

Konzeptualisierung, Erfassung und Entwicklung von Kompetenzen im Klassenmanagement

Dissertation

zur Erlangung eines
Doktors der Philosophie (Dr. phil.)

am Fachbereich Erziehungswissenschaft und Psychologie
der Freien Universität Berlin

vorgelegt von Dipl.-Psych. Valentina Piwowar
Berlin, 2013

Erstgutachterin: Prof. Dr. Felicitas Thiel, Freie Universität Berlin
Zweitgutachterin: Prof. Dr. Bettina Hannover, Freie Universität Berlin

Tag der Disputation: 27.11.2013

Diese Dissertation ist im Rahmen des Projekts „Kompetenzen des Klassenmanagements (KODEK). Entwicklung und Evaluation eines Fortbildungsprogramms für Lehrkräfte zum Klassenmanagement“ entstanden, welches im Rahmenprogramm zur Förderung der empirischen Bildungsforschung (Schwerpunktprogramm: Professionalisierung des Pädagogischen Personals in Bildungseinrichtungen) durch das Bundesministerium für Bildung und Forschung unter dem Förderkennzeichen 01JH0920 ermöglicht wurde.

Inhalt

Zusammenfassung	7
1. Einleitung und theoretischer Rahmen der Arbeit.....	12
1.1. Professionelle Kompetenz von Lehrkräften.....	16
1.2. Kompetenzen im Klassenmanagement	22
1.3. Erfassung von Kompetenzen im Klassenmanagement	37
1.4. Entwicklung von Kompetenzen im Klassenmanagement.....	46
1.5. Empirische Überprüfung von Kompetenzentwicklung.....	56
1.6. Zusammenfassung und Ableitung der Forschungsfragen	61
1.7. Literaturverzeichnis.....	65
2. Multidimensionale Erfassung von Kompetenzen im Klassenmanagement. Konstruktion und Validierung eines Beobachter- und eines Schülerfragebogens für die Sekundarstufe.	102
2.1. Einleitung	102
2.2. Methode.....	107
2.3. Ergebnisse	112
2.4. Diskussion	119
2.5. Literaturverzeichnis.....	124
2.6. Anhang	129
2.7. Autorinnenhinweis	133
3. Training Inservice Teachers' Competencies in Classroom Management. A Quasi-Experimental Study with Teachers of Secondary Schools.....	137
3.1. Introduction	137
3.2. Method	146
3.3. Results	151
3.4. Discussion	155
3.5. References	160
3.6. Acknowledgements	169

4. Evaluating Response Shift in Training Evaluation. Comparing the Retrospective Pre-Test with a Structural Equation Modeling Approach in a Classroom Management Training Program.	173
4.1. Introduction	173
4.2. Method	179
4.3. Results	183
4.4. Discussion	187
4.5. References	192
5. Gesamtdiskussion	198
5.1. Zusammenfassung der Ergebnisse	198
5.2. Grenzen und Stärken der Arbeit.....	213
5.3. Ausblick	214
5.4. Literaturverzeichnis.....	217
Danksagung	225
Erklärung	227
Lebenslauf.....	229

Zusammenfassung

In der vorliegenden Arbeit wurden vier zentrale Fragestellungen entwickelt und im Rahmen von drei empirischen Studien operationalisiert: 1) Wie kann Klassenmanagement theoretisch gefasst werden; 2) wie können Kompetenzen im Klassenmanagement mit Hilfe von Fragebogenverfahren empirisch erfasst werden; 3) wie können Kompetenzen systematisch entwickelt werden und 4) wie kann die Entwicklung von Kompetenzen empirisch nachgewiesen werden? Zur Beantwortung dieser Fragen wurde zunächst auf Basis einer systematischen Integration von Forschungsbefunden sowie verschiedener Forschungstraditionen ein taxonomisches Modell abgeleitet, das Strategien und Dimensionen von Klassenmanagement systematisiert. Die Taxonomie unterscheidet die drei Dimensionen *Management des Schülerverhaltens*, *Management der Instruktion* und *Management der Lehrer-Schüler-Beziehung*, welchen die Strategien *Regeln*, *Störungsintervention* und *Monitoring*; *Prozeduren*, *Gruppenmobilisierung*, *Zeitmanagement* und *Klarheit des Handlungsprogramms* sowie *Arbeitsbündnis* und *Umgang mit Konflikten unter Schülern* zugeordnet wurden. Insbesondere diese dritte Dimension wird im deutschsprachigen Raum in der Regel nicht als Teil des Klassenmanagements gefasst. Auf Grundlage dieses Modells wurden zwei Fragebogenverfahren als Fremdbesicht zur Erfassung von Kompetenzen (Schülerfragebogen, KODEK-S und Beobachterfragebogen, KODEK-B) und ein Selbstbesicht zur Erfassung von Wissen im Klassenmanagement (SEWIK) entwickelt. KODEK-S und KODEK-B wurden an einer Stichprobe mit 88 Lehrkräften und 1811 Schülerinnen und Schülern an Berliner und Brandenburger Oberschulen überprüft (Teilstudie 1); SEWIK kam im Rahmen von zwei Interventionsstudien ($n = 37$ und $n = 146$) zum Einsatz (Teilstudien 2 und 3). Die entwickelten Instrumente wiesen gute bis sehr gute Testeigenschaften auf. Es sind weitere Validitätsprüfungen und die Überarbeitung einzelner Items des KODEK-S notwendig. Darüber hinaus wurde ein Training für Lehrkräfte im Beruf und ein Training für Studierende des Lehramts konzipiert und implementiert. Beide Trainings wurden im Rahmen von Prä-Post Designs evaluiert. Dabei zeigten sich positive Effekte auf Wissen, Können und die Schülerbeteiligung bei den Lehrkräften (Teilstudie 2) sowie Mittelwertsveränderungen im Wissen bei den Studierenden (Teilstudie 3). Implikationen für die Weiterentwicklung der Trainings sowie die weitere empirische Prüfung in Bezug auf die Wirksamkeit der Trainings werden diskutiert. Schließlich wurde das Phänomen des Response Shift im Rahmen von Teilstudie 3 mit Hilfe verschiedener statistischer Verfahren analysiert. Es konnte gezeigt werden, dass ein Response Shift beim Einsatz von Selbstbesichtungsverfahren sehr wahrscheinlich ist, insbesondere im Rahmen von Interventionsstudien. Sowohl der Vergleich

8 ZUSAMMENFASSUNG

von konventioneller Prä- und Postmessung als auch der Vergleich von retrospektiver Prä- und Postmessung (indirekte Veränderungsmessung) ist vor dem Hintergrund des unzureichenden Invarianzniveaus problematisch. Es wird auf potentielle Validitätsprobleme beim Einsatz von konventioneller und retrospektiver Prätests im Rahmen von indirekter Veränderungsmessung hingewiesen. In einem abschließenden Ausblick werden auf Basis der Ergebnisse aus den drei Teilstudien mögliche Implikationen für Praxis und Forschung abgeleitet.

Schlagwörter: Klassenmanagement – Fragebogen – Training – Kompetenzentwicklung – Response Shift.

