

Schneider, Christoph; Bodensohn, Rainer
**Zur Wahrnehmung von Lehrerhandeln. Versuch des Perspektivenabgleichs
zwischen Lehrern, Schülern und Fremdbeobachtern**

Lehrerbildung auf dem Prüfstand 4 (2011) 2, S. 319-348



Quellenangabe/ Reference:

Schneider, Christoph; Bodensohn, Rainer: Zur Wahrnehmung von Lehrerhandeln. Versuch des Perspektivenabgleichs zwischen Lehrern, Schülern und Fremdbeobachtern - In: Lehrerbildung auf dem Prüfstand 4 (2011) 2, S. 319-348 - URN: urn:nbn:de:0111-pedocs-147261 - DOI: 10.25656/01:14726

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-147261>

<https://doi.org/10.25656/01:14726>

in Kooperation mit / in cooperation with:

V E P

www.vep-landau.de

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Herausgeber

Rainer Bodensohn, Andreas Frey und Reinhold S. Jäger
Bürgerstraße 23, 76829 Landau/Pfalz
Telefon: +49 6341 280 32165, Telefax: +49 6341 280 32166

Verlag

Empirische Pädagogik e. V.
Bürgerstraße 23, 76829 Landau/Pfalz
Telefon: +49 6341 280 32180, Telefax: +49 6341 280 32166
E-Mail: info@vep-landau.de
Homepage: <http://www.vep-landau.de>

Umschlaggestaltung

© Harald Baron

Druck

DIFO Bamberg

Alle Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung, werden vorbehalten. Kein Teil des Werks darf in irgendeiner Form (durch Fotografie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung des Verlags reproduziert oder unter Verwendung elektronischer Systeme verbreitet werden.

ISSN 1867-2779

© Verlag Empirische Pädagogik, Landau 2011

Inhalt

Editorial	153
Originalarbeiten	
Keller-Schneider, M.: Die Bedeutung von Berufswahlmotiven von Lehrpersonen in der Bewältigung beruflicher Anforderungen in der Berufseingangsphase	157
König, J. & Herzmann, P.: Lernvoraussetzungen angehender Lehrkräfte am Anfang ihrer Ausbildung. Erste Ergebnisse aus der wissenschaftlichen Begleitung des Kölner Modellkollegs Bildungswissenschaften	186
Konrad, K.: Qualifizierung von Lehrkräften zur Förderung von selbst- gesteuertem Lernen: Konzeption und erste Ergebnisse einer Fortbildungsreihe	211
Leonhard, T. & Rihm, T.: Erhöhung der Reflexionskompetenz durch Begleitveranstaltungen zum Schulpraktikum? – Konzeption und Ergebnisse eines Pilotprojekts mit Lehramtsstudierenden	240
Lohmann, V., Seidel, V. & Terhart, E.: Bildungswissenschaften in der universitären Lehrerbildung: Curriculare Strukturen und Verbindlichkeiten	271
Mohr, S. & Ittel, A.: Zum Zusammenhang von Kompetenzeinschätzung und individuellen Merkmale bei Lehramtsstudierenden technisch-orientierter Fächer	303
Schneider, C. & Bodensohn, R.: Zur Wahrnehmung von Lehrerhandeln. Versuch des Perspektivenabgleichs zwischen Lehrern, Schülern und Fremdbeobachtern	319
Weiß, S., Lerche, T. & Kiel, E.: Der Lehrberuf: Attraktiv für die Falschen?	349
Informationen	
Blömeke, S., Zlatkin-Troitschanskaia, O. & Buchholtz, C.:	368
Impressum	372

Contents

Articles

- Keller-Schneider, M.:
The influence of career choice motives of teachers in coping
with professional demands in career entry phase 157
- König, J. & Herzmann, P.:
Learning preconditions of future teachers at the beginning of
pre-service teacher education. First results of the evaluation of a
model program at the University of Cologne 186
- Konrad, K.:
Qualifying teachers to promote self-regulated learning: Concept and
first results of a teacher training program 211
- Leonhard, T. & Rihm, T.:
Can courses accompanying school internships foster the development
of the reflective competence of student teachers? – Conception and
results of a pilot project 240
- Lohmann, V., Seidel, V. & Terhart, E.:
Educational themes and topics in teacher education: an analysis
of curricula in German universities 271
- Mohr, S. & Ittel, A.:
The association between student teachers' competencies, age, and
personality factors in STEM education 303
- Schneider, C. & Bodensohn, R.:
Perception of teacher behaviour. An attempt to model convergence
between ratings of teachers themselves, external observers, and the
pupils' view 319
- Weiß, S., Lerche, T. & Kiel, E.:
Does the teacher profession attract the wrong students? 349

Originalarbeiten

Christoph Schneider und Rainer Bodensohn

Zur Wahrnehmung von Lehrerhandeln. Versuch des Perspektivenabgleichs zwischen Lehrern, Schülern und Fremdbeobachtern

Zusammenfassung: Zur Beurteilung von Lehrerverhalten aus der Perspektive der Lehrer selbst und aus der Perspektive von Fremdbeurteilern können Systematiken pädagogisch-fachlicher Kompetenzen wie die von Oser (2001) oder der KMK (2004) herangezogen werden, zusätzlich kann die Betrachtung um allgemein für berufliches Handeln bedeutsame Sozial- und Methodenkompetenzen angereichert werden. Die entsprechenden Instrumentarien sind allerdings für die Beurteilung von Lehrerhandeln aus Schülersicht nicht geeignet. In dieser Studie wird das Handeln angehender Lehrkräfte in Schulpraktika in Selbst- und Fremdsicht anhand der Systematiken von Oser (2001) und Frey (2008) beurteilt, für die Schülersicht wird das Instrument von Schneider & Bodensohn (2008) verwendet. Es wird untersucht, ob sich zwischen diesen drei Perspektiven eine Konvergenz in der Wahrnehmung von Lehrerhandeln ausmachen lässt. In Anlehnung an die MTMM-Methodik kann in einem ersten Schritt gezeigt werden, dass es innerhalb der Perspektiven zwar starke Methodeneffekte gibt, diese Methoden aber korreliert sind. Strukturhypothesen dazu, wie Facetten der professionellen Sicht und der Schülersicht inhaltlich zusammenhängen, werden in einem zweiten Schritt anhand einer Expertenbefragung generiert. Diese Strukturhypothesen können allerdings in einem dritten Analyseschritt die empirischen Zusammenhänge zwischen den Perspektiven nicht beschreiben. Daraus kann gefolgert werden, dass eine Konvergenz zwischen den Perspektiven zumindest auf diesem Weg nicht darstellbar ist.

Schlagwörter: Lehrerhandeln – MTMM – Praktika – Wahrnehmung

Perception of teacher behaviour. An attempt to model convergence between ratings of teachers themselves, external observers, and the pupils' view

Abstract: Inventories of teacher's pedagogical competencies (Oser, 2001; KMK, 2004) can be employed for the assessment of teacher behaviour in the classroom, along with social and methodological competencies not specific to the teaching domain. Such instruments, however, are not adequate for the assessment of teacher behaviour from the pupils' perspective. Within this study, in-classroom behaviour of university students in teacher training is observed by means of the Oser (2001) and the Frey (2008) instruments. The pupils' view is assessed with the Schneider and Bodensohn (2008) questionnaire. The research question is if there is convergence between the three perspectives of the to-be-teachers' self-ratings, ratings by external observers, and the pupils' view. Within an approach resembling the fundamental thought of the MTMM method, it is first shown that each perspective encompasses a strong method effect. These methods are correlated, however. In a second step, expert ratings are employed to generate structural hypotheses on contentual overlap of facets of professional observation and the pupils' view. In a third step though, these hypotheses prove to be inadequate for modelling the empirical interrelations. It is concluded that, at least with our approach, a convergence between the perspectives cannot be yielded.

Key words: MTMM – perception – teacher acting – teacher training

1. Erkenntnisstand

Die Wirkung des Lehrerhandelns in Schulen und darüber hinaus in der Lehrerbildung wissenschaftlich in den Blick zu nehmen, kann heute als allgemein anerkanntes Desiderat bezeichnet werden. Dazu mangelt es nicht an Orientierungen. Wang, Haertel und Walberg (1993) haben in ihrer Metaanalyse die Erkenntnisse über Wirkungen im Unterricht bilanziert. Helmke (2003) hat Studienergebnisse in einem Lehrbuch zur Unterrichtsqualität zusammengetragen. In einer Expertise im Auftrag der Kultusministerkonferenz (KMK) hat Terhart (2002) ausgeführt, dass sich die Evaluation auf die gesamte Wirkungskette des Lehrerhandelns in den Phasen der Lehrerbildung bis zu den Lernerfahrungen und Wirkungen bei den Schülern¹ beziehen muss. Folglich sind alle Beteiligten der Lehrerbildung an geeigneter Stelle in die Evaluation einzubinden. Eine Systematik hierzu bieten etwa die von der KMK (2004) herausgegebenen „Standards für die Lehrerbildung“.

Wie Studierende und Ausbilder in Schulpraktika Kompetenzen beobachten bzw. beurteilen, wird empirisch seit gut einem Jahrzehnt vielfach in den Blick genommen (vgl. Balzer, Bodensohn & Frey, 2004; Bodensohn & Schneider, 2006, 2008b; Frey, 2006, 2008; Hascher, 2011; Hascher & Moser, 2001; Oser, 2001). Dass dabei das Einbeziehen von Schülerurteilen in Evaluationsprozesse unverzichtbar ist, wird heute kaum noch ernstlich angezweifelt (vgl. Clausen, 2002; Kunter & Baumert, 2006; Schneider & Bodensohn 2008; Wubbels, Brekelmans & Hoymayers, 1992). Gruehn (2000) konnte in den Studien des MPI aufzeigen, dass Schülerurteile differenziert unterrichtliche Bedingungen und Lernarrangements bewerten können. Bezüglich der Schülerurteile fassen Schwarz und Prange (1997, S. 134) zusammen, dass deren Validität nicht grundsätzlich bestritten wird und sich die Kontroversen eher auf messtheoretische Probleme beziehen.

Die Ergebnisse der Landauer Studie zur Struktur der Schülerurteile (Bodensohn & Schneider, 2008a) lieferte Anhaltspunkte dafür, dass Schüler offensichtlich in breiten, gut gegeneinander abgrenzbaren Kategorien denken, um ihre Lehrkräfte im Praktikum zu beurteilen. Auf die Nähe der gewonnenen Kategorien zu etablierten führungspsychologischen Konstrukten (Iowa-Studien: Kurt Lewin: Lück, 1996; Pennsylvania-Studies: Landy und Lamiell-Landy, 1978; Ohio-State-Studies: Fleishman, 1953; Michigan Studies: Katz & Kahn, 1952; Vroom & Jago, 1991) ist bei Schneider und Bodensohn (2008) hingewiesen worden.

Derzeit wird geklärt, wie Schüler als Experten von Unterricht Unterrichtsqualität und Feedbackkulturen unterstützen können (Helmke et al., 2009), nicht geklärt ist dagegen, wie und ob Schülerurteile für die Lehrerbildung genutzt werden können

¹ Die Verwendung der männlichen Form dient der verbalen Vereinfachung. Gemeint sind selbstverständlich beide Geschlechter.

(vgl. Helmke, 2009). Einige Studien zur Unterrichtsqualität (Clausen, 2002; Kunter & Baumert, 2006) deuten darauf hin, dass es wenig Übereinstimmung zwischen den Einschätzungen von Lehrkräften, externen Experten und Schülern gibt. In dieser Studie wird der Frage nachgegangen, ob spezifisch bezüglich Facetten des Lehrerhandelns eine Konvergenz zwischen den verschiedenen Perspektiven existiert.

