HELMUT (2001): Informationsgehalt und Validität von Studierendenbefragungen als Evaluation. In: D. Müller-Böling, S. Hornbostel u. S. Berghoff (Hrsg.): Hochschulranking. Aussagefähigkeit, Methoden, Probleme. Gütersloh: Bertelsmann-Stiftung. 43-65.

KROMREY, HELMUT (2001a): Evaluation von Lehre und Studium - Anforderungen an Methodik und Design. In: C. Spiel (Hrsg.): Evaluation universitärer Lehre – zwischen Qualitätsmanagement und Selbstzweck. S. 21-60. Münster: Waxmann.

LEVACIC, ROSALIND & RON GLATTER (2001): 'Really Good Ideas'? Developing Evidence-Informed Policy and Practice in Educational Leadership and Management. *Educational Management and Administration*. 29(1), 5-25.

MÜLLER-BÖLING, DETLEF (2000): *Die entfesselte Hochschule*. Gütersloh: Bertelsmann Stiftung.

NUTLEY, SANDRA, ISABEL WALTER & HUW T. O. DAVIES (2003): From Knowing to Doing. A Framework for Understanding the Evidence-into-Practice Agenda. *Evaluation* 9(2): 125-148.

POHLENZ, PHILIPP U. KAREN TINSNER (2004): *Determinanten des Studienabbruchs. Eine empirische Untersuchung zu Gründen und Verantwortlichkeiten*. Reihe Potsdamer Beiträge zur Lehrevaluation, 1. Potsdam: Universitätsverlag.

PROOJEN, JAN-WILLEM VAN U. WILLEM A. VAN DER KLOOT (2001): Confirmatory Analysis of Exploratively Obtained Factor Structures. *Educational and Psychological Measurement* 61(5): 777-792.

RINDERMANN, HEINER (2001): Die studentische Beurteilung von Lehrveranstaltungen - Forschungsstand und Implikationen. In: C. Spiel (Hrsg.): Evaluation universitärer Lehre – zwischen Qualitätsmanagement und Selbstzweck. S. 61-88. Münster: Waxmann.

SANDERS, JAMES R. (1999): Handbuch der Evaluationsstandards - Die Standards des Joint Committee on Standards for Educational Evaluation. Opladen: Leske und Budrich.

SCHNELL, RAINER, PAUL B. HILL U. ELKE ESSER (1999): *Methoden der empirischen Sozialforschung*. München, Wien: Oldenbourg.

SCHUBARTH, WILFRIED, KARSTEN SPECK, ULRIKE GROBE, ANDREAS SEIDEL U. CHARLOTTE CHUDOBA (2005): Die zweite Phase der Lehrerbildung aus Sicht der Brandenburger Lehramtskandiatinnen und Lehramtskandidaten (Potsdamer LAK-Studie 2004/05). Abschlussbericht. Potsdam: Universität Potsdam.

SPIEL, CHRISTIANE (2001): Der differentielle Einfluss von Bias-Variablen auf studentische Lehrveranstaltungsbewertungen In: U. Engel (Hrsg.) *Hochschul-Ranking. Zur Qualitätsbewertung von Studium und Lehre*. 61-82. Frankfurt/Main: Campus.

STEYER, ROLF & MICHAEL EID (1993): Messen und Testen. Berlin u.a.: Springer.

WISSENSCHAFTSRAT (1996): *Empfehlungen zur Stärkung der Lehre in den Hochschulen durch Evaluation*. Bonn.