Abstract

The present thesis develops the following four research questions, which were addressed in three empirical studies: 1) How can we conceptualize classroom management competencies coherently, 2) how can we assess classroom management competencies via questionnaires adequately, 3) how can we develop classroom management competencies systematically and 4) how can we identify improvement in competencies on the basis of empirically sound designs and measures? In order to answer those questions, first a theoretical model was developed that integrates both empirical findings and diverse theoretical approaches. This model distinguishes between the dimensions of *behavior management*, *instructional management* and *management of teacher-student relationships*. Respective strategies that are assigned to those dimensions are as follows: *rules, dealing with disruptions and monitoring; procedures, group mobilization, time management and clarity of program of action* as well as *working alliance and conflicts among students*. The third dimension is often neglected in Germany. Three questionnaires, a student rating of competencies (KODEK-S), an observer rating of competencies (KODEK-B) and a self-rating of knowledge (SEWIK) were constructed on the basis of this model. The KODEK-S and the KODEK-B were implemented in a study of 88 teachers and 1811 students of secondary schools in Berlin and Brandenburg (Study 1); the SEWIK was implemented in two intervention studies (Study 2, n = 37 and Study 3, n = 146). The instruments revealed good to very good psychometric properties. Further research is needed with regard to the analysis of validity as well as a reformulation of single KODEK-S items. In addition, an inservice teacher training and a preservice teacher training on classroom management were developed and evaluated. Positive effects on inservice teachers' knowledge, competencies and student engagement (Study 2), and positive change in mean knowledge of preservice teachers (Study 3) were identified. Directions for further adaption of the two trainings and additional research on their effectiveness are given. Eventually, in Study 3 the phenomenon of response shift was analyzed using both a structural equation procedure (analysis of measurement invariance over time) and the retrospective pretest procedure (comparison of post measures' means). Results indicate that response shift is likely to occur when using self-ratings, particularly in intervention studies. Since both traditional and retrospective pretest measures did not show a sufficient level of measurement invariance for the post test measures, a pre-post comparison is questionable in either case. The results of the three studies are finally interpreted with regard to their implications for both practical and research issues.

Key words: classroom management – rating instruments – training program – development – response shift.

1

Einleitung und theoretischer Rahmen der Arbeit

1. Einleitung und theoretischer Rahmen der Arbeit

Der Beruf des Lehrers und der Lehrerin¹ hat in den letzten Jahren deutlich an medialer, bildungspolitischer und bildungswissenschaftlicher Aufmerksamkeit gewonnen, insbesondere in Folge der Veröffentlichungen der zum Teil ernüchternden Ergebnisse großer nationaler und internationaler Schulleistungsstudien (z. B. PISA, IGLU, DESI, TIMMS, VERA, vgl. Köller, 2007). Lehrkräfte gelten als „die wichtigsten Akteure im Bildungswesen“ (Baumert & Kunter, 2011, S. 29) und werden als zentrale Vermittler zwischen den Bedürfnissen der Schülerinnen und Schüler und den Anforderungen des Curriculums gesehen (Bransford, Darling-Hammond & LePage, 2005). In Hatties (2009) breit rezipierter Metaanalyse wies er dem „Faktor Lehrer“ mit einer Gesamteffektstärke von $d = .49$ den größten Einfluss auf den Lernzuwachs von Schülerinnen und Schülern zu (im Vergleich zu den Faktoren Schüler, Familie, Schule, Curricula und Unterricht). Die Qualifizierung von Lehrkräften gilt damit als entscheidender Beitrag zur Optimierung von Bildungsprozessen.

Das gegenwärtige bildungspolitische Interesse spiegelt sich in der systematischen Forschungsförderung von diversen Schwerpunktprogrammen zur Bildungsforschung wider: unter anderem in Programmen zur konzeptuellen Ausdifferenzierung und Modellierung einzelner Kompetenzbereiche (Klieme & Leutner, 2006) sowie zur systematischen Weiterentwicklung von Lehrerkompetenzen (u.a. TEDS-M, *Teacher Education and Development Study in Mathematics*, Blömeke, Kaiser & Lehmann, 2008). Als erste Ergebnisse wurden bereits eigens darauf fokussierte Themenhefte (Bauer, Gräsel & Prenzel, 2012; Fleischer, Leutner & Klieme, 2012; Prenzel, Gogolin & Krüger, 2007; Retelsdorf & Südkamp, 2012) sowie umfangreiche Forschungsberichte publiziert (z. B. Blömeke et al., 2008; Kunter, Baumert, Blum, Klusmann, Krauss & Neubrand, 2011).

Welche Aspekte des Lehrerhandelns als zentral erachtet werden (Stichwort: *evidence-based practice and knowledge base*, z. B. Gillies, 2009; Hattie, 2009), ergibt sich zum einen aus Forschungen zu Einflussfaktoren für die Leistungs- und Persönlichkeitsentwicklung von Schülerinnen und Schülern (insbesondere aus der Prozess-Produkt-Forschung), zum anderen aus einer Anforderungsanalyse der Tätigkeiten von Lehrkräften, welche vor allem durch die Expertenforschung vorangetrieben wurde (vgl. Bromme, 1992). Wie insbesondere für den

¹ In der vorliegenden Arbeit wird versucht, weitestgehend gendergerecht zu schreiben. Insbesondere bei zusammengesetzten Wörtern oder wenn Formulierungen durch gendergerechte Sprache unübersichtlich werden, wurde auf konsequentes Gendersplitting verzichtet. Grundsätzlich sind immer beide Geschlechter gemeint (Schülerinnen und Schüler, Lehrerinnen und Lehrer, Beobachterinnen und Beobachter, Teilnehmerinnen und Teilnehmer usw.).

Mathematikunterricht nachgewiesen werden konnte, ist das Fachwissen von Lehrkräften zwar eine notwendige Voraussetzung für Schülerlernen, jedoch nicht ausreichend, um dieses auch tatsächlich zu gewährleisten (Blömeke et al., 2008; Kunter & Baumert, 2011)– fachdidaktische und pädagogisch-psychologische Kompetenzen sind zentrale Vermittler dieser Lerninhalte. Bereits 1963 machte Carroll auf den Umstand aufmerksam, dass das Verhältnis von gewährter, benötigter und genutzter Lernzeit entscheidend für die Effektivität von Unterricht ist, also wesentlich an eine effektive Zeitnutzung gekoppelt ist. Dieses Verständnis wird in Angebots-Nutzungs-Modellen und Modellen der Unterrichtsqualität aufgegriffen, nach denen die aktive Lernzeit (*time on task*), also Zeit, in denen Schülerinnen und Schüler aktiv in kognitive Lernprozesse eingebunden sind, als Schlüsselvariable für erfolgreichen Unterricht gilt (vgl. Helmke, 2009).

Kompetenzen, die Lehrkräfte darin unterstützen, die aktive Lernzeit der Schülerinnen und Schüler zu optimieren, werden überwiegend dem Bereich des *Klassenmanagements* zugeschrieben (u.a. Bereitstellung einer sicheren Lernumgebung, Monitoring des Schülerverhaltens und Umgang mit Störverhalten). Um genau diesen Aspekt erfolgreichen² Lehrerhandelns soll es in der folgenden Arbeit gehen. Die Bedeutung von effektivem Klassenmanagement für erfolgreichen Unterricht ist in zahlreichen Studien und Metaanalysen belegt worden: Es zeigen sich nicht nur Zusammenhänge mit den Leistungen von Schülerinnen und Schülern sondern auch mit deren sozio-emotionaler und Persönlichkeitsentwicklung (Emmer, Evertson & Anderson, 1980; Kunter, Baumert & Köller, 2007; Oliver, Wehby & Reschly, 2011; Seidel & Shavelson, 2007; Wang, Haertel & Walberg 1993). Aber auch berufliche Selbstwirksamkeitserwartungen und das berufliche Belastungsempfinden von Lehrkräften (Brouwers & Tomic, 2000; Clunies-Ross, Little & Kienhuis, 2008; Friedman, 2006) sind mit Kompetenzen im Klassenmanagement assoziiert.

Klieme und Hartig (2007) formulieren drei zentrale Forschungs- und Anwendungsprobleme in Bezug auf die professionelle Kompetenz von Lehrkräften: Die Frage nach der Reichweite von Kompetenzen, ihrer Modellierung, Messung, Bewertung und Zertifizierung sowie ihrer Entwicklung und Förderung (vgl. auch Blomberg, Seidel & Prenzel, 2011; König, 2012). Diese Probleme werden in der vorliegenden Arbeit auf den Bereich Klassenmanagement übertragen und folgende Fragestellungen formuliert:

1. Was konstituiert Klassenmanagement und wodurch ist gelingendes bzw. erfolgreiches Klassenmanagement gekennzeichnet?

² Der Begriff des „erfolgreichen“ Lehrers kommt von Bromme (1993) und überträgt den Expertenbegriff in die Lehrerforschung.