1.1 Lehrerhandeln aus professioneller und wissenschaftlicher Sicht

1.1.1 Beschreibungssysteme zu professionellem Lehrerhandeln

Zur Beschreibung des Praxishandelns und der dabei relevanten Kompetenzen angehender Lehrkräfte existieren im deutschen Sprachraum zumindest zwei dominierende Ansätze. Erstens ist die bekannte Systematisierung von Oser (2001) zu nennen, die originär zur Beurteilung der Leistungsfähigkeit der Lehrerbildungssysteme entworfen wurde. Sie beschreibt – aufbauend auf Expertenübereinkünften bezüglich minimal von praktisch tätigen Lehrkräften im Schulalltag einzufordernden Handlungskompetenzen – insgesamt 88 verhaltensnah formulierte Standards, die zwölf Standardbereichen zugeordnet sind. Hierbei handelt es sich vornehmlich um eine normative, an den realen Anforderungen im „Alltagsgeschäft“ der Lehrkraft orientierte Zuordnung. Durch Befragungen Schweizer Lehramtsstudierender, die Oser (2001) durchführte, kommt er zu dem Schluss: „Die Professionalisierung als komplexer Kompetenzerwerb in der Ausbildung zum Lehrerberuf liegt im Argen. Man ist erstaunt, feststellen zu müssen, dass die Verarbeitungstiefe bei zentralen Fähigkeiten zur Bewältigung von Aufgaben in diesem Beruf kaum ausgeschöpft wird ...“ (S. 310) und stellt damit der Lehrerbildung (zu diesem Zeitpunkt) ein schlechtes Zeugnis aus.

Zweitens bewog auch in Deutschland das Fehlen einer empirischen Evaluation und Wirkungsanalyse (Terhart, 2002) dazu, eine Aufstellung dessen zu schaffen, was die Lehrerbildung zu leisten in der Lage sein sollte (KMK, 2004). Obwohl verschiedentlich angemahnt wurde, dass diese „Standards der Lehrerbildung“ nicht wissenschaftliche Begrifflichkeiten abbilden (vgl. Baumert & Kunter, 2006), haben sie doch in den vergangenen Jahren die Sensibilität für die Standardorientierung in der Lehrerbildung insgesamt wecken bzw. steigern können.

Professionsstandards, die nicht explizit auf den Lehrerberuf, sondern breiter auf berufliches Handeln im Allgemeinen abzielen, sind Gegenstand des hierarchischen Ansatzes der Strukturierung von Kompetenzen von Frey (2006; Frey, Jäger & Renold, 2005). Insbesondere die Klassen der Sozial- und Methodenkompetenzen sind in diesem Ansatz den (etwa im Sinne des Ansatzes von Oser konzipierten) pädagogischen Fachkompetenzen beim Blick auf die Kompetenzen von Lehrkräf-

ten gleichberechtigt zur Seite gestellt. Zur Erfassung dieser beiden Kompetenzklassen wurde von Frey und Balzer (2003) ein Erhebungsbogen veröffentlicht.

1.1.2 Evaluation der Kompetenzen angehender Lehrkräfte in den Landauer Studien

Im Rahmen des Projekts REBHOLZ² wurden in den Jahren 2005-2010 zur längsschnittlichen Evaluation der Kompetenzentwicklung Lehramtsstudierender in Vollerhebung Instrumente eingesetzt, die unter anderem pädagogische Fachkompetenzen im Sinne von Oser (2001) sowie Sozial- und Methodenkompetenzen in der Fassung von Frey (2008) messen. Neben einer Eingangsbefragung, in der neben soziodemografischen Angaben auch Motive zur Aufnahme eines Lehramtsstudiums und weitere Selbstauskünfte erhoben wurden, umfasste das Projekt jeweils zum Ende zweier längerer Schulpraktika die Selbsteinschätzung der Studierenden und die Fremdeinschätzung durch die Praktika betreuende Mentoren jeweils bezüglich der Ausprägung der Fach-, Sozial und Methodenkompetenzen. Für eine genauere Beschreibung des Projektdesigns, der gemessenen Merkmale und im Längsschnitt nachweisbarer Kompetenzentwicklungen sei auf Bodensohn und Schneider (2008b; vgl. auch Bodensohn & Schneider, 2009; Schneider & Bodensohn, 2008) verwiesen. Befunde zu wahrgenommener Wichtigkeit der erhobenen Fachkompetenzen und ihrer tatsächlichen Relevanz im Hinblick auf die Praktikumsituation aus der Sicht der Praktikanten und ihrer Mentoren wurden von Schneider und Bodensohn (2007; Bodensohn & Schneider, 2010) vorgelegt.

1.1.3 Zur Binnenstruktur des in REBHOLZ eingesetzten Instrumentariums

Angesichts der oben kurz umrissenen Entstehungsgeschichte der von Oser (2001) vorgelegten Systematisierung zu fachlich-pädagogischen Elementen erfolgreichen Lehrerhandelns und der Einteilung derselben in zwölf Standardbereiche ist nicht davon auszugehen, dass sich die zwölf Bereiche in empirischen Strukturanalysen als orthogonale, d. h. unabhängig voneinander variiierende Faktoren darstellen lassen. Diese Erwartung lässt sich sehr deutlich an den Daten einer der Befragungskohorten aus REBHOLZ bestätigen. Zwischen den Selbsteinschätzungen der Studierenden ($N = 569$)³ auf den insgesamt 13 zur Erhebung pädagogischer Fachkompetenzen verwendeten Skalen⁴ finden sich ausnahmslos sta-

² Projekt „Report beruflicher Handlungskompetenz im Organisationsbereich der Lehrerbildung des Zentrums für Lehrerbildung in Landau“; www.uni-landau.de/schulprakt-studien/rebholz.htm

³ Dies bezeichnet die Größe des Untersuchungskollektivs. Aufgrund fehlender Werte liegen die Fallzahlen für die auf der Grundlage paarweisen Fallausschlusses berechneten bivariaten Korrelationen teils erheblich darunter.

⁴ Es handelt sich um Skalen, die in enger Anlehnung an die Oser'schen Standardbereiche entwickelt wurden. „Schule und Öffentlichkeit“ fällt in REBHOLZ weg, da dieser Kompetenzbereich in nur wenige Wochen dauernden Schulpraktika keine Rolle spielt. Anstelle des bei Oser (2001) inhaltlich sehr

tistisch bedeutsame (alle $p < .01$), positive Interkorrelationen. Die mittlere Interkorrelation zwischen den 13 Skalen beträgt hierbei .47, die geringste Interkorrelation .19 und die größte .78. Noch deutlicher sind die Interdependenzen zwischen den Skalen im Fremdurteil ($N = 404$) ausgeprägt: Hier beträgt die mittlere Skaleninterkorrelation .64 (*Range*: .45 bis .88).

Im Kontrast zu den hohen Interkorrelationen der Skalen *innerhalb* einer Beobachtungsperspektive sind die Korrelationen *zwischen* den Perspektiven der Selbst- und der Fremdeinschätzung ($N = 380$) deutlich geringer ausgeprägt: Die Korrelationen inhaltsgleicher Skalen zwischen Selbst- und Fremdeinschätzung – im Sinne der Multitrait-Multimethod-Analyse (MTMM, vgl. Ostendorf et al, 1986) handelt es sich hier um die „konvergenten“ Korrelationen – sind zwar alle positiv, betragen aber im Mittel nur .22 (*Range*: .13 bis .30). Nur teilweise niedriger sind diejenigen Korrelationen zwischen Selbst- und Fremdeinschätzung, die sich auf unterschiedliche Skalen beziehen. Diese betragen im Mittel .19 (*Range*: .03 bis .31).

Bezüglich der Binnenstruktur der pädagogischen Fachkompetenzen lässt sich angesichts dieser Interkorrelationen das Zwischenfazit ziehen, dass (a) sowohl Selbst- als auch Fremdwahrnehmung der Kompetenzbereiche hochgradig interdependent sind, so dass jeweils eine allen Kompetenzfacetten gemeinsame Varianzquelle im Sinne eines Generalfaktors (bzw. „Methodenfaktors“) anzunehmen ist und dass (b) Selbst- und Fremdwahrnehmung nur lose miteinander assoziiert sind.

Ähnliche Befunde berichten Frey und Balzer (2003) für die Bereiche der Sozial- und der Methodenkompetenzen in Bezug auf Selbstbeschreibungen. Unter den insgesamt sieben Skalen, die die Aspekte der Sozialkompetenzen abbilden, beträgt die mittlere Korrelation .59, unter den fünf Skalen zu Methodenkompetenzen .63. Auch variieren Sozial- und Methodenkompetenzen nicht unabhängig voneinander (mittlere Korrelation .53). Für eine zumindest konzeptionelle Abgrenzbarkeit der beiden Bündel Sozial- und Methodenkompetenzen spricht jedoch, dass es den Autoren gelang, die beiden Kompetenzbündel in einer Faktorenanalyse mit obliquen Rotation auf zwei Faktoren abzubilden. Auch für die beiden Bündel der Sozial- und der Methodenkompetenzen kann angesichts dieser Bezüge folgerichtig ebenfalls jeweils eine gemeinsame Varianzquelle im Sinne eines Methodenfaktors angenommen werden.

breit gefassten Bereichs „Allgemeindidaktische und fachdidaktische Kompetenzen der Lehrkraft“ wurden drei separate Skalen („Fachdidaktische Gesichtspunkte“, „Allgemeindidaktische Gesichtspunkte: Unterrichtsplanung“ und „Allgemeindidaktische Gesichtspunkte: Unterrichtsdurchführung“) eingesetzt.

1.2 Lehrerhandeln aus der Sicht der Schüler

1.2.1 Lehrerhandeln und Unterrichtsqualität – austauschbare Begriffe?

Im Angebots-Nutzen-Modell zur Wirkungsweise des Unterrichts (vgl. Fend, 2008; Helmke, 2009) werden die Ergebnisse des Unterrichts als Funktion einerseits der Qualität und Quantität der Lernangebote durch die Lehrkraft und andererseits der Qualität und Quantität der Wahrnehmung bzw. Abfrage dieser Lernangebote durch die Schüler verstanden. Bemisst man die Qualität des Unterrichts an seinem Ertrag, der sich vornehmlich in fachlichen und überfachlichen Lernergebnissen auf der Seite der Schüler niederschlägt, so ist im Rahmen des Modells klar, dass dieser Ertrag nicht alleine auf die durch das Verhalten der Lehrkraft geschaffenen unterrichtlichen Bedingungen zurückzuführen sein kann. Vielmehr handelt es sich beim Lehrerhandeln nur um *eine* – wenn auch möglicherweise die Wesentlichste – der Entstehungsbedingungen für eine hohe Qualität des Unterrichts.

Dass Unterrichtsqualität sich nicht singular auf „richtige“ Verhaltenselemente der Lehrkraft zurückführen lässt, sondern ein Produkt des Interaktionsprozesses – und damit auch der „Passung“ – zwischen Lehrkraft und Schülern darstellt, wird auch deutlich, wenn man „Kataloge“ empirisch gesammelten Wissens zu Merkmalen guten Unterrichts (Helmke, 2009; Meyer, 2003) als Grundlage der Einschätzung heranzieht. Dekliniert man inhaltlich, welche dieser Merkmale alleine „gutes“ Lehrerhandeln reflektieren, so wird deutlich, dass dies zwar auf die Mehrzahl, nicht aber auf sämtliche der zusammengetragenen Merkmale zutrifft, da etwa ein „lernförderliches Klima“ eben nicht allein auf den durch die Lehrkraft gegebenen Input zurückzuführen sein kann. Insofern ist (gutes) Lehrerhandeln möglicherweise als notwendige, nicht aber als hinreichende Bedingung für das Zustandekommen einer hohen Unterrichtsqualität anzusehen. Für die Praxis birgt dies die Folgerung, dass spezifisch zur Beobachtung bzw. Messung der Qualität von Lehrerhandeln nicht unreflektiert Instrumente, die den Anspruch der Messung von Unterrichtsqualität erheben, eingesetzt werden sollten.