2. Wie können Kompetenzen im Klassenmanagement empirisch fassbar gemacht werden? Liegen entsprechende Verfahren vor, die eine standardisierte, reliable, valide und auch ökonomische Erfassung von Klassenmanagement ermöglichen?
3. Wie können Kompetenzen im Klassenmanagement gezielt weiterentwickelt werden? Gibt es ausreichend bzw. angemessene Aus- und Weiterbildungsmöglichkeiten für Studierende bzw. Lehrkräfte in unterschiedlichen beruflichen Entwicklungsstadien?
4. Wie kann wissenschaftlich fundiert überprüft werden, ob Ergebnisse von Lehreraus- und -weiterbildung auf implementierte Interventionen zurückzuführen sind?

Damit sind die zentralen Fragestellungen der vorliegenden Arbeit umrissen, die im Rahmen von drei Studien untersucht wurden. Einleitend werden zunächst die theoretische Rahmung sowie relevante Forschungsbefunde skizziert (Kapitel 1). Dafür wird in Kapitel 1.1 auf das zu Grunde gelegte Begriffsverständnis der professionellen Kompetenz von Lehrkräften nach Weinert (2001b) eingegangen, welches maßgeblich die theoretische und empirische Operationalisierung sowie die Analyse der Entwicklung von Kompetenzen prägt. Durch die Darstellung von Kompetenzmodellen und Standards in der Lehrerbildung wird eine strukturelle Rahmung entworfen, an der sich die Analyse spezifischer Lehrerkompetenzen–hier Klassenmanagement–orientiert.

Obwohl Klassenmanagement sowohl als Professionsstandard als auch im Rahmen von kompetenzorientierten Ansätzen operationalisiert wird, gibt es in Deutschland nur wenig systematische Forschung zum diesem Teil der Lehrexpertise (Ophardt & Thiel, 2008; Schönbacher, 2008). Es liegen unterschiedliche und häufig eher reduzierte Auffassungen darüber vor, was Klassenmanagement konstituiert, so dass in Kapitel 1.2 eine konzeptuelle und Begriffsbestimmung in Anlehnung an Doyle (2006) sowie Ophardt und Thiel (2013) vorgenommen wird, in der Dimensionen und Strategien des Klassenmanagement differenziert werden.

Nachdem die zentralen Begriffe der Arbeit definiert sind, wird in Kapitel 1.3 die empirische Operationalisierung von Kompetenzen im Klassenmanagement thematisiert. Zwar gilt die systematische und theoriegeleitete Erfassung auf Basis psychometrisch belastbarer Instrumente als zentraler Bestandteil der Professionalisierung und empirischen Erforschung von Lehrerkompetenzen, derzeit existieren jedoch nur wenige etablierte Messverfahren, die Klassenmanagement und dessen zeitliche Entwicklung adäquat abbilden können (vgl. Blomberg et al., 2011; Brookhart, 2004). Klassenmanagement kann indirekt an seinen Wirkungen gemessen werden (Kriterien wie Leistungen, Einstellungen, Haltungen) oder auch über Stan-

dards idealtypisch bzw. normativ operationalisiert werden. In der vorliegenden Arbeit wird der Fokus auf die Befragung über standardisierte, geschlossene Ratingverfahren gelegt.

Trotz des Stellenwerts für schulisches Lernen und das gesundheitliche Wohlbefinden von Lehrkräften wurde die systematische Entwicklung von Klassenmanagement bislang kaum erforscht (Kunter & Baumert, 2011; Ophardt & Thiel, 2013). Dies ist vor dem Hintergrund, dass Lehrkräfte in verschiedenen Phasen der Berufsausbildung und -ausübung, aber auch Auszubildende und die Schulaufsicht im Bereich des Klassenmanagements den größten Entwicklungsbedarf sehen (Jones, 2006; Melnick & Meister, 2008; Pigge & Marso, 1997), durchaus erstaunlich. Dieser Bedarf spiegelt sich auch in der steigenden Anzahl von deutschsprachigen Handbüchern zum Klassenmanagement und verwandten Themen wider (Eichhorn, 2008; Keller, 2008; Lohmann, 2003; Nolting, 2009; Ophardt & Thiel, 2013; Rogers & Kühn, 2013; Toman, 2007). Aufgrund der spezifischen Beschaffenheit von Wissen und Können im Klassenmanagement sind speziell darauf abgestimmte, wissenschaftlich fundierte Lerngelegenheiten unabdingbar, die eine systematische, evidenzbasierte und langfristig wirksame Entwicklung von Kompetenzen im Klassenmanagement gewährleisten. In Kapitel 1.4 werden daher die verschiedenen Wissenstypen und ihre mentale Repräsentation, darauf abgestimmte adäquate Lerngelegenheiten sowie bereits existierende Programme zur Vermittlung von Klassenmanagement in der Lehreraus- und Lehrerfortbildung dargestellt.

Ob Trainings und spezielle Fortbildungskurse erfolgreich sind, ist zunächst in deren Konzipierung über die theoretische Fundierung der Inhalte und Lehr-Lern-Methoden angelegt. Darüber hinaus muss jedoch ein empirischer Nachweis gelingen, dass die erhofften Veränderungen eintreten (kurzfristiger Transfer sowie langfristige Persistenz) und dass diese Veränderungen auf die Intervention zurückzuführen sind (Gollwitzer & Jäger, 2009). Die Qualität der Wirksamkeitsevaluation ist damit ein entscheidender Prädiktor dafür, dass beobachtete Effekte die wahren Effekte zuverlässig abbilden. Je kleiner der Interpretationsspielraum der empirischen Befunde ist, desto eindeutiger können Effekte (oder fehlende Effekte) interpretiert werden. Dabei gibt es spezifische methodische Probleme, die noch nicht eindeutig gelöst worden sind. In Kapitel 1.5 werden daher abschließend angemessene methodologische und statistische Verfahren dargestellt, die die interne, externe und statistische Validität von Interventionsstudien beeinflussen.

In einer Zusammenfassung der nach aktuellem Stand offenen Forschungsfragen werden im letzten einleitenden Kapitel 1.6 die zentralen Fragestellungen der Arbeit herausgearbeitet sowie die einzelnen Studien vorgestellt, die diese Fragestellungen in konkrete Hypothesen übersetzen und empirisch prüfen. Im ersten Manuskript wird der Frage der Erfassung von

Klassenmanagement nachgegangen. Dabei werden zwei Instrumente vorgestellt, die die Erfassung in Form von standardisierten Fremdberichten ermöglichen (Kapitel 2). Im zweiten Manuskript wird ein entwickeltes Klassenmanagement-Training für erfahrene Lehrkräfte und dessen Evaluation vorgestellt (Kapitel 3). Im dritten Manuskript werden die messmethodischen Probleme von Prä-Post Designs eruiert, die auf Selbstauskünften beruhen (Kapitel 4). Abschließend erfolgt eine integrierende Diskussion auf Basis der gefundenen Ergebnisse und im Hinblick auf die formulierten Desiderata (Kapitel 5).

1.1. Professionelle Kompetenz von Lehrkräften

Klassenmanagement als Teil professioneller Kompetenz zu konzeptualisieren impliziert, eine Anlehnung an die Kompetenzforschung vorzunehmen. Innerhalb kompetenzorientierter Ansätze liegen verschiedenste Auffassungen von Kompetenz vor. Entsprechend werden unterschiedliche dimensionale, strukturelle und genetische Modelle von Lehrerkompetenz postuliert. Unter Rückgriff auf empirische Befunde und theoretische Konzepte verschiedener Forschungstraditionen (v.a. Professionsforschung, Kompetenzforschung, Expertenparadigma) wird im Folgenden der Kompetenzbegriff näher spezifiziert (Kapitel 1.1.1), Standards und Modelle für Lehrerkompetenzen dargestellt (Kapitel 1.1.2) sowie abschließend eine Verortung von Klassenmanagement in Standards und Modellen vorgenommen (Kapitel 1.1.3).