1.2.2 Messung von Unterrichtsqualität und Übereinstimmung der Wahrnehmung über verschiedene Perspektiven

In Studien, die gezielt Aspekte der Unterrichtsqualität in den Blick nehmen (etwa Clausen, 2002; Gruehn, 2000; Kunter & Baumert, 2006; Wagner, 2008) werden zumeist in verschiedenen Large-Scale-Assessments eingesetzte Skalen verwendet, deren Inhalte und Akzentuierungen aus Inventarisierungen des Wissens zu Merkmalen guten Unterrichts abgeleitet sind. Explizit mit der Frage nach der perspektivenübergreifenden Gleichheit der Wahrnehmungsstruktur beschäfti-

gen sich Kunter und Baumert (2006) und Clausen (2002). Zur Erfassung der verschiedenen Perspektiven der Beteiligten (Schüler, Lehrkraft und externe Beobachter) werden in diesen Arbeiten im Sinne der Vergleichbarkeit zumeist Items mit gleichem oder ähnlichem, lediglich verbal an die Perspektive angepasstem, Inhalt verwendet.

Obwohl die genannten Studien in ihrem Anspruch der Messung von Merkmalen der Unterrichtsqualität durchaus unterschiedliche Wege gehen, was die Anzahl und Inhalte der Skalen angeht, kommen sie doch bezüglich der Frage nach der Vergleichbarkeit der Wahrnehmungsstrukturen zu vergleichbaren Ergebnissen: Von Clausen (2002) wurde zur Prüfung der Übereinstimmung der faktoriellen Struktur der Unterrichtswahrnehmung über die drei Perspektiven hinweg im Sinne der Multitrait-Multimethod-Analyse das Verfahren der Constrained-Component-Analysis (CCA; Kiers, Takane & Berge, 1996) eingesetzt. Bei diesem Verfahren handelt es sich um einen Ansatz, der die bei der Anwendung der klassischen Verfahren der konfirmatorischen Faktorenanalyse auf MTMM-Fragestellungen oft auftretenden Konvergenz- und Schätzprobleme umgeht, andererseits aber nicht den gängigen und in kommerziell vertriebenen Softwarepaketen enthaltenen Auswertungsverfahren zugeordnet werden kann. Anhand dieser Auswertungsmethodologie kommt Clausen (2002) zu dem Schluss, dass sowohl die Wahrnehmungen zu Unterrichtsqualität über die drei Perspektiven hinweg unterschiedliche Binnenstrukturen aufweisen, als auch die Übereinstimmungen bezüglich der betrachteten Merkmale gering sind.

Auch von Kunter und Baumert (2006) werden geringe Übereinstimmungen zwischen der Wahrnehmungsstruktur von Lehrern und Schülern berichtet. Auf der Basis von 28 gleichlautenden Items konnten hier in explorativen Faktorenanalysen bei der Lehrerperspektive acht, bei der Schülerperspektive jedoch nur fünf Faktoren gefunden werden, von denen lediglich zwei („Ineffektive Klassenführung“ und „Inadäquates Interaktionstempo“) vergleichbare Inhalte aufwiesen. Weiterführende konfirmatorische Analysen legten hier den Schluss nahe, dass den Items insbesondere in der Schülersicht ein Generalfaktor zugrunde liegt, während dies auf die Lehrersicht nicht zutrifft. Diese Befunde deuten insgesamt darauf hin, dass zwischen der Sicht von Lehrkräften, Schülern und externen Beobachtern auf die Unterrichtsqualität davon ausgegangen werden kann, dass (a) die Kategorien der Wahrnehmung nicht universell sind und dass (b) eine Konvergenz der Sichtweisen („Kommen die verschiedenen Gruppen von Beteiligten zu ähnlichen Schlüssen bezüglich konkreten Unterrichts?“) kaum gegeben ist.

1.2.3 Messung des Lehrerverhaltens aus Schülersicht

In Abschnitt 1.1.2 wurde das in den Landauer Studien verwendete Instrumentarium zur Erfassung der Kompetenzen angehender Lehrkräfte in den schulprakti-

schen Phasen des ersten Ausbildungsabschnitts beschrieben. Der Einsatz der Skalen zu fachlichen (d. h. hier: pädagogischen) Kompetenzen in Anlehnung an Oser (2001) und allgemeinen für berufliches Handeln bedeutsamen Sozial- und Methodenkompetenzen (nach Frey, 2008) setzt die Fähigkeit zur professionellen Reflexion des in den ersten Unterrichtsversuchen im Praktikum gezeigten Handelns durch die Brille eines auf der Basis von Empirie und Expertenbefragungen generierten Instrumentariums voraus. Angesichts der Menge der hiermit vorzunehmenden Kompetenzeinschätzungen und der verbalen Komplexität der Iteminhalte wäre es jedoch verfehlt, den von den Praktikanten unterrichteten Schülern – überwiegend Sekundarstufe I –, das gleiche Instrumentarium zur Einschätzung des Verhaltens „ihres“ Praktikumslehrers an die Hand zu geben.

Da gleichwohl in den Landauer Studien frühzeitig der Wunsch entstand, neben der „professionellen“ Kompetenzeinschätzung der angehenden Lehrkräfte in Selbstbeschreibung und durch die betreuenden Mentoren in den Praktikumschulen in Fremdbeschreibung auch die von den Praktikanten unterrichteten Schüler selbst zu Wort kommen zu lassen, ergab sich der Bedarf, hierfür ein geeignetes Instrument heranzuziehen. In der Sichtung der auf dem Markt verfügbaren Instrumente (vgl. Ditton & Arnold, 2004; Gerstenmaier, 1975; Haecker & Werres, 1983; Kahl, Buchmann & Witte, 1977; Martin, 1995; Mayer & Nickolaus, 1998;) stellte sich keines als uneingeschränkt geeignet für den speziellen Befragungszweck heraus. Vor diesem Hintergrund wurde im Rahmen von REBHOLZ ein eigenes Instrument zur Messung von Lehrerhandeln (genauer: Handeln der angehenden Lehrkräfte im Praktikum) geschaffen. Dieses Instrument, dessen Konstruktion im Detail bei Bodensohn und Schneider (2008) beschrieben ist, besitzt sechs Skalen die auf der Basis der Ergebnisse explorativer Faktorenanalysen auf Klassenebene aggregierter ipsatierter Daten gebildet wurden. Die Benennungen, Reliabilitäten und Interkorrelationen der Skalen sind in Tabelle 1 dargestellt. Obwohl es sich um auf der Basis ipsatierter Daten erhaltene orthogonale Faktoren handelt, die der Skalenbildung zugrunde liegen, sind die auf der Basis der nicht-ipsatierten Rohwerte gebildeten Skalen auf Klasseneben substanzial interkorreliert. Insofern kann in der Folge konzeptionell nicht davon ausgegangen werden, dass sich die Schülerwahrnehmung in voneinander unabhängigen Aspekten vollzieht. Es ist anzunehmen, dass in Übereinstimmung mit den oben beschriebenen Ergebnissen von Kunter und Baumert (2006) zur Schülerwahrnehmung von Unterrichtsqualität auch den Aspekten der Schülerwahrnehmung von Lehrerhandeln eine generelle Beobachtungstendenz in Form eines Generalfaktors zugrunde liegt.

Tabelle 1: Innere Konsistenzen (Cronbachs α) und Interkorrelationen der anhand der explorativen Faktorenanalyse gebildeten Skalen

Nr.	Name der Skala	Anzahl Items	α	Korrelation* zu Skala Nr,				
				1	2	3	4	5
1	Autokratie	3	.80					
2	Anregungsgehalt des Unterrichts	4	.92	-.52				
3	Lernkultur	5	.88	-.48	.82			
4	Wertschätzung	3	.91	-.76	.72	.66		
5	Klare Strukturvorgaben	4	.86	-.48	.81	.74	.67	
6	Einbindung der Schüler	4	.89	-.63	.81	.76	.83	.82

* Berechnung auf der Basis auf Klassenebene aggregierter Rohwerte, $N = 321$; für alle Korrelationen $p < .01$.

2. Fragestellung und Methodik

2.1 Globale Fragestellung

Das Verhalten der angehenden Studierenden in den schulpraktischen Phasen des universitären Ausbildungsabschnitts wird in den Landauer Studien aus drei Perspektiven betrachtet: Selbsteinschätzung der Studierenden, Fremdeinschätzung durch Betreuer und Sicht der Schüler. Während Selbst- und Fremdeinschätzung in die Aspekte der pädagogisch-fachlichen Kompetenzen, der Sozialkompetenzen und der Methodenkompetenzen, jeweils repräsentiert in Form einiger Einzelmerkmale, untergliedert sind, ist die Erhebung der Schülersicht anhand von sechs Skalen nicht in diese drei Kompetenzaspekte geteilt.

Anliegen dieser Studie ist die Klärung der Frage, ob sich – trotz der verschiedenen Erhebungszugänge – Konvergenzen in der Beurteilung des Verhaltens über die drei Perspektiven hinweg ausmachen lassen. Allerdings ist bekannt, dass bereits Fremd- und Selbsteinschätzung der Praktikanten (trotz gleichlautender Erhebungsinstrumente) nur lose miteinander assoziiert sind (vgl. Abschnitt 1.1.3), was bereits hier auf eine mangelnde Konvergenz hinweist. Im Zentrum unseres Interesses steht jedoch die Frage, ob sich die Schülersicht mit der „professionellen“ Sicht aus Selbst- und/oder Fremdwahrnehmung überschneidet. Um die evidenten Probleme zu vermeiden, die sich aus der fehlenden Übereinstimmung zwischen Selbst- und Fremdwahrnehmung ergeben, wird in zwei Teilfragestellungen separat geprüft, ob sich (a) die Schülersicht mit der Fremdbeschreibung „verträgt“ und/oder ob dies (b) zwischen Schülersicht und der Selbstbeschreibung Fall ist.

2.2 Datenstruktur und Stichprobe

Die Studierenden durchlaufen in der universitären Phase der Ausbildung jeweils zwei Praktika, das erste in aller Regel von zweiwöchiger, das zweite von vierwöchiger Dauer. Zum Ende jedes dieser Praktika schätzen sich die Studierenden anhand des oben beschriebenen Instrumentariums bezüglich ihrer fachlich-pädagogischen Kompetenzen auf 13⁵ Skalen in Anlehnung an Oser (2001) sowie auf sieben Skalen zu Sozialkompetenzen und auf fünf Skalen zu Methodenkompetenzen (Instrument von Frey, 2008) ein. Parallel dazu werden die Studierenden anhand bedeutungsgleicher Skalen von jeweils einem in der Schule für die Betreuung des Praktikanten verantwortlichen Mentor beurteilt. Die betreuenden Mentoren haben dabei überwiegend eine auf die Betreuungsaufgaben im Allgemeinen und die Evaluation der Praktikanten anhand dieses Instrumentariums im Speziellen vorbereitende Fortbildungsveranstaltung besucht. Ergänzt werden die Selbsteinschätzungen der Praktikanten und die Fremdeinschätzungen durch die Mentoren durch auf die auf Klassenebene aggregierte Wahrnehmung der Schüler mittels der in Abschnitt 1.2.3 beschriebenen sechs Skalen.

Da in dieser Studie die methodische Frage nach der Konvergenz der drei Perspektiven bezüglich der Qualität des Handelns der Praktikanten im Vordergrund steht und Fragen der längsschnittlichen Entwicklung der Handlungskompetenzen nicht berührt werden, werden ausschließlich zum zweiten Messzeitpunkt erhobene Daten ausgewertet. Es handelt sich mithin um die querschnittliche Betrachtung des „Status Quo“ zum Ende der schulpraktischen Anteile der universitären Phase der Lehrerausbildung. Anzumerken ist ferner, dass sich die Fremd- und die Selbstbeobachtung hier auf den gesamten Zeitraum des Praktikums, in dem die angehende Lehrkraft typischerweise in mehreren Klassen unterrichtet, bezieht, während die Schülerwahrnehmung Ausdruck der Sicht einer einzelnen Klasse ist, die im Praktikum besonders häufig unterrichtet wurde. Als Minimalkriterium wurde hier formuliert, dass für den Einschluss in die Studie die Lehrkraft zumindest drei Stunden in der Klasse unterrichtet haben muss, durch die ihr Verhalten beurteilt wird. Im Mittel hatten die in die Studien aufgenommenen Studierenden 5.7 (\pm 3.9) Stunden in der beurteilenden Klasse unterrichtet.

In die Analysen eingehen können nur jene Fälle, bei denen zum zweiten Messzeitpunkt (a) sowohl eine Selbsteinschätzung als auch eine Schülerbeurteilung bzw. (b) sowohl eine Fremdeinschätzung als auch eine Schülerbeurteilung vorliegt. Obwohl es sich bei den Studien in REBHOLZ dem Wesen nach um Vollerhebungen handelt, trifft dieses synchrone Vorliegen der verschiedenen Beobach-

⁵ In die Modellierungen aufgenommen wurden nur elf dieser 13 Skalen, die Skalen „Zusammenarbeit in der Schule“ und „fachdidaktische Gesichtspunkte“ wurden wegen sehr hoher Anteile fehlender Daten von den Analysen ausgenommen.

tungsmodalitäten nur auf einen Teil der Studierenden zu. Dies liegt zum einen daran, dass zwar im Rahmen der Organisation der Schulpraktika von den Studierenden selbst die Selbstbeurteilung verpflichtend eingefordert werden kann, bereits die Fremdbeurteilung durch die Mentoren, obwohl als integraler Bestandteil des Praktikums angekündigt, auch vom guten Willen und von der Gewissenhaftigkeit der Partner in den Schulen abhängt. Ausdrücklich als Angebot an die Studierenden war die Schülerbefragung konzipiert, denn zum einen kann deren Umsetzung vor Ort im Praktikum kaum von der Universität aus kontrolliert werden, zum anderen bedeutet die Durchführung der Befragung mit der Notwendigkeit, die Daten selbst vor der Auswertung einzugeben, netto einen gewissen Mehraufwand⁶.

Vor dem Hintergrund dieser Restriktionen beträgt bezüglich Teilfragestellung (a) die Anzahl der in die Analyse zur Konvergenz zwischen Fremdurteil und Schülersicht eingehende Zahl an Praktikanten $N = 177$. Vor der Berechnung der Korrelationsmatrix als Grundlage der konfirmatorischen Modellierungen in *LISREL 8.72* wurden in *PASW/SPSS 18.0* vorab bei einer Overall-Missing-Quote von 10,4 % fehlende Werte multipel imputiert. Bei der Teilfragestellung (b) der Analyse zur Konvergenz zwischen Selbsturteil und Schülersicht gingen in die analoge Berechnung $N = 321$ Fälle ein (Overall-Missing-Quote 6,7 %).

2.3 Die Multitrait-Multimethod-Analyse (MTMM) als gedankliches Leitmodell zum Vorgehen

Der MTMM-Ansatz wurde erstmals von Campbell und Fiske (1959; vgl. Ostendorf, Angleitner & Ruch, 1986) beschrieben und gilt seither als bedeutendster Ansatz zur empirischen Konstruktvalidierung. Einen umfassenden Überblick zur Konzeption des Ansatzes und aktuellen methodischen – vornehmlich konfirmatorischen – Herangehensweisen geben Schermelleh-Engel und Schweizer (2007). Der Grundgedanke ist die Forderung, dass ein valides Konstrukt grundsätzlich mit unterschiedlichen methodischen Zugangswegen erfassbar sein muss. Diese Zugangswege können verschiedene Messinstrumente mit der gleichen Messintention, aber auch verschiedene Beobachter oder Beobachtergruppen mit gleichen Messinstrumenten sein. Dargestellt werden die Zusammenhänge in einer Interkorrelationsmatrix, darin wird von „konvergenter Validität“ gesprochen, wenn die Messungen des gleichen Konstrukts über verschiedene Erhebungsmethoden hinweg (*monotrait-heteromethod*) positiv korreliert sind. Zusätzlich werden die Zusammenhänge zwischen den Messungen verschiedener Konstrukte innerhalb einer Methode betrachtet (*heterotrait-monomethod*). Sind diese Korrelationen unbedeutend, wird von „diskriminanter Validität“ gesprochen, die im

⁶ Dies soll nicht den Blick darauf verstellen, dass von den Studierenden, die Schülerbefragungen durchführten, deren individuelle Auswertungen in aller Regel als ausgesprochen informativ und bereichernd wahrgenommen wurden.

Sinne diskriminanter Validität unerwünschten gemeinsamen Varianzanteile innerhalb einer Methode oder Beobachtungsmodalität werden in diesem Zusammenhang als „Methodeneffekte“ bezeichnet.

Bezogen auf den Gegenstandsbereich der Konvergenz der Perspektiven verschiedener Gruppen von Beobachtern sieht der MTMM-Ansatz in der Regel vor, dass für die verschiedenen Perspektiven (z. B. Selbst- und Fremdbeschreibung) gleich lautende oder nur minimal verbal angepasste Items und Skalen zur Messung eines Konstrukts eingesetzt werden. Dieses Vorgehen ist etwa in der Arbeit von Clausen (2002) abgebildet. Da jedoch in der hier beschriebenen Studie für die Erfassung der Schülersicht einerseits und der „professionellen“ Sicht andererseits vor dem Hintergrund der praktischen Handhabbarkeit insbesondere für die Schüler völlig unterschiedliche Instrumentarien eingesetzt werden, handelt es sich hier nicht um eine MTMM-Analyse im engeren Sinne. Gleichwohl wird hier dem MTMM-Ansatz das methodische Herangehen in der Form konfirmatorischer Faktorenanalysen „entlehnt“.

2.4 Mehrstufiges Vorgehen zur Analyse der Konvergenz

2.4.1 Erster Schritt: Gibt es überhaupt eine Konvergenz zwischen den Perspektiven?

Angesichts der in Abschnitt 1.1.3 berichteten massiven Interkorrelationen zwischen den einzelnen Skalen innerhalb der drei Kompetenzbündel „fachlich-pädagogische Kompetenzen“, „Sozialkompetenzen“ und „Methodenkompetenzen“ sowohl in der Selbst- wie auch in der Fremdbeschreibung muss davon ausgegangen werden, dass jedem dieser Bündel ein großer – im Sinne der MTMM-Denklogik per se „unerwünschter“ – Methodeneffekt innewohnt. Gleiches muss für die Binnenstruktur der Schülerwahrnehmung (vgl. 1.2.3) angenommen werden. In Abschwächung der durch MTMM zu beantwortenden Grundfrage, ob es *überhaupt* solche die Validität der Konstrukte in Frage stellenden Methodeneffekte gibt, ist hier folglich die Frage zu stellen, ob es *zusätzlich* zu den ohnehin zu erwartenden starken Methodeneffekten überhaupt Anteile gemeinsamer Varianz *zwischen* diesen Methodenfaktoren gibt. Bevor also über die verschiedenen Beobachtungsmodalitäten hinweg inhaltliche Faktoren (Im Sinne von MTMM „Traits“ oder „Traifaktoren“ genannt) modelliert werden können, stellt sich hier die Frage, ob es Varianzanteile gibt, die dies überhaupt erst rechtfertigen würden. Falls dies nicht gegeben ist, wäre die Frage nach der *Beschaffenheit* der Traifaktoren von vorne herein obsolet.

Diese Frage wird in einem ersten Analyseschritt anhand konfirmatorischer Faktorenanalysen geklärt indem – jeweils getrennt für die beiden Teilfragestellungen – einem ersten Modell mit unkorrelierten nur-Methoden-Faktoren ein zweites Modell

mit korrelierten nur-Methoden-Faktoren gegenüber gestellt wird. Die beiden Modelle sind in ihrer Struktur in Abbildung 1 dargestellt. Dabei wird jeweils unterstellt, dass allen Skalen eines jeden Bündels der „professionellen“ Wahrnehmung (also Fach-, Sozial- und Methodenkompetenzen) sowie allen Aspekten des Schülerurteils jeweils eine gemeinsame Varianzquelle im Sinne eines Generalfaktors unterliegt. Erweist sich Modell 2 gegenüber Modell 1 als überlegen, ist dies ein erster Hinweis auf mögliche systematische Traitvarianz. Erst hieran kann sich die Frage anschließen, *wie* diese Traitvarianz modelliert werden kann.

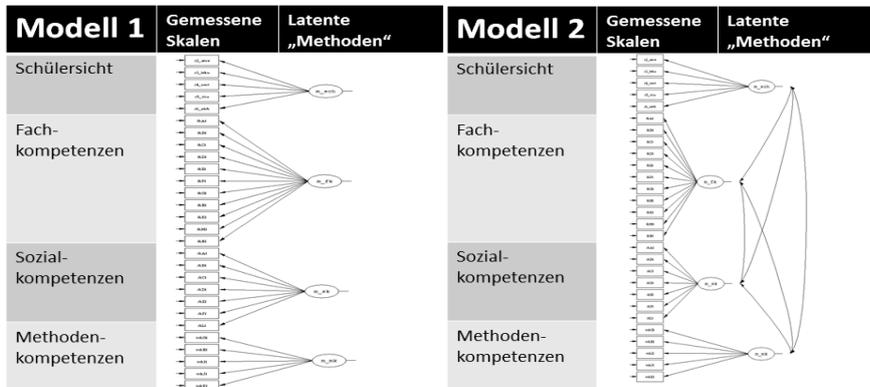


Abbildung 1: Strukturelle Darstellung der nur die Methodenfaktoren berücksichtigenden Modelle der konfirmatorischen Faktorenanalysen: links Modell 1 mit unkorrelierten Methodenfaktoren, rechts Modell 2 mit korrelierten Methodenfaktoren

2.4.2 Zweiter Schritt: Expertenbefragung zur Generierung von Strukturhypothesen die weiterführende Modellierung von Traitfaktoren

In typischen MTMM-Analysen zur Konvergenz der Perspektiven verschiedener Beobachtungsmodalitäten ist quasi im Ansatz selbst festgelegt, dass gleiche manifest gemessene Traits über verschiedene Methoden hinweg jeweils einen der latenten Traitfaktoren konstituieren. Da jedoch in dieser Studie zur Erfassung der „professionellen Sicht“ (gemeint ist Fremd- bzw. Selbstsicht) insgesamt 23⁷ Skalen zum Einsatz kommen, die wesentlich differenzierter (und überdies in drei Kompetenzbündel untergliedert) Aspekte des Lehrerhandelns in den Blick nehmen, als dies die sechs Skalen zur Erfassung der Schülersicht leisten, ist hier *nicht* offensichtlich, wo überhaupt Konvergenzen zwischen den Perspektiven zu

⁷ Diese setzen sich zusammen aus elf Skalen zu Fachkompetenzen sowie sieben Skalen zu Sozial- und fünf Skalen zu Methodenkompetenzen.

erwarten sind. Um dies vorab zu klären und damit Strukturhypothesen überhaupt erst zu generieren wurde eine Expertenbefragung durchgeführt, in der im Rahmen von Fortbildungs- und Informationsveranstaltungen Akteure aus verschiedenen Bereichen des Lehrerbildungs- und Schulwesens (Lehrende an Universitäten und Studienseminaren sowie praktisch tätige Lehrkräfte) gebeten wurden, im Einzelnen zu „deklinieren“, wo Ihrer Einschätzung nach inhaltliche Überlappungen bzw. Übereinstimmungen zwischen den Aspekten der „professionellen Sicht“ und jeweils den sechs Aspekten der Schülerwahrnehmung bestehen. Um das Vorgehen der Befragung zu verdeutlichen, ist in Abbildung 2 ein Ausschnitt aus dem Erhebungsbogen dargestellt. Die Einschätzungen der Experten wurden in Anschluss gemittelt. Fiel die mittlere Einschätzung der Überlappung in den Bereich der Antwortkategorien „stimmt genau“ oder „stimmt meistens“ (dies entspricht bei einer an der Schulnotenskala angelehnten Codierung einem numerischen Wert ≤ 2.5), so wurde davon ausgegangen, dass es sich um eine bedeutsame Überlappung handelt.