1.1.1. Begriffsbestimmung und theoretische Konzeptualisierung von professioneller Kompetenz

Der Kompetenzbegriff erfreut sich derzeit großer Popularität in den Erziehungs- und Bildungswissenschaften sowie in Soziologie und Psychologie, wodurch zum einen die theoretische Ausdifferenzierung aber auch die systematische Kompetenzmodellierung vorangetrieben werden konnten (Baumert & Kunter, 2006; Klieme & Hartig, 2007; Prenzel et al., 2007). Gleichsam wird die „inflationäre“ Verwendung und damit Unschärfe des Begriffs beklagt (Bodensohn, 2005; vgl. auch Erpenbeck & von Rosenstiel, 2007; Hartig, 2008; Schneider & Bodensohn, 2010; Weinert, 2001a). Was genau unter Kompetenz–genauer: professioneller Kompetenz von Lehrkräften (Blömeke, Felbrich & Müller, 2008b; Seifert, Hilligus & Schaper, 2009; Terhart, 2007)–gefasst und verstanden wird, variiert nicht nur national (Kaufhold, 2006; Kemna, 2012; Klieme & Hartig, 2007; Terhart, 2007; Weinert, 2001b), sondern auch international stark (Eraut, 1994; Weinert, 2001a).

Trotz der Anwendungsbreite des Begriffs konnte sich in der deutschsprachigen Forschung die Definition Weinerts (2001b) etablieren (vgl. Klieme & Hartig, 2007), wonach Kompetenzen definiert werden als

die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können. (S. 27)

Kompetenz wird demnach nicht ausschließlich als „kognitive Disposition“ verstanden (Klieme & Leutner, 2006), sondern breiter angelegt (vgl. Kaufhold, 2006) und als Handlungskompetenz konkretisiert (Weinert, 1999): „The concept of action competence systematically combines constructs of the cognitive and motivational approach related to the goals, demands and tasks of a particular action context (e.g., a profession)“ (S. 14). Das potentiell verfügbare Wissen und Können ist damit per definitionem an nonkognitive Bereitschaften und Fähigkeiten gebunden; die professionelle Handlungskompetenz äußert sich erst in der Bewältigung von konkreten Anforderungssituationen (Bromme, 1993, 1997; Klieme & Hartig, 2007; Weinert, 2001a) auf der Ebene der Performanz (vgl. Blömeke, 2007; Kemna, 2012; Klieme & Hartig, 2007; Maag Merki & Grob, 2005).³ Dies gilt insbesondere für fächerübergreifende, pädagogisch-psychologische Kompetenzen, die zum Teil implizit bleiben und somit erst in Handlungssituationen sichtbar werden.

In der kognitionspsychologisch geprägten Expertenforschung sind die terminologischen Äquivalente für die rein kognitiven Kompetenzen als deklaratives *Wissen* bzw. *formal knowledge* zu sehen (vgl. Baartman & Bruijn, 2011; Bransford et al., 2005; Fenstermacher, 1994; Renkl, 2006; Schraw, 2006; Weinert, 2001b). Die Handlungskompetenz betrifft demgegenüber die Anwendung von Wissen (Hartig, 2008) in einer realen Anforderungssituation und manifestiert sich im *Können* der Lehrkraft (Baer, 2009; Weinert, 2001b). Das Können von Lehrkräften (*practical knowledge*, Fenstermacher, 1994; Gholami & Husu, 2010; van Tartwijk, den Brok, Veldman & Wubbels, 2009) resultiert aus dem Zusammenspiel von kognitiven und nonkognitiven Bereitschaften und Fähigkeiten und rekuriert wiederum auf prozedurales Wissen (Schraw, 2006).

³ Während sich Kompetenz eher auf den Dispositionsbegriff bezieht (also professionelles Handeln unter Idealbedingungen, vgl. Bromme, 1993), manifestiert sich Handlungskompetenz performativ in der Bewältigung von realen Anforderungssituationen (Baartman, Bastiaens, Kirschner & van der Vleuten, 2007) und ist somit über die Zeit und über Situationen potentiell instabil (Maag Merki & Grob, 2005). Zwischen Kompetenz und Performanz besteht nicht ein deterministischer, sondern ein vermittelter, stochastischer Zusammenhang (ebd.). In der vorliegenden Arbeit ist nachfolgend mit *Kompetenz* also immer dieses handlungsorientierte Verständnis–im Sinne von *Handlungskompetenz*–gemeint.

Damit wird ein funktional-pragmatisches Konzept professioneller Handlungskompetenz zu Grunde gelegt, das durch vier Eigenschaften konstituiert ist (in Anlehnung an Kaufhold, 2006; Klieme & Hartig, 2007, vgl. auch Kemna, 2012):

- Kompetenz zeigt sich in der Bewältigung von (komplexen) Handlungssituationen.
- Kompetenz ist situations- und kontextabhängig. Ob eine Person kompetent handelt, hängt von motivationalen und volitionalen Aspekten der Nutzung und Umsetzung seiner Fähigkeiten in bestimmten Situationen und Kontexten ab (Hartig, 2008).
- Kompetenz ist subjektgebunden. Kompetenzen sind hypothetische Konstrukte in Form dispositioneller Voraussetzungen (Erpenbeck & von Rosenstiel, 2007), die in bestimmten Bereichen, Kontexten oder Anforderungssituationen realisiert werden können, aber nicht müssen (Terhart, 2007).
- Kompetenzen sind veränderbar und damit erlern- und vermittelbar (Bransford et al., 2005; Klieme & Leutner, 2006; Terhart, 2007; Weinert, 2001a).

1.1.2. Standards und Kompetenzstrukturmodelle

Strukturelle Einordnungen bzw. Systematisierungen zur Ausdifferenzierung professioneller Kompetenz von Lehrkräften lassen sich aktuell im Wesentlichen zwei Gruppen zuordnen: Einerseits wurden die in Standards unterteilten Modelle für die Lehrerbildung entwickelt sowie andererseits die auf den Arbeiten Shulmans (1986, 1987) basierenden dimensional (im Deutschen häufig: topologischen) Strukturmodelle.⁴

Standards

Die Formulierung von Professionsstandards für die Lehrerbildung wurde vor allem in den USA vorangetrieben (Cochran-Smith & Zeichner, 2005; Darling-Hammond, 2001; Darling-Hammond, Wise & Klein, 1999; Mitchell, Robinson, Plake & Knowles, 2001). Im deutschsprachigen Raum gaben zunächst Oser (1997a) bzw. Oelkers und Oser (2001) sowie darauf aufbauend Terhart (2002a) entscheidende Impulse zur Entwicklung und Implementation von nationalen Professionsstandards (vgl. Ettlín & Fuchs, 2007). „Standards sind optimal ausgeführte bzw. optimal beherrschte und in vielen Situationen anwendbare Fähigkeiten und Fertigkeiten, die nur von Professionellen Verwendung finden können, nicht aber von Laien oder Personen anderer Professionen“ (Oser, 1997a, S. 28). Standards benennen also, was

⁴ Eine in Deutschland in den Arbeits- und Berufswissenschaften verbreitete Modellierung orientiert sich an der Unterteilung in 4 Kompetenzklassen: Fachkompetenz, Methodenkompetenz, Personale Kompetenz, Soziale Kompetenz (und Varianten davon, vgl. Achtenhagen & Baethge, 2008; Edelmann & Tippelt, 2007; Erpenbeck & von Rosenstiel, 2007; Frey, 2005). Da diese Modelle nicht professionspezifisch sind und Klassenmanagement hier nicht eindeutig zu verorten ist, wird auf eine Darstellung der Unterscheidung in Kompetenzklassen verzichtet.

kompetentes Lehrerhandeln konstituiert. Sie dienen dabei gleichzeitig als Steuerungsinstrument zur Qualitätssicherung, indem sie Vorgaben und Ziele für das Leistungs- und Qualitätsniveau der Aus- und Weiterbildung schaffen (ebd.; König & Blömeke, 2009).

Inhaltlich entwickelte zunächst Oser (1997a) durch Gespräche mit Verantwortlichen der Lehrerbildung 88 Standards (*Kompetenzprofile*), die er in zwölf Standardgruppen zusammenfasste. Osers Standards wurden zum Teil kontrovers diskutiert, jedoch von Terhart (2002a) aufgegriffen, weiterentwickelt und neu strukturiert. Terhart ordnet den Bereichen *Unterrichtsfach*, *Fachdidaktik* und *Erziehungswissenschaften* je zehn und *schulpraktische Studien* fünf Standards zu, die am Ende der ersten Ausbildungsphase erreicht werden sollten. Diese werden in der zweiten Ausbildungsphase, dem Referendariat, um zehn weitere Standards ergänzt. Die definierten Ausbildungsstandards sollen sich nach Terhart an den Kompetenzebenen *Wissen*, *Reflexion*, *Kommunikation* und *Beurteilen* sowie in der zweiten Ausbildungsphase zusätzlich *Können* orientieren. Zusätzlich formulierte er Standards für die mit der Lehrerausbildung beauftragten Institutionen und das darauf gerichtete Steuerungssystem sowie qualitative Stufen bei der Kompetenzgenese (Terhart, 2002a; Sekretariat der Ständigen Konferenz der Kulturlminister der Länder in der Bundesrepublik Deutschland [KMK], 2004a, 2004b).

Die starke Rezeption von Standards wurde mit deren verbindlicher Einführung durch die KMK (2004a, 2004b; vgl. auch Klieme et al., 2003) eingeleitet. Mit Standards ist die Hoffnung verbunden, klare Ansprüche an die Prozesse und Ergebnisse von Schule und Lehrerbildung zu formulieren, die evaluierbar und standardisiert sind (Terhart, 2002a). Ohne die Erarbeitung von Standards lasse sich nicht feststellen, ob die Ziele von Lehrerbildung erreicht worden sind oder nicht (ebd.; Oelkers & Oser, 2001). Die Schwächen von Standards sehen Baumert & Kunter (2006) in der fehlenden Rahmung eines professionellen Handlungsmodells für Lehrkräfte (vgl. auch Cramer, 2010), das eine theoretisch begründete Auswahl der Kompetenzprofile und Standardgruppen erlaubt. Von Seiten der Praktiker wurde vor allem der normative Charakter von Standards kritisiert und damit Widerstände aber auch Ängste formuliert, unter anderem die daraus folgende Einschränkung der Selbstbestimmtheit einzelner Schulen (Terhart, 2002b).

Kompetenzstrukturmodelle

Um zu einer nicht bloß normativen Festlegung von Lehrerkompetenzen zu kommen, legte Shulman (1986, 1987) auf Basis von Fallstudien ein heuristisches Wissensmodell vor (*categories of knowledge base*, 1987, S. 8), das heute–mit entsprechenden Anpassungen–den Grundstein für die verbreitetsten Kompetenzstrukturmodelle bildet (Baumert & Kunter, 2006;

Oser, Duggeli & Heinzer, 2010). Shulman selbst verwies in unterschiedlichen Arbeiten auf unterschiedliche Wissensdomänen (1986, 1987, 2002); am häufigsten rezipiert wird die Unterscheidung von 1987 in:

- Fachwissen (*content knowledge*),
- Pädagogisches Wissen (*general pedagogical knowledge*),
- Wissen über das Fachcurriculum (*curricular knowledge*),
- Fachdidaktisches Wissen (*pedagogical content knowledge*),
- Psychologie des Lerners (*knowledge of learners and their characteristics*),
- Organisationswissen (*knowledge of educational context*) und
- Erziehungsphilosophisches, bildungstheoretisches und bildungshistorisches Wissen (*knowledge of educational's end, purposes values, their philosophical and historical grounds*).

Hinsichtlich der strukturbildenden Wissensdimensionen hat sich heute überwiegend die Unterteilung in Fachwissen, Fachdidaktisches Wissen und Psychologisch-pädagogisches Wissen durchgesetzt (vgl. Blömeke et al. 2008b; Bransford et al., 2005; Shulman, 2002), auch wenn diese zum Teil durch je „eigene“ Wissensdomänen ergänzt werden (z. B. Beratungswissen, Baumert & Kunter, 2011; Philosophie des Schulfachs, Bromme, 1997; Kontextwissen, Grossman, 1995). Diese Dreiteilung der Wissensdomänen entspricht der curricularen Strukturierung von Lehramtsstudiengängen in fachliche, fachdidaktische und erziehungswissenschaftliche Ausbildungsanteile (Seifert et al., 2009) sowie der Struktur der Ausbildungsstandards nach Terhart (2002a).

In Deutschland griff zunächst Bromme (1992, 1993, 1997) Shulmans Topologie auf und trieb deren Weiterentwicklung maßgeblich voran. In der Tradition der Expertenforschung nahm er eine Anforderungsanalyse des Lehrerberufs vor und kam so schließlich zu einer auch theoretisch begründeten Topologie des professionellen Wissens. Auf Grundlage des Weinert'schen Kompetenzbegriffs, den topologischen Wissensmodellen und den Arbeiten von Bromme entstanden in den letzten Jahren diverse, wenn auch sehr ähnliche Kompetenzstrukturmodelle, die neben den kognitiven Wissensdomänen immer auch die nonkognitiven persönlichen Voraussetzungen von Lehrkräften modellieren (vgl. Baumert & Kunter, 2011; Blömeke et al., 2008b; Seifert et al., 2009) und damit den in Kapitel 1.1.1 beschriebenen dispositionellen Charakter von Kompetenz theoretisch operationalisieren. Im Vorgriff auf das folgende Kapitel 1.1.3 werden die beiden verbreitetsten Kompetenzmodelle nachfolgend skizziert.

Das in COACTIV (*Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung mathematischer Kompetenz*) entwickelte generische Strukturmodell der professionellen Kompetenz von Lehrkräften ist hierarchisch strukturiert und unterscheidet zunächst verschiedene Kompetenzaspekte (*Professionswissen; Überzeugungen, Werthaltung und Ziele; Motivationale Orientierungen; Selbstregulation und Metakognitive Fähigkeiten*). Dem Professionswissen werden weiter ausdifferenzierende Kompetenzbereiche zugeordnet (*Fachwissen; Fachdidaktisches Wissen; Pädagogisch-psychologisches Wissen; Organisationswissen; Beratungswissen*).

Ähnlich strukturiert sind das in TEDS-M (Blömeke et al., 2008) bzw. MT21 (*Mathematics Teaching in the 21st Century*, Schmitt, Blömeke & Tatto, 2011) international etablierte Kompetenzstrukturmodell, in dem die beiden Dimensionen Professionswissen (*Fachwissen; Fachdidaktisches Wissen; Erziehungswissenschaftliches Wissen*) und Überzeugungen (*epistemologische Überzeugungen; Überzeugungen zum Lehrern und Lernen; Schul- und professionstheoretische Überzeugungen*) unterschieden werden (Blömeke et al., 2008b).

Während das vorrangige Ziel der Professionalisierungs- bzw. Qualifizierungsforschung eine Beschreibung kriterialer Normen oder Standards bzw. der Strukturen für das Erreichen dieser Standards ist, wird in Kompetenzstrukturmodellen eine stärker differenzielle Perspektive eingenommen (Baumert & Kunter, 2011), um die Prädiktoren und Konsequenzen individueller Kompetenzunterschiede untersuchen zu können. Kompetenzstrukturmodelle sind gegenüber Standards eher aus der Perspektive von Lehrkräften konzipiert, da sie die Voraussetzungen für die erfolgreiche Bewältigung des Lehrerberufs anforderungsbezogen und aus einer mehrdimensionalen Perspektive formulieren. Ihnen werden damit eher Anknüpfungspunkte für die Instrumententwicklung zugesprochen (Kemna, 2012); jedoch bleibt die inhaltliche Ausgestaltung konkreter Kompetenzen qua Modell zunächst offen. Spätestens für die Entwicklung von Instrumenten zur Kompetenzdiagnose sind auch hier normative Quantifizierungen unumgänglich.

1.1.3. Klassenmanagement als Standard und in Kompetenzstrukturmodellen

Klassenmanagement wird sowohl in den Standards (explizit, vgl. Terhart, 2002a bzw. implizit, vgl. Oser, 1997a, 2004) als auch in allen genannten Kompetenzstrukturmodellen als eigener Bereich professioneller Kompetenz definiert (vgl. Baumert & Kunter, 2006; Blömeke, Felbrich & Müller, 2008a; Bromme, 1997). Oser führte bereits 1997(a) die Standardgruppe *Bewältigung von Disziplinproblemen und Schülerrisiken* ein, die Themen des Klassenmanagements mit je einem Standard erfasst (u.a. Auslagerung von Konflikten; Umgang mit Dis-

ziplinproblemen; Umgang mit Mobbing; Konflikte unter Schülern). Nach Terhart (2002a) wird „Unterrichts-/Klassenführung“ im Rahmen der Ausbildungsstandards jedoch erst als Entwicklungsziel des Referendariats formuliert (S. 35).