Die pädagogische Kompetenz.. (PÄD. K. 3)	..stimmt überein oder überlappt inhaltlich mit der Schülerwahrnehmung	stimmt genau	stimmt meistens	stimmt manchmal	stimmt oft nicht	stimmt meistens nicht	stimmt gar nicht
Bewältigung von Disziplinproblemen und Schülerrisiken	Autokratie	1	2	3	4	5	6
	Anregungsgehalt des Unterrichts	1	2	3	4	5	6
	Lernkultur	1	2	3	4	5	6
	Wertschätzung	1	2	3	4	5	6
	Klare Strukturvorgaben	1	2	3	4	5	6
	Einbindung der Schüler	1	2	3	4	5	6

Abbildung 2: Ausschnitt aus dem den Experten vorgelegten Zuordnungsbogen zur Generierung von Strukturhypothesen

2.4.3 Dritter Schritt: Formulierung von Modellen mit Traifaktoren auf der Basis der Ergebnisse der Expertenbefragung

Die Ergebnisse der Expertenbefragung können unmittelbar in Strukturhypothesen umgesetzt werden, indem in den entsprechenden konfirmatorischen Modellen dort, wo die Experten – anhand der oben beschriebenen Wahl eines Cut-Off-Wertes – im Mittel der Meinung sind, dass substanzielle inhaltliche Überlappun-

gen bestehen, Pfade frei geschätzt werden können und andererseits dort, wo dies nicht der Fall ist, Pfade restringiert bzw. nicht zugelassen werden. Eine notwendige Vorannahme bei diesem Vorgehen ist bezüglich der durch die Ergebnisse der Expertenbefragung nicht erfragten potenziellen Anzahl latenter Traitfaktoren zu treffen. Um zu erreichen, dass die Struktur des Schülerurteils mit ihren sechs inhaltlichen Aspekten auch in den konfirmatorischen Analysen möglichst unverzerrt abgebildet wird, wurde hier die Vorannahme getroffen, dass sechs latente Traitfaktoren⁸ zugelassen und modelliert werden, von denen jeweils ein Pfad zu einer der Schülerskalen verläuft.

Für die konfirmatorische Modellierung von MTMM-Matrizen gibt es eine Reihe verschiedener Spielarten, die sich darin unterscheiden, ob zwischen einerseits den Traitfaktoren und andererseits zwischen den Methodenfaktoren Interkorrelationen angenommen werden. Eine kompakte Übersicht über unterschiedlichen Zugänge der Modellierung findet sich bei Schermelleh-Engel und Schweizer (2007). Für die hier beabsichtigten Modellierungen empfehlen sich als Ausgangspunkt wenig restriktive Ansätze, da einerseits (in Anschluss an Modell 2) davon auszugehen ist, dass die Methodenfaktoren untereinander korreliert sind und andererseits angesichts der hohen Skaleninterkorrelationen im Schülerurteil (vgl. Tab. 1) damit zu rechnen ist, dass auch die latenten Traits, die sich an die manifesten Skalen des Schülerurteils anlehnen (s.o.) nicht voneinander unabhängig sind. Das Modell, das dies umsetzt, wird in der einschlägigen Terminologie als CTCM-Modell (*correlated trait – correlated method*) bezeichnet. Von Schermelleh-Engel und Schweizer (2007; vgl. auch Clausen, 2002) wird berichtet, dass derartige Modellierungsansätze mit der konfirmatorischen Faktorenanalyse aufgrund der großen Zahl zu schätzender Parameter oft mit Schätzproblemen behaftet sind, die sich in der mangelnden Konvergierbarkeit der Modelle und sog. „Heywood-Cases“⁹ niederschlagen. Von Eid (2000) wurde zur Vermeidung bzw. Minimierung derartiger Schätzprobleme ein Modell minimaler Annahmen vorgestellt, in dem auf der Methodenseite ein Faktor weniger modelliert wird, als tatsächlich Methoden eingesetzt werden, dieses wird als CTC(M-1)-Modell (*correlated trait – correlated method minus one*) bezeichnet. Bei der Interpretation dieses Modelltyps ist allerdings zu berücksichtigen, dass es sich nicht um symmetrische Modelle handelt, d. h. dass sich Parameterschätzungen im gesamten Modell in Abhängigkeit davon ändern, *welcher* der Methodenfaktoren nicht modelliert wird. Bezogen auf die Fragestellung dieses Beitrags besteht der dritte Modellie-

⁸ Diese theoretisch gerechtfertigte Annahme ließ sich später in der Modellierung aufgrund eines methodischen Problems bei der Expertenbefragung nicht aufrecht erhalten (vgl. Ergebnisteil).

⁹ Es handelt sich meist um unplausible Parameterschätzungen wie negative Fehlervarianzen und standardisierte Koeffizienten, die numerische Werte größer eins annehmen.

rungsschritt darin, die Modelle, in die die aus der Expertenbefragung gewonnenen Strukturhypothesen einfließen, zunächst anhand des CTCM-Ansatzes und bei etwa auftretenden Schätzproblemen anhand des CTC(M-1)-Ansatzes zu prüfen.

3. Ergebnisse

3.1 Ergebnisse des ersten Analyseschritts – sind die Methodenfaktoren korreliert, und wenn ja, wie hoch?

Gegenstand des ersten Analyseschritts ist die Klärung folgender aufeinander aufbauender Fragen: An erster Stelle soll geprüft werden, ob generell die Annahme gerechtfertigt ist, dass den vier verschiedenen Kompetenzaspekten (die drei Kompetenzbündel Fach-, Sozial- und Methodenkompetenzen der „professionellen“ Sicht sowie die Schülersicht) jeweils eine Varianzquelle im Sinne eines General- bzw. Methodenfaktors gemeinsam ist und ob weiterhin angenommen werden kann, dass diese vier Methodenfaktoren voneinander unabhängig sind. Diese Hypothese entspricht dem in Abbildung 1 dargestellten Modell 1. Gerechtfertigt wäre diese Annahme dann, wenn Modell 1 einen guten Modellfit aufweisen würde. Die zweite darüber hinausgehende Frage lautet, ob Modell 2 mit korrelierten Methodenfaktoren die Daten besser zu beschreiben vermag als Modell 1. Wenn dies zutrifft, ist zudem zu prüfen, wie groß die in den Modellen gefundenen Interkorrelationen zwischen den Methodenfaktoren sind.

In Tabelle 2 sind die Maße der Anpassungsgüte dargestellt, wobei die Interpretationsmuster für die beiden betrachteten Teilfragestellungen sehr ähnlich sind. In beiden Fragestellungen verfehlt Modell 1 (Annahme unabhängiger Methodenfaktoren) nach einigen der angelegten üblichen Beurteilungskriterien eine akzeptable Anpassungsgüte, während sich Modell 2 in der ersten Teilfragestellung durch einen guten, in der zweiten Teilfragestellung zumindest durch einen akzeptablen Modellfit auszeichnet. Im direkten Vergleich der Modelle stellt sich Modell 2 in beiden Teilfragestellungen sowohl im absoluten Vergleich per AIC als auch im statistischen Vergleich (obere Teilfragestellung: $\Delta X^2 = 309.3$; $\Delta df = 6$; $p < .001$; untere Teilfragestellung $\Delta X^2 = 483.1$; $\Delta df = 6$; $p < .001$) als überlegen gegenüber Modell 1 heraus. Dies ist als Hinweis auf systematische Anteile gemeinsamer Varianz über die Beobachtungsmodalitäten hinweg zu werten, die es rechtfertigen, die folgenden Analyseschritte zu unternehmen.

Tabelle 2: Anpassungsgüte von Modell 1 (unkorrelierte Methodenfaktoren) und Modell 2 (korrelierte Methodenfaktoren) für die beiden Teilfragestellungen zur gemeinsamen Modellierung von Fremdurteil und Schülersicht (oben) sowie von Selbsturteil und Schülersicht (unten)

Teilfragestellung (a): Gemeinsame Modellierung von Fremdurteil und Schülersicht (N = 177)							
	X²	df	RMSEA	NNFI	CFI	sRMR	AIC
Modell 1	1050.0	350	.11	.94	.94	.39	1162.0
Modell 2	740.7	344	.08	.97	.98	.05	864.7
Teilfragestellung (b): Gemeinsame Modellierung von Selbsturteil und Schülersicht (N = 321)							
	X²	df	RMSEA	NNFI	CFI	sRMR	AIC
Modell 1	1569.1	350	.10	.93	.93	.30	1681.1
Modell 2	1086.0	344	.08	.96	.97	.06	1210.0

Interessant ist an dieser Stelle aber auch der Blick darauf, welche absolute Höhe die Schätzungen der Interkorrelationen zwischen den Methodenfaktoren in Modell 2 erreichen. Diese sind in Tabelle 3 dargestellt. Zum einen ist hier zu entnehmen, dass die drei Kompetenzbündel der „professionellen“ Beurteilung (Fach-Sozial- und Methodenkompetenzen) sehr hoch korreliert sind. Hieran vermag nicht die bloße Tatsache zu überraschen, dass hier substanzielle Korrelationen bestehen. Angesichts der Höhe der Bezüge ist es allerdings legitim, die Frage zu stellen, ob denn „professionelle“ Beurteiler – und dies betrifft sowohl die in Tabelle 3 links dargestellten Fremdurteile als auch die rechts dargestellten Selbsturteile – überhaupt in der Lage sind zur kognitiven Trennung der drei Kompetenzbündel bei der Einschätzung von Lehrerverhalten. Interessanter sind indes die moderat hohen Korrelationen (.28 bis .36), die in der Teilfragestellung (a) Fremdurteil – Schülersicht zwischen dem Methodenfaktor Schülerurteil und den drei Methodenfaktoren des Fremdurteils bestehen. Diese deuten auf wesentliche Anteile gemeinsamer Varianz zwischen Schülern und Fremdbeobachtern hin. Für die Bezüge zwischen Selbsturteil und Schülersicht (b) gilt dies analog, angesichts der deutlich niedrigeren Interkorrelationen (.04 bis .20) allerdings nur in abgeschwächter Form.

Tabelle 3: Interkorrelationen der Methodenfaktoren in Modell 2 in den beiden Teilfragestellungen

Nr.	Methodenfaktor	Interkorrelationen in Teilfragestellung					
		(a) Fremdurteil – Schülersicht			(b) Selbsturteil – Schülersicht		
		1	2	3	1	2	3
1	Schülerurteil						
2	Fachkompetenzen	.28			.20		
3	Sozialkompetenzen	.36	.83		.15	.81	
4	Methodenkompetenzen	.35	.92	.91	.04	.83	.87

3.2 Ergebnisse des zweiten Analyseschritts – wo sehen Experten Überlappungen zwischen den „professionellen“ Kategorien und der Schülersicht?