Strukturell wird Klassenmanagement einheitlich fachunabhängig verortet und damit dem Bereich des erziehungswissenschaftlichen Wissens (Blömeke et al., 2008a; Seifert et al., 2009) bzw. dem pädagogischen (Bromme, 1997) oder dem pädagogisch-psychologischen Wissen (Baumert & Kunter, 2011) zugeordnet. Bereits an den unterschiedlichen Bezeichnungen deuten sich unterschiedliche Auffassungen darüber an, welche weiteren Aspekte dieser dritten Wissensdomäne zugeordnet werden. Auch innerhalb von Forschergruppen eines Modells werden hierunter verschiedene Aspekte subsumiert. In COACTIV werden überwiegend die Kompetenzfacetten *Klassenführung*, *Unterrichtsmethoden*, *Diagnostik*, *Lernprozesse* und *Heterogenität* differenziert (Baumert & Kunter, 2011; Voss & Kunter, 2011; Voss, Kunter & Baumert, 2011), in TEDS-M sind es die Bereiche *Strukturierung von Unterricht*, *Motivierung*, *Umgang mit Heterogenität*, *Klassenführung* und *Leistungsbeurteilung* (Blömeke et al., 2008a; König & Blömeke, 2009).

Obwohl diese Wissensdomäne im Gegensatz zu Fach- und Fachdidaktischem Wissen als weitestgehend fachunabhängig gilt, wird auch betont, dass das erziehungswissenschaftliche Wissen und Können eng an fachspezifische und fachdidaktische Aspekte gekoppelt ist (Voss & Kunter, 2011). Wie im nachfolgenden Kapitel dargestellt wird, erfüllt Klassenmanagement eine wichtige Stützfunktion für die Instruktion und ist stets auf diese bezogen. In Modellen der Unterrichtsqualität (insbesondere aus der Lehr-Lern-Forschung zum Fach Mathematik) wird Klassenmanagement daher neben konstruktiver Unterstützung (bzw. Schülerorientierung, Klieme, Schümer & Knoll, 2001 oder Motivierung, Ophardt & Thiel, 2013) und kognitiver Aktivierung als Basisdimension der Unterrichtsqualität konzipiert (Baumert & Kunter, 2006; Helmke, Helmke, Heyne, Hosenfeld, Schrader & Wagner, 2010; Kunter et al., 2005).

1.2. Kompetenzen im Klassenmanagement

Wie im vorangegangenen Kapitel herausgearbeitet wurde, stellt Klassenmanagement einen eigenständigen Bereich professioneller Handlungskompetenz dar. Aufgrund der uneinheitlichen begrifflichen sowie strukturellen Konzeptualisierung von Klassenmanagement wird zunächst eine terminologische Präzisierung und theoretische Verortung von Klassenmanagement vorgenommen sowie eine daraus abgeleitete und neu entwickelte Taxonomie vorgestellt, die die Klassenmanagement konstituierenden Handlungsstrategien systematisiert und in

Dimensionen ordnet (Kapitel 1.2.1). Die diesen Dimensionen zugeordneten Strategien des Klassenmanagements werden anschließend herausgearbeitet und ihre empirische Relevanz für schulisches und sozioemotionales Lernen der Schülerinnen und Schüler sowie für Merkmale der Lehrergesundheit und des Berufserfolgs nachgewiesen (Kapitel 1.2.2, 1.2.3 und 1.2.4).

1.2.1. Begriffsbestimmung und theoretische Konzeptualisierung von Klassenmanagement

Der Begriff Klassenmanagement–übersetzt aus dem englischen *classroom management*–findet in der angloamerikanischen Forschung bereits zu Beginn des 20. Jahrhunderts in der psychologischen und pädagogischen Fachliteratur Verwendung (Arnold, 1908; Bagley, 1907; Paulu, 1928), auch wenn die systematische Erforschung erst ein halbes Jahrhundert später verortet wird (vgl. Brophy, 2006). Entsprechend haben sich diverse theoretische wie methodische Paradigmen des Themas angenommen und verschiedene Auffassungen von Klassenmanagement entwickelt (für einen Überblick siehe u. a. Brophy, 2006; Cipani, 2008; Freiberg, 1983; Hardin, 2008; Manning & Bucher, 2007; Taubert, 2007). In der vorliegenden Arbeit wird eine Integration unterschiedlicher Forschungsansätze vorgenommen, die sich an den Arbeiten von Ophardt und Thiel (2007, 2008, 2013) sowie der im angloamerikanischen Sprachraum vertretenen, breiter angelegten Auffassung von Klassenmanagement orientieren (Evertson & Weinstein, 2006b).⁵

Als Ausgangspunkt dient die kontextualistisch-ökologische Sichtweise auf Klassenmanagement (Doyle, 1986, 2006; Gump, 1967; Kounin, 1970, 1976). Klassenmanagement bezieht sich hiernach nicht direkt auf Lernen, sondern wird funktional definiert als „how order is established and maintained in classroom environments“ (Doyle, 2006, S. 99). Die organisierende Grundeinheit dieser Lernumgebung ist die Aktivität bzw. *Aktivitätsstruktur* (Berliner, 1983; Doyle, 1984; Gump, 1967; Leinhardt, Weidman & Hammond, 1987), z. B. Stillarbeit, Unterrichtsgespräch oder Übergänge. Ordnung im Klassenzimmer herzustellen bedeutet dabei nicht, dass die Schülerinnen und Schüler sich ausschließlich passiv, ruhig und regelkonform zu verhalten haben, sondern, dass die Schülerinnen und Schüler–innerhalb akzeptierbarer Grenzen–dem von der Lehrkraft vorgegebenen Handlungsprogramm folgen (Doyle, 1986). In

⁵ Im Deutschen ist sowohl der Begriff *Klassenmanagement* (Eigenmann, 2009; Merkens, 2010; Rogers & Kühn, 2013; Ophardt & Thiel, 2013; Schönbächler, 2008) als auch *Klassenführung* (Baumert & Kunter, 2006; Dubs, 1995; Gold & Holodysnki, 2011; Haag & Streber, 2012; Helmke, 2007a; Mayr, 2006; Neuenschwander, 2006a; Seidel, 2009) verbreitet. Beide Begriffe–Führung und Management–sind mit einer eingeschränkten Auffassung konnotiert (vgl. Eigenmann, 2009; Schönbächler, 2008). Wesentlich für den hier verwendeten Begriff des Klassenmanagements ist der Umstand, dass Klassenführung häufiger für ein engeres Verständnis von classroom management aufgegriffen wurde (z. B. Baumert & Kunter, 2006; Helmke, Hosenfeld, Schrader & Wagener, 2002; Seidel, 2009). Dennoch werden beide Begriffe in der vorliegenden Arbeit synonym verwendet.

Abhängigkeit kontextualer Faktoren (der etablierten Aktivitätsstrukturen, der Voraussetzungen der Schülerinnen und Schüler oder den Überzeugungen der Lehrkraft) variiert damit, was als Ordnung verstanden und welches Schülerverhalten als angemessen oder erwünscht gilt. Als Ergebnis der Analyse der spezifischen Anforderungsstruktur des Klassenkontextes postulierte Doyle (1986) sechs Dimensionen, die unabhängig von der Unterrichtsform und den beteiligten Personen die Ursprünge für die Aufgaben des Klassenmanagements charakterisieren:

- *Multidimensionalität*: Im Unterricht treffen unterschiedliche Personen mit verschiedenen Zielen, Interessen und Fähigkeiten in unterschiedlichen Situationen aufeinander. Dies muss von der Lehrkraft berücksichtigt und orchestriert werden.
- *Simultaneität*: Viele Dinge geschehen gleichzeitig im Unterricht und erfordern von der Lehrkraft simultane Aufmerksamkeit.
- *Unmittelbarkeit*: Lehrerhandeln muss häufig unmittelbar erfolgen, ohne dass dieses proaktiv reflektiert werden kann.
- *Unvorhersagbarkeit*: Das Unterrichtsgeschehen kann häufig nicht antizipiert und damit nur begrenzt geplant werden. Auch bei sorgfältiger Vorbereitung bleibt Unterricht situationsabhängig.
- *Öffentlichkeit*: Interaktionen sind in der Regel nicht privat sondern öffentlich und können von den Anwesenden beobachtet werden; jeder wird Zeuge dessen, was der andere sagt und tut (insbesondere die Lehrkraft).
- *Geschichte*: Im Verlauf der Zeit entwickelt sich ein gemeinsames System an Erfahrungen, Routinen und Normen; Ereignisse sind damit nicht nur ein gegenwärtiges Phänomen, sie resultieren aus der Vergangenheit und wirken sich auf die Zukunft aus.

Diese anforderungsorientierte Sichtweise sowie die im ökologischen Paradigma eingeführten Begriffe *Handlungsprogramm* und des *Handlungsvektor* (Doyle, 1984, 2006; Merritt, 1982; Ophardt & Thiel, 2013)⁶ werden dem hier vorgenommenen Verständnis von Klassenmanagement als theoretische Bezugspunkte zu Grunde gelegt. Das Handlungsprogramm stellt den von der Lehrkraft intendierten Unterrichtsplan dar und ist stets sowohl bezüglich der sozialen Teilhabe als auch hinsichtlich der akademischen Aufgaben definiert (*social participation dimension* und *academic work dimension*, Doyle, 2006). Der Handlungsvektor ist der kontinuierliche Versuch der Lehrkraft, dieses Programm im Unterricht zu implementieren (ver-

⁶ Merritt (1982) führte den Begriff des Handlungsvektors ein, Doyle (1984, 2006) griff diesen auf, verwendete ihn jedoch nicht disjunkt sondern weitestgehend synonym mit dem Begriff des Handlungsprogramms. Die hier verwendete, eindeutige Unterscheidung wurde erst in den Arbeiten von Ophardt und Thiel vorgenommen (2007, 2008, 2013).

gleiche *Momentum* bei Kounin, 1970); er ist durch seine Richtung und Stärke charakterisiert (Doyle, 2006; Merritt, 1982). Störungen werden demnach als sekundäre bzw. konkurrierende Handlungsvektoren verstanden, die die Schülerinnen und Schüler vom primären Handlungsvektor–dem „eigentlichen“ Lerngegenstand–ablenken (Hastie, Sinelnikov & Brock, 2007; Merritt, 1982; Pollard, 1980). Diese konkurrierenden Handlungsvektoren können von Schülerinnen und Schülern sowie Lehrkräften (vgl. *Sprunghaftigkeit* bei Kounin, 1976, z. B. durch Unvermitteltheiten, Reizabhängigkeit oder Überproblematisierung) gleichermaßen initiiert werden. Strategien des Klassenmanagements dienen folglich der Stabilisierung und dem Schutz des primären Handlungsvektors.

Dennoch greift eine ausschließlich ökologische Sichtweise auf Klassenmanagement zu kurz. Nicht nur individuelle motivationale und kognitive Prozesse der Schülerinnen und Schüler bleiben unberücksichtigt (Doyle, 2006; Weinstein, 1991), sondern auch die potentielle Reichweite unterrichtlicher Ziele (u.a. die moralische und soziale Entwicklung der Schülerinnen und Schüler, vgl. Brophy, 1985, 1998; Freiberg, 1983). Dementsprechend wird Klassenmanagement nachfolgend umfassender verstanden als „the actions teachers take to create an environment that supports and facilitates both academic and social-emotional learning“ (Evertson & Weinstein, 2006a, S. 4, orientiert an Brophy, 1988, S. 2). Mit dieser Begriffsauffassung gehen drei Erweiterungen des Konzepts Klassenmanagement einher, die handlungstheoretische Aspekte, behavioristische Prinzipien und Ergebnisse klimaorientierter Ansätze mit einbeziehen.

Nach einer handlungstheoretischen Auffassung (*activity theory*) erfolgt Lehrerhandeln zielgerichtet, beruht auf strategischem Wissen (vgl. Kapitel 1.4.1) und wird als selbstreguliertes Handeln gesehen (Doyle, 2006; Evertson & Harris, 1992; Martin, 2004; Neuenschwander, 2006b). Erfolgreiches Klassenmanagement umfasst demnach sowohl Planungskompetenzen, als auch die „Kompetenz zum raschen situationsangemessenen Handeln“ (Bromme, 1992, S. 198), also adaptive Handlungskompetenz (Bischoff, Brühwiler & Baer, 2005; Weinert, 1996). Während Lehrkräfte in einem ökologischen Verständnis situational–im Kontext von Strukturen–regulieren, gilt die Lehrkraft hier als „Gestalterin sozialer Verhaltenserwartungen“ (Ophardt & Thiel, 2008, S. 268). Die Etablierung von Verhaltenserwartungen (z. B. durch Regeln und Prozeduren) dient der Komplexitätsreduktion, für die Lehrkraft sowie die Schülerinnen und Schüler gleichermaßen (Bromme, 1992): Im Unterricht kann situativ auf Schemata für sozial geteiltes Handeln zurückgegriffen werden (Ophardt & Thiel, 2008). Diese Schemata bzw. Interaktionsskripte sind dann erfolgreich, wenn sie zu Beginn des Schuljahres eingeübt und etabliert wurden (Emmer et al., 1980). Während das ökologische Paradigma ein Gegen-

entwurf zu der vorherrschenden behavioristischen Auffassung war (vgl. auch Freiberg, 1999), werden im handlungstheoretischen Paradigma auch Prinzipien des systematischen Verhaltensauf- und -abbaus durch Mechanismen operanter Konditionierung (positive und negative Verstärkung, positive und negative Bestrafung, Löschung) wieder aufgegriffen (Emmer, Evertson & Worsham, 2003; Landrum & Kauffmann, 2006). Jedoch rückte *gewünschtes* Verhalten in den Mittelpunkt (statt Kontrolle unerwünschten Verhaltens, vgl. Canter & Canter, 1992) und damit der Einsatz von Lob und positiven Verstärkern (vgl. Brophy, 1996) sowie die Forderung nach motivadäquater und bedürfnisorientierter, also individualisierter Verhaltensmodifikation (Evertson & Weinstein, 2006a; Marzano, Gaddy, Foseid, Foseid & Marzano, 2005; Weinstein, 2007).

Schließlich entwickelte sich in humanistischen und schülerzentrierten Forschungsansätzen eine Erweiterung der Konzeptualisierung von Klassenmanagement um die Ebene der Lehrer-Schüler-Beziehung (Battistich, Solomon, Watson & Schaps, 1997; Emmer & Stough, 2001; Freiberg, 1983; Gordon, 1974; Johnson & Johnson, 2006; Jones, 1996; Mayr, 2006; Nie & Lau, 2009; Pianta & Hamre, 2006; Schönbacher, 2008; Tausch & Tausch, 1998; Wubbels & Brekelmans, 2005). So äußert sich Martin (2004) unter Bezug auf Jones (1996) noch vor wenigen Jahren überrascht darüber, dass die Forschungsbefunde zum Stellenwert positiver Peer-Interaktionen, Kooperation, Konfliktmanagement und Selbstbestimmung—also auch Befunde der Klassen- und Schulklimaforschung—bislang vom Klassenmanagement ausgeklammert blieben, da diese in Bezug auf die Definition von Klassenmanagement als Herstellung einer adäquaten Lernumgebung von zentraler Bedeutung sind.