Zur Generierung von Strukturhypothesen zur Konstituierung konfirmatorischer Modelle mit inhaltlichen Faktoren wurden insgesamt $N = 16$ Experten anhand des in Abschnitt 2.4.2 dargestellten Verfahrens befragt. Die Gruppe dieser Experten setzte sich in etwa zu gleichen Teilen aus Lehrenden in der universitären Lehrerbildung (sechs Personen), Ausbildern an Studienseminaren (vier Personen) und in der Schule tätigen Lehrkräften (vier Personen) zusammen. Drei der Experten verfügten über eine Schulerfahrung von bis zu zehn Jahren, sechs über eine Erfahrung von bis zu 20 Jahren und fünf über eine darüber hinausgehende Erfahrung. Die Ergebnisse der Befragung sind in Tabelle 4 dargestellt.

Sichtet man die in Tabelle 4 hervorgehobenen „bedeutsamen“ Überlappungen zwischen professionellen Kategorien der Kompetenzeinschätzung und den Aspekten der Schülersicht, so wird deutlich, dass es (a) wahrgenommene inhaltliche Übereinstimmungen durchaus in angemessener Zahl auftreten und (b) diese überwiegend eine hohe inhaltliche Plausibilität aufweisen. Insofern erscheint es gerechtfertigt, diese Übereinstimmungen in konfirmatorischen Modellen als Strukturhypothesen zur Konstituierung der latenten Traits zu verwenden. Ein bei der Expertenbefragung aufgetretenes Problem wird allerdings bezüglich des Aspekts „Autokratie“ der Schülersicht deutlich: Diesem Aspekt werden – inhaltlich hochgradig unplausibel – keine Überlappungen zugeschrieben. Eine mögliche Erklärung ist, dass hier die negative Konnotation des Merkmals „Autokratie“ zu Missverständnissen geführt hat, etwa in dem Sinne der fehlenden Einsicht, dass z. B. die *Abwesenheit* von „Autokratie“ aus der Sicht der Schüler sehr wohl etwas mit einer positiv gedachten Lehrer-Schüler-Beziehung zu tun haben mag. Da es aber post-hoc nicht möglich ist, Nicht-Übereinstimmungen, die auf derlei Missverständnisse zurückgehen, von Nicht-Übereinstimmungen, die tatsächliche Ferne

der Inhalte zum Ausdruck bringen zu trennen, wurde das Ziel verfehlt, spezifisch bezüglich des Aspekts „Autokratie“ Strukturhypothesen zu gewinnen. Es wird daher im sich anschließenden Analyseschritt auf die Modellierung eines so überschriebenen latenten Traits verzichtet.

Tabelle 4: Ergebnisse der Expertenbefragung (Mittelwerte) zu inhaltlichen Überlappungen zwischen „professionellen“ Beobachtungskategorien, die in Selbst- und Fremdeinschätzungen zum Einsatz kommen und Aspekten des Schülerurteils

	Schülerurteile* „prof.“ Kat.	Auto- kratie	Anre- gung	Lern- kultur	Wert- schät- zung	Struktur	Einbin- dung
Fachkompetenzen	Lehrer-Schüler-Bez.	4.31	2.93	2.27	1.13	2.94	1.63
	Beob. & Diag.	4.69	3.31	1.94	2.25	2.81	2.13
	Disziplin	3.44	2.50	2.00	1.56	2.06	2.19
	Förderung v. soz. Verhalten	4.31	4.06	2.63	1.37	3.44	1.44
	Lernstrategien	4.81	2.38	1.44	2.60	2.31	2.25
	Gestaltung und Methoden	4.56	1.50	1.38	3.06	1.63	2.06
	Leistungsmessung	4.19	3.94	2.50	2.19	3.06	3.56
	Medien	5.19	1.68	2.25	3.69	3.75	3.56
	Selbstorganisationskomp.	4.50	3.69	3.12	3.27	1.81	3.37
	Unterrichtsplanung	4.75	1.93	2.06	3.00	2.00	2.63
Unterrichtsdurchführung	4.38	2.00	2.13	2.44	1.75	2.06	
Sozialkompetenzen	Selbstständigkeit	4.38	2.69	2.31	1.92	2.46	1.92
	Kooperation	4.75	3.00	1.69	1.50	3.19	1.25
	Soz. Verantwortung	3.94	3.38	2.50	1.44	3.63	1.69
	Konfliktfähigkeit	2.69	3.19	2.75	2.19	2.56	2.25
	Kommunikation	4.19	3.00	2.88	2.38	2.88	1.75
	Führungsfähigkeit	2.94	3.75	3.69	2.00	1.63	2.69
	Situationsger. Auftreten	4.19	3.50	2.31	1.69	2.63	2.63
Methoden- kompetenzen	Reflexivität	4.88	4.63	2.50	2.50	3.38	2.38
	Analysefähigkeit	5.00	2.88	2.44	3.06	2.94	2.81
	Flexibilität	4.69	2.40	2.06	2.13	2.81	1.81
	Zielorientierung	3.50	3.19	2.88	3.44	1.63	3.13
	Arbeitstechnik	4.56	3.12	1.94	3.63	2.00	2.81

Die Einschätzungen der Experten wurden auf einer Skala von 1 – „stimmt genau“ bis 6 „stimmt gar nicht“ vorgenommen, Nach dem in Abschnitt 2.4.2 definierten Kriterium „bedeutsame“ Überlappungen sind hervorgehoben; * Für die vollständigen Benennungen der Aspekte des Schülerurteils vgl. Tabelle 1.

3.3 Ergebnisse des dritten Analyseschritts – Versuch der Modellierung latenter Traits

3.3.1 Modelle der konfirmatorischen Faktorenanalyse mit Methodenfaktoren und latenten Traits

Gemäß dem in Abschnitt 2.4.3 beschriebenen Vorgehen wurde zunächst versucht, getrennt für die beiden Teilfragestellungen (a) Konvergenz Fremdurteil – Schülersicht und (b) Konvergenz Selbsturteil – Schülersicht Modelle mit korrelierten Methoden und korrelierten Traits (CTCM) zu berechnen, wobei die in Tabelle 4 dargestellten Hervorhebungen in der Modelldefinition der Freisetzung von Pfaden zu den fünf latenten Traits¹⁰ entsprechen. Hierbei traten erhebliche Probleme auf: Während das CTCM-Modell zu Teilfragestellung (a) nicht konvergiert, ist das analoge Modell zu Teilfragestellung (b) mit sog. „Heywood-Cases“ (hier: Interkorrelationen zwischen den latenten Traits und standardisierte Faktorladungen, die den Wert 1 überschreiten) behaftet. Da diese als Befunde als unspezifische Hinweise auf eine mangelnde Passung zwischen Modell und Daten zu verstehen sind, kann keine inhaltliche Interpretation erfolgen.

Erwartungskonträr bleiben die Schätzprobleme allerdings in gleicher Weise auch dann bestehen, wenn die von Eid (2000) beschriebene CTC(M-1)-Modellierung¹¹ zur Anwendung gebracht wird. Obwohl diese Modellierungstechnik gerade mit dem Ziel entwickelt wurde, derlei Konvergenz- und Schätzprobleme zu reduzieren, führt sie bei der hier beabsichtigten Modellierung in beiden Teilfragestellungen nicht zu diesem Ziel. Als Erklärung hierfür kommt in Betracht, dass die Varianzanteile, die auf Methodeneffekte zurückgehen, gegenüber denjenigen, die „echte“ Traitvarianz ausdrücken, so dominant sind, dass bei Nicht-Modellierung eines der Methodenfaktoren ganz wesentliche Anteile gemeinsamer Varianz zwischen den Indikatoren dieses Bereichs nicht mehr darstellbar sind.

Interpretatorisch bleibt angesichts der gescheiterten Modellierungsversuche nur die allgemeine Schlussfolgerung, dass es mit der gewählten Herangehensweise und den für die Betrachtung des Lehrerhandelns in dieser Studie verwendeten Instrumenten *pauschal nicht möglich* war, Konvergenzen in der Wahrnehmung von (a) Fremdbeurteilern und Schülern sowie (b) dem Selbsturteil der angehenden Lehrkräfte und Schülern darzustellen.

¹⁰ Aufgrund der in Abschnitt 3.2 berichteten Probleme bei der Hypothesengewinnung zum Aspekt „Autokratie“ wurde auf die Modellierung eines entsprechenden latenten Traits verzichtet.

¹¹ Eid (2000) empfiehlt, die Auswahl des Methodenfaktors, der nicht modelliert wird, nach theoretischen Erwägungen vorzunehmen. Bei unserem Vorgehen wurde darauf verzichtet, den Methodenfaktor der Aspekte des Schülerurteils zu modellieren.

3.3.2 „Ursachenforschung“ zur mangelnden Konvergenz der Modelle

Eine Möglichkeit der Suche nach den Ursachen für die mangelnde Passung zwischen Daten und Modell auf deskriptiver Ebene ist die Betrachtung der Korrelationsmuster zwischen den manifesten Messungen. Soll das Modell in der Lage sein, die Daten hinreichend gut zu beschreiben, so ist es eine Grundanforderung, dass dort, wo die Strukturhypothesen dies vorsehen, auch empirisch die relativ betrachtet höchsten Zusammenhänge auftreten. Eine solche Gegenüberstellung von Modell und den Daten beider Teilfragestellungen ist in Tabelle 5 dargestellt.

Tabelle 5: Gegenüberstellung der Ergebnisse der Expertenbefragung (Strukturhypothesen) mit den relativ betrachtet höchsten Skaleninterkorrelationen in den beiden Teilfragestellungen (a) Konvergenz Fremdurteil – Schülersicht und (b) Konvergenz Selbsturteil – Schülersicht

Schülerurteile		Anre-gung	Lern-kultur	Wert-schät.	Struk-tur	Einbin-dung
„prof.“ Kat.						
Lehrer-Schüler-Bez.	Hypothese*					
	Korr. bei TF (a)**					
	Korr. bei TF (b)***					
Beob. & Diag.	Hypothese*					
	Korr. bei TF (a)**					
	Korr. bei TF (b)***					
Disziplin.	Hypothese*					
	Korr. bei TF (a)**					
	Korr. bei TF (b)***					
Förderung v. soz. Verhalten	Hypothese*					
	Korr. bei TF (a)**					
	Korr. bei TF (b)***					
Lernstrategien	Hypothese*					
	Korr. bei TF (a)**					
	Korr. bei TF (b)***					

* Hypothese: dunkelgrau hinterlegt sind diejenigen Felder, bei denen die Experten im Mittel eine „bedeutsame“ Übereinstimmung zwischen professioneller Kategorie und Aspekten der Schülersicht annehmen; ** Korr bei TF (a): dargestellt sind die relativ höchsten Korrelationen der Korrelationsmatrix von Fremdurteil und Schülersicht, hellgrau hinterlegt sind Zellen, in denen Korrelationen $\geq .25$ zu finden sind; *** Korr bei TF (b): dargestellt sind die relativ höchsten Korrelationen der Korrelationsmatrix von Selbsturteil und Schülersicht, grau hinterlegt sind Zellen, in denen Korrelationen $\geq .15$ zu finden sind.

Fortsetzung Tabelle 5

Schülerurteile		Anre- gung	Lern- kultur	Wert- schät.	Struk- tur	Einbin- dung
„prof.“	Kat.					
Gestaltung und Methoden	Hypothese*	■	■		■	■
	Korr. bei TF (a)**				■	
	Korr. bei TF (b)***	■	■	■	■	■
Leistungsmessung	Hypothese*		■	■		
	Korr. bei TF (a)**					
	Korr. bei TF (b)***	■	■			
Medien	Hypothese*	■	■			
	Korr. bei TF (a)**					
	Korr. bei TF (b)***	■				
Selbstorganisations- kompetenz	Hypothese*				■	■
	Korr. bei TF (a)**		■			
	Korr. bei TF (b)***					
Unterrichtsplanung	Hypothese*	■	■		■	
	Korr. bei TF (a)**					
	Korr. bei TF (b)***	■	■		■	
Unterrichts- durchführung	Hypothese*	■	■	■	■	■
	Korr. bei TF (a)**	■	■		■	
	Korr. bei TF (b)***	■			■	
Selbstständigkeit	Hypothese*		■	■	■	■
	Korr. bei TF (a)**				■	■
	Korr. bei TF (b)***					
Kooperation	Hypothese*		■	■		■
	Korr. bei TF (a)**	■	■	■	■	■
	Korr. bei TF (b)***					
Soz. Verantwortung	Hypothese*		■	■		■
	Korr. bei TF (a)**	■	■	■	■	■
	Korr. bei TF (b)***				■	
Konfliktfähigkeit	Hypothese*			■		■
	Korr. bei TF (a)**	■	■	■	■	■
	Korr. bei TF (b)***					

Fortsetzung Tabelle 5

Schülerurteile „prof.“ Kat.		Anre- gung	Lern- kultur	Wert- schät.	Struk- tur	Einbin- dung
Kommunikation	Hypothese*					
	Korr. bei TF (a)**					
	Korr. bei TF (b)***					
Führungsfähigkeit	Hypothese*					
	Korr. bei TF (a)**					
	Korr. bei TF (b)***					
Situationsgerechtes Auftreten	Hypothese*					
	Korr. bei TF (a)**					
	Korr. bei TF (b)***					
Reflexivität	Hypothese*					
	Korr. bei TF (a)**					
	Korr. bei TF (b)***					
Analysefähigkeit	Hypothese*					
	Korr. bei TF (a)**					
	Korr. bei TF (b)***					
Flexibilität	Hypothese*					
	Korr. bei TF (a)**					
	Korr. bei TF (b)***					
Zielorientierung	Hypothese*					
	Korr. bei TF (a)**					
	Korr. bei TF (b)***					
Arbeitstechnik	Hypothese*					
	Korr. bei TF (a)**					
	Korr. bei TF (b)***					

Da generell die in Teilfragestellung (b) auftretenden Korrelationen zwischen den Skalen des Selbsturteils und den Aspekten der Schülersicht geringer sind, als in Teilfragestellung (a), beziehen sich in Tabelle 5 die hellgrauen Hinterlegungen zur Kennzeichnung der *relativ* höchsten Korrelationen bei (a) auf $r \geq .25$, bei (b) dagegen auf $r \geq .15$. Aus dem visuellen Mustervergleich in Tabelle 5 gehen zwei Gegebenheiten hervor: Zum einen treten die empirisch relativ höchsten Korrelationen nicht systematisch dort auf, wo dies gemäß der Hypothesen anzunehmen ist, in Anlehnung an die MTMM-Terminologie wäre dies als Fehlen konvergenter Bezüge aufzufassen. Andererseits treten aber auch empirische Korrelationen dort auf, wo dies gemäß Hypothese nicht anzunehmen ist, dies kann als Fehlen dis-

kriminanter Validität verstanden werden. Insgesamt wird im Mustervergleich der Hypothesen mit den Daten deutlich, dass hier nur sehr geringe Übereinstimmungen existieren. Insofern ist einsehbar, dass diese Datenlage nicht zu den hypothetisch angenommenen Modellen der Konvergenz zwischen der Einschätzung des Lehrerhandelns Perspektiven Schüler- Fremdbeurteiler bzw. Schüler-Selbstbeurteiler zu führen vermag.

4. Bewertung der Ergebnisse und Ausblick

4.1 Einschätzung fehlender Konvergenz der Perspektiven *Experten, Studierende und Schüler*

Prämisse der vorliegenden Studie zur Einschätzung von Lehrerhandeln war die Notwendigkeit der Verwendung unterschiedlicher Beurteilungsskalen für Ausbilder und Studierende einerseits sowie Schüler andererseits. Der Literatur gibt darüber Auskunft, dass bezüglich der Unterrichtsqualität auch bei Verwendung gleicher Iteminhalte nur mäßige Übereinstimmungen zwischen den verschiedenen Beurteilungsperspektiven bestehen (vgl. Clausen, 2002, S. 233; Kunter und Baumert, 2006, S. 239f.; Wubbels, Brekelmans & Hooymayers, 1992, S. 49f.). Die Ergebnisse unserer Studie deuten stark darauf hin, dass sich dies sinngemäß auch auf den Gegenstand „Lehrerhandeln“ übertragen lässt.

Die Binnenstruktur der Landauer Daten legt den Schluss nahe dass, vergleichbar zu Clausen (2002), die Schüler zu einem generalisierenden Urteil neigen. Die hohen Interkorrelationen der Skalen deuten darauf hin, dass sie auf einen Generalfaktor „guter Lehrer“ rekurrieren. Dennoch: Auch wenn die Schülerperspektive sich nur durch annähernd eindimensionales „schwarz-weiß-Denken“ auszeichnet, wird sie dadurch ja nicht entwertet, im Gegenteil. Die praktische Relevanz der Wahrnehmung und Aufarbeitung der Schülerperspektive ist (vergl. Helmke et al., 2009) besonders im Schülerfeedbackprozess zu sehen.

Geringe Konvergenz und Probleme bei der Modellierung in der vorliegenden Studie führen zu der Frage: Welche der drei Perspektiven ist nun die „richtige“ bzw. repräsentiert den „wahren Wert“ der Qualität des Lehrerhandelns? Oder ist bereits die Frage falsch gestellt? Cronbach folgert (1995, S. 146, zit. nach Clausen, 2002) Konvergenz sei zwar ein vernünftiges Ideal, aber ein entferntes. Man solle sich lieber der Untersuchung der Bedeutung perspektivenspezifischer Varianzanteile zuwenden.¹² Diese scheinen tatsächlich verschiedene Gewichte zu besitzen, wie Kunter und Baumert (2006) dokumentieren.

¹²Vergleichbar kam auch die soziolinguistische Diskussion um „Defizit“ (Bernstein, 1958) oder „Differenz“ (Labov, 1969) schichtenspezifischen Sprachverhaltens bis heute nicht zum Abschluss und offenbart praktische Relevanz erst in konkreten Bildungsansätzen und deren Evaluation.

Wie Clausen (2002, S. 187f.) salomonisch formuliert, kann angenommen werden, dass jede der drei Perspektiven ihren eigenen Wahrheitsanspruch hat. Insofern ist auch die Folgerung nachvollziehbar, dass die Brauchbarkeit der „Informationsquellen“ vom Ziel der konkreten Anwendung abhängt. Die klassische „fehlerorientierte“ Herangehensweise, in der mangelnde Konvergenz und fehlende Validität gleichgesetzt werde, wird von Clausen in Frage gestellt. Dieser Einschätzung schließen sich vor dem Hintergrund der oben berichteten Ergebnisse auch die Autoren dieses Beitrags ungeteilt an.

Literaturlage und Realität der Lehrerausbildung in Deutschland stellen jedoch einen offenkundigen Widerspruch dar: Wubbels, Brekelmans und Hooymayers (1992, S. 49) fassen das Zutrauen in die übergeordnete Gültigkeit der Schülerurteile zusammen, „because of evidence, that student perceptions indeed are a good measure für the actual teacher behavior.“ Weiter folgern Wubbels, Brekelmans und Hooymayers (1992) auf der Basis mehrerer Studien, dass das Schülerurteil sich nicht nur von anderen Perspektiven unterscheidet, sondern auch besonders reliabel und valide sei. Im deutlichen Gegensatz zu dieser – möglicherweise einseitigen – Sichtweise steht jedoch der „Status Quo“ in den beiden Phasen der Lehrerausbildung, sind es doch zumindest im deutschen Ausbildungssystem einzig und allein die „geweihten“ Experten (weniger) in Universität und (mehr im) Vorbereitungsdienst, welche über die Qualität des Lehrerhandelns und damit praktisch über Stellenchancen, Lebensbiografien etc. bei Schülern und Lehrkräften befinden, familiäre Auswirkungen „schulmüder“ Kinder hier noch gar nicht mitgedacht.

Ein weiteres Ergebnis der Studie ist, dass bezüglich der Beziehungen der Urteilsgruppen Schüler-Ausbildungslehrkräfte-Experten untereinander die höchsten Korrelationen zwischen Ausbilder- (hier Klassen- oder Fachlehrer) und Schülerurteilen bestehen. Korrelationen zum Studierendenurteil fehlen oder sind vernachlässigungswürdig. Da die Schülerurteile auf Klassenebene aggregiert sind, könnte auch auf das Wirken einer Klassen-Beurteilungskultur geschlossen werden, welche sich von der der Studierendenperspektive deutlich unterscheidet.

4.2 Offene Fragen

Für die Verbesserung der Ausbildungsverläufe in der Lehrerbildung erscheint Wahrnehmung, Abgleich und die prozesshafte Entwicklung der Beurteilungskulturen verschiedener an Ausbildung beteiligter Gruppen unerlässlich. Gleich welcher der Perspektiven ein höherer Wahrheitsgehalt zugemessen wird, ist die mangelnde Konvergenz der Urteile in Ausbildungssituationen Anstoß zum gezielten Handeln in Richtung der Implementierung von „gemeinsamen“ Urteilskulturen in Schule, Hochschule und gemeinsamer Ausbildungssituation. Dabei ist zunächst die Frage wissenschaftlich noch nicht geklärt, ob es sinnvoller ist, mit der Ver-

schiedenartigkeit der Perspektiven weiter zu arbeiten oder auf die Implementierung intersubjektiv vereinbarter Kriterien abzuheben. Es muss auch geklärt werden, welche der drei Perspektiven Schüler, Ausbildungslehrkräfte und Experten gemessen an etablierten Außenkriterien¹³ welchen Wahrheitsgehalt tragen.

Konkret für REBHOLZ (2005-2010) bedeutet dies: Die durch acht Begleitveranstaltungen (pro Semester) für Studierende und Ausbildungslehrer gefüllte Implementierungszeit im Projekt REBHOLZ über 5 Jahre hinweg hat sich angesichts der Voraussetzungen in den Praktika an Schulen zwar bemerkbar gemacht, aber als noch nicht ausreichend erwiesen. Weder schienen die Ausbildungslehrkräfte genügend in Beurteilungskulturen involviert, noch schienen sie sich mit der Vereinbarung von Grundsätzen der Unterrichtsqualität in den Klassen ausreichend befasst zu haben. In den Gesprächen der Lehrerfortbildungen war auch bis 2010 erkennbar, dass die Befassung der Lehrkräfte mit dem „Orientierungsrahmen Schulqualität (ORS)“ in Rheinland-Pfalz noch wenig ausgeprägt war. Es ist also zu klären, in welchem Maße intersubjektiv vereinbarte Beurteilungskriterien in den Klassen selbst etabliert werden müssen und dass es wissenschaftlich begründete „Essentials“ dazu gibt, von deren allgemeiner Bekanntheit und Anerkennung vermutlich noch eingeschränkt auszugehen ist.

In der Lehrerbildung bisher wenig beachtet scheint das in der industriellen Fertigung dominierende Paradigma der *Kundenorientierung*. Übertragen auf das Schul- und Lehrerbildungswesen, würde dieses Muster ein Primat der Pflege der Schülerurteilkulturen nahelegen. Lehramtsstudierende sind aber auch „Kunden“ der Hochschul- und Seminarbildung. Sollte es zutreffen, dass die Bewertungskulturen der Hochschule, Studienseminare und Studierenden nicht konvergieren, wären auch klare bildungspolitische Entscheidungen notwendig. Die künstliche Trennung von Hochschulen und Seminarbildung sollte ja durch Beschluss der KMK und HRK (2008) verhindert werden, wird aber mancherorts in der Praxis tätig vorangetrieben. Wo die Trennung von Hochschul- und Seminar - Lehrbewertung, Ausbildungserfolg und Wirkungen in den Schulen noch „offenes oder heimliches Programm“ ist, muss diese schnellstens korrigiert werden. Indem geeignete Maßnahmen der Qualitätssicherung in der Lehrerbildung greifen (Bodensohn & Schneider, 2008c), können auch verschiedene Perspektiven unterrichtlicher Beurteilung praktische Relevanz entfalten.

¹³Bei Helmke u. a. (2009, S. 237f.) sind externe Kriterien aus PISA dargestellt: Student achievement, student satisfaction, teacher enthusiasm, quality of homework tasks.

Literatur

- Balzer, L., Bodensohn, R. & Frey, A. (2004). Diagnose und Rückmeldung von Handlungskompetenzen von Studierenden im Blockpraktikum – das Projekt VERBAL. *Journal für LehrerInnenbildung*, 4, 30-36.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *ZfE*, 9, 469-520.
- Bernstein, B. (1958). *Soziale Struktur, Sozialisation und Sprachverhalten*. Aufsätze 1958-1970. Amsterdam: Contact-Press.
- Bodensohn, R. & Schneider, C. (2006). Weiterentwicklung der Evaluationskultur Schulpraktischer Studien. In M. Rotermund (Hrsg.), *Schulpraktische Studien. Evaluationsergebnisse und neue Wege der Lehrerbildung* (S. 87-115). Leipzig: Leipziger Universitätsverlag.
- Bodensohn, R. & Schneider, C. (2008a). Lehrerhandeln aus Schülersicht – eine Analyse der Wahrnehmungsstruktur von Schülern bei der Beurteilung von angehenden Lehrkräften in Schulpraktika. *Lehrerbildung auf dem Prüfstand*, 1, 699-718.
- Bodensohn, R. & Schneider, C. (2008b). Was nützen Praktika? Evaluation der Block-Praktika im Lehramt – Erträge und offene Fragen nach sechs Jahren. *Empirische Pädagogik*, 22.
- Bodensohn, R. & Schneider, C. (2008c). Lehrerbildung auf den Prüfstand erfordert die Sicherstellung von Qualität. *Empirische Pädagogik*, 22, 430-459.
- Bodensohn, R. & Schneider, C. (2009). Vier Studien zu Blockpraktika als Ausgangsbasis für die Entwicklung standardbezogener Evaluation. In R. Bolle & M. Rotermund (Hrsg.), *Schulpraktische Studien in gestuften Studiengängen* (S. 206-238). Leipzig: Leipziger Universitätsverlag.
- Bodensohn, R. & Schneider, C. (2010). Präzisierung von Zielkriterien – Wahrgenommene Wichtigkeit von Standards des Lehrerhandelns im Schulalltag und deren Relevanz in den schulpraktischen Phasen des ersten Ausbildungsabschnitts. In J. Abel (Hrsg.), *Wirkt Lehrerbildung? Antworten aus der empirischen Forschung* (S. 235-242). Münster: Waxmann.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminat validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Clausen, M. (2002). *Unterrichtsqualität: eine Frage der Perspektive?: Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität*. Münster: Waxmann.
- Cronbach, L. J. (1995). Giving method variance its due. In D. W. Fiske, P. E. ShROUT & S. T. Fiske (Eds.), *Personality research, methods, and theory. A festschrift honoring Donald W. Fiske* (pp. 145-157). Hillsdale, N.J: L. Erlbaum Associates.

- Ditton, H. & Arnold, B. (2004). Schülerbefragung zum Fachunterricht – Feedback an Lehrkräfte. *Empirische Pädagogik*, 18, 115-139.
- Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, 65, 241-261.
- Fend, H. (2008). *Schule gestalten: Systemsteuerung, Schulentwicklung und Unterrichtsqualität*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Fleishman, E. A. (1953). The description of supervisory behavior. *Personnel Psychology*, 37, 1-6.
- Frey, A. (2006). Strukturierung und Methoden zur Erfassung von Kompetenz. *Bildung und Erziehung*, 59, 125-145.
- Frey, A. (2008). Kompetenzstrukturen von Studierenden in der ersten und zweiten Phase der Lehrerbildung.: Eine nationale und internationale Standortbestimmung. Landau: Verlag Empirische Pädagogik.
- Frey, A. & Balzer, L. (2003). Soziale und methodische Kompetenzen – der Beurteilungsbogen smk: Ein Messverfahren für die Diagnose von sozialen und methodischen Kompetenzen. *Empirische Pädagogik*, 17, 148-175.
- Frey, A., Jäger, R. S. & Renold, U. (2005). *Kompetenzdiagnostik. Theorien und Methoden zur Erfassung und Bewertung von beruflichen Kompetenzen*. Landau: Verl. Empirische Pädagogik.
- Gerstenmaier, J. (1975). Urteile von Schülern über Lehrer: Eine Analyse ausgewählter empirischer Untersuchungen. *Erziehungswissenschaft: Bd. 6*. Weinheim: Beltz.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen: Schüler als Quellen der Unterrichtsbeschreibung*. Münster: Waxmann.
- Haecker, H. & Werres, W. (1983). *Schule und Unterricht im Urteil der Schüler: Bericht einer Schülerbefragung in der Sekundarstufe I*. Frankfurt a. M.: P. Lang.
- Hascher, T. (2011). Forschung zur Wirksamkeit der Lehrerbildung. In E. Terhart (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 418-440). Münster: Waxmann.
- Hascher, T. & Moser, P. (2001). Betreute Praktika – Anforderungen an Praktikumslehrerinnen und -lehrer. *Beiträge zur Lehrerbildung*, 19, 217-231.
- Helmke, A. (2009). *Unterrichtsqualität*. Seelze-Velber: Klett-Kallmeyer.
- Helmke, A., Piskol, K., Pikowsky, B. & Wagner, W. (2009). Schüler als Experten von Unterricht. *Lernende Schule*, 46/47, 98-103.
- Kahl, T., Buchmann, M. & Witte, E. (1977). Ein Fragebogen zur Schülerwahrnehmung unterrichtlicher Lernsituationen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 9, 277-285.

- Katz, D. & Kahn, R. L. (1952). Some recent findings in human relations research, In E. Swanson, T. Newcombe & E. Hartley (Eds.), *Readings in social psychology* (pp. 114-121), NY: Holt, Reinhart and Winston.
- Kiers, H. A. L., Takane, Y. & Berge, J. M. F. (1996). The analysis of multitrait-multimethod matrices via constrained component analysis. *Psychometrika*, 61, (pp. 601-628).
- KMK (2004). Standards für die Lehrerbildung: Bildungswissenschaften: Beschluss der Kultusministerkonferenz vom 16.12.2004. Verfügbar unter: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf [1.11.2011].
- KMK/HRK (2008). Empfehlung der Kultusministerkonferenz und der Hochschulrektorenkonferenz zur Vergabe eines Masterabschlusses in der Lehrerbildung bei vorgesehener Einbeziehung von Leistungen des Vorbereitungsdiens-tes. Verfügbar unter: <http://www.hrk.de/bologna/de/download/dateien/KMK-HRK-Empfehlung08-07-08.pdf> [1.11.2011].
- Kunter, M. & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction, 9, 231-251.
- Labov, W. (1969). *The study of nonstandard English*. Washington DC.: National Council of Teachers of English.
- Landy, F. J & Lamiell-Landy, A. (1978) Dimensions of teacher behaviour. *Journal of Applied Psychology*, 63, 522-526.
- Lück, H. E.(1996). Kurt Lewin. Weinheim: Beltz.
- Martin, L. (1995). Klassenlehrer und Tutoren: Was Schüler von ihnen erwarten?: Ergebnisse von Schülerbefragungen 1982-1993. *Pädagogische Rundschau*, 49, 199-217.
- Mayer, J. K. & Nickolaus, R. (1998). Bewertung von Unterricht durch Schüler. Ein standardisierter Unterrichts-Beurteilungsbogen als Rückmeldungsinstrument für Lehrende. *Die berufsbildende Schule*, 50, 297-302.
- Meyer, H. (2003). Zehn Merkmale guten Unterrichts. *Pädagogik*, 10, 36-43.
- Oser, F. (2001). Standards: Kompetenzen von Lehrpersonen. In F. Oser & J. Oelkers (Hrsg.), *Die Wirksamkeit der Lehrerbildungssysteme* (S. 215-342). Zürich: Rüegger.
- Ostendorf, F., Angleitner, A. & Ruch, W. (1986). Die Multitrait-Multimethod-Analyse: Konvergente und diskriminante Validität der Personality-Research-Form. Göttingen: Verlag für Psychologie Hogrefe.
- Schermelleh-Engel, K. & Schweizer, K. (2007). Multitrait-Multimethod-Analysen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 325-341). Springer.

- Schneider, C. & Bodensohn, R. (2007). Fachkompetenzen in der Schulpraxis – Zur Bedeutung der Oserschen Standards professionellen Lehrerhandelns für den Berufsalltag und zur Kompetenzeinschätzung in Schulpraktischen Studien der ersten Phase der Lehrerbildung. In D. Flammeyer (Hrsg.), *Mehr Praxis in der Lehrerbildung - aber wie? Möglichkeiten zur Verbesserung und Evaluation* (S. 149-176). Leipzig: Leipziger Universitätsverlag.
- Schneider, C. & Bodensohn, R. (2008). Berufliche Handlungskompetenzen in der ersten Phase der Lehrerausbildung – Ergebnisse zur Entwicklung im Längsschnitt. In M. Rotermund (Hrsg.), *Bologna verändert die Lehrerbildung. Auswirkungen der Hochschulreform* (Schriftenreihe der Bundesarbeitsgemeinschaft Schulpraktische Studien, S. 32-63). Leipzig: Leipziger Universitätsverlag.
- Schwarz, B. & Prange, K. (1997). *Schlechte Lehrer/innen: Zu einem vernachlässigten Aspekt des Lehrberufs*. Beltz grüne Reihe. Weinheim: Beltz.
- Terhart, E. (2002). *Standards für die Lehrerbildung: Eine Expertise für die Kultusministerkonferenz*. Münster: Institut für Schulpädagogik und Allgemeine Didaktik, Universität Münster.
- Vroom, V. H. & Jago, A. G. (1991). *Flexible Führungsentscheidungen*. Stuttgart: Poeschel Verlag.
- Wagner, W. (2008). *Methodenprobleme bei der Analyse der Unterrichtswahrnehmung aus Schülersicht*. Universität Koblenz-Landau, Campus Landau (Pfalz). Verfügbar unter: [http://kola.opus.hbz-nrw.de/volltexte/2008/234/pdf/Diss_\(Publikation\).pdf](http://kola.opus.hbz-nrw.de/volltexte/2008/234/pdf/Diss_(Publikation).pdf) [1.11.2011].
- Wang, M. C., Haertel, G. D. & Walberg, H. J. (1993). Toward a knowledge base: Why, how, for whom? *Review of Educational Research*, 63, 365-376.
- Wubbels, T., Brekelmans, M. & Hooyman, H. P. (1992). Do teacher ideals distort the self-reports of their interpersonal behavior? *Teaching & Teacher Education*, 8, 47-58.

Anschrift der Autoren:

Dr. Christoph Schneider, Dipl. Psych., Institut für Bildung im Kindes- und Jugendalter, Universität Koblenz-Landau, Campus Landau,
E-Mail: schneider@uni-landau.de

Dr. Rainer Bodensohn, Dipl. Päd., Zentrum für Lehrerbildung, Universität Koblenz-Landau, Campus Landau,
E-Mail: bodensohn@uni-landau.de