Es liegt eine Fülle von Modellen vor, um *Anforderungsbereiche*, häufiger jedoch *Handlungsstrategien* (*practices*) des Klassenmanagements zu systematisieren (einen Überblick geben u. a. Freiberg, 1983; Hardin, 2008; Manning & Bucher, 2007; Taubert, 2007). Keines dieser Modelle hat sich jedoch verbindlich durchgesetzt. Ophardt und Thiel (2007, 2008) fassen die Anforderungsbereiche des Klassenmanagements zu drei Bereichen zusammen: Einführung von Regeln und Einübung von Verhalten, Steuerung von Unterrichtsprozessen sowie Bearbeitung von Konflikten in der Schulklasse.

Ausgehend von diesen Anforderungsbereichen und mit Blick auf die angestrebte Instrumentenwicklung wird ein neues Modell zur Kategorisierung von relevanten Klassenmanagementstrategien entworfen, das auf die Klassifizierungen von Mayr, Eder & Fartacek (2005) zur Klassenführung (*Verhalten kontrollieren, Unterricht gestalten, Beziehung för-*

dern)⁷ sowie von Martin, Yin und Baldwin (1998) zu klassenmanagementbezogenen Lehrerüberzeugungen (*behavior management, instructional management, people management*, vgl. auch Martin & Yin, 1999; Martin, Yin & Mayall, 2007) zurückgreift und sich in ähnlicher Form auch bei anderen Autoren wiederfindet (Hastie et al., 2007; Wubbels, den Brok, Veldman & van Tartwijk, 2006). Dabei werden drei Dimensionen von Klassenmanagement unterschieden, die in den nachfolgenden Kapiteln genauer beschrieben werden (vgl. Abbildung 1):

1. *Management von Schülerverhalten*
2. *Management der Instruktion*
3. *Management der Lehrer-Schüler-Beziehung*

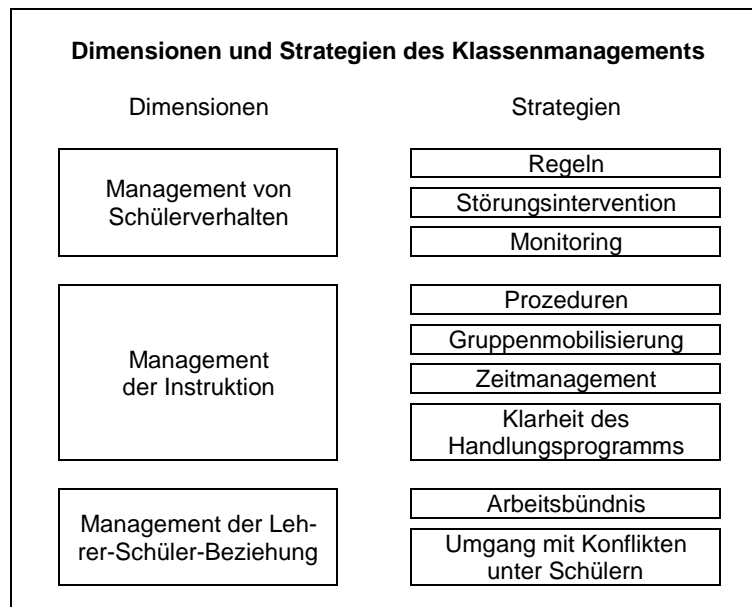


Abbildung 1. *Taxonomie zur Systematisierung von Strategien des Klassenmanagements*

Die Ausdifferenzierung der unter diesen Dimensionen gefassten Subdimensionen bzw. Strategien erfolgt unter Berücksichtigung der theoretisch und empirisch fundierten Relevanz, mit Hinblick auf die Operationalisierbarkeit und einen ähnlich hohen Abstraktionsgrad der

⁷ Insgesamt werden 21 Handlungsstrategien drei Bereichen der Klassenführung zugeordnet: 1) Verhaltenskontrolle (disziplinbezogene Strategien)–*klare Verhaltenserwartungen, Beschäftigung der Schüler, Kontrolle der Lernarbeit, Allgegenwärtigkeit, Verstärkung erwünschten Verhaltens, Eingreifen bei Störungen, Bestrafung unerwünschten Verhaltens*; 2) Unterrichtsgestaltung (pädagogisch-didaktische Strategien)–*Bedeutsame Lernziele, Strukturierter Unterricht, Klare Arbeitsanweisungen, Interessanter Unterricht, Fachkompetenz, Positive Erwartungshaltung, Verlässlichkeit*; 3) Beziehungsförderung (pädagogische Strategien)–*Wertschätzung, Verstehen, Authentizität, Humor, Gemeinschaftsförderung, Schülermitbestimmung, Kommunikation* (vgl. Mayr et al., 2005).

Konstrukte, sowie hinsichtlich der theoretischen Trennschärfe zwischen den Subdimensionen.⁸

1.2.2. Dimensionen des Klassenmanagements: Management des Schülerverhaltens

Der Umgang mit (vor allem störendem und nicht erwartungskonformem) Schülerverhalten galt lange Zeit als Kernelement des Klassenmanagements und wird in einer engen Auffassung auch heute noch als solches betrachtet (z. B. Cipani, 2008; vgl. Brophy, 2006; Weinstein, 1999). Unter der Prämisse, dass Schülerverhalten durch geeignete Techniken modifiziert und gelenkt werden kann, wurden zahlreiche Studien durchgeführt, die operante Konditionierungsprinzipien auf den Kontext Schule und Unterricht übertrugen. Dass erfolgreiches Lehrerhandeln sich nicht auf die—in der Regel reaktive—*Verhaltenskontrolle* beschränken kann, sondern auch präventiv und proaktiv auf das Verhalten der Schülerinnen und Schüler Einfluss nehmen sollte, zeigte erstmalig Kounin in seinen Beobachtungsstudien (1970, 1976) auf Grundlage videografierten Unterrichts. Er wies nach, dass nicht der Umgang mit Störungen per se zu hoher aktiver Lernzeit und Engagement der Schülerinnen und Schüler führte. Vielmehr eignete sich vorbeugendes Lehrerverhalten wie *Überlappung (overlapping)* und *Allgegenwärtigkeit (withitness)* zur Vorhersage des erwünschten Schülerverhaltens. In Klassen, in denen Lehrkräfte mehr reaktive als proaktive Strategien zur Steuerung des Schülerverhaltens einsetzten, wurde häufiger gestört und die Lehrkräfte berichteten ein höheres Stresserleben (Clunies-Ross et al., 2008, vgl. auch Admiraal, Korthagen & Wubbels, 2000; Emmer et al., 1980; Friedman, 2006; Little & Akin-Little, 2008). Insbesondere die Formulierung von sozialen und akademischen Verhaltenserwartungen sowie der gezielte Aufbau von Selbstregulationskompetenzen und *self-guidance* (Brophy, 1985) können proaktiv auf Schülerverhalten Einfluss nehmen (Evertson & Harris, 1992; Hardin, 2008; Martin, 2004). Dem Management von Schülerverhalten werden demnach die Strategien *Regeln*, *Störungsintervention* und *Monitoring* zugeordnet.

Regeln

Eine Regel identifiziert generelle Erwartungen und Standards, die durch konkrete Verhaltensanforderungen beschrieben werden (Emmer et al., 2003). Regeln dienen nicht der Disziplinierung sondern der vorbeugenden Stabilisierung des Handlungsvektors; Regelsysteme

⁸ Dabei führt nicht die additive bzw. isolierte Beherrschung dieser Strategien zu erfolgreichem Klassenmanagement, sondern deren kontext- und situationsadäquate Kombination, Reflexion und Modifikation bei unerwarteten Ereignissen (Berliner, 1983; Martin & Yin, 1999; Westermann, 1991). Es werden daher auch verschiedene „Muster“ (Mayr, 2006) und „Typologien“ (Neuenschwander, 2006b) von Klassenmanagement diskutiert (vgl. auch Weinert, Schrader & Helmke, 1990).

gestalten die Lernumgebung, die das Lernen der Schülerinnen und Schüler organisiert (Brophy, 1998). Ein wesentlicher Bestandteil der Regelarbeit ist damit die systematische Auswahl und Etablierung von Regeln bei Übernahme einer neuen Klasse bzw. zu Beginn der Schuljahres (Emmer et al., 1980; Evertson & Emmer, 1982). Wenn auch nur wenig systematische Forschung zum Einsatz von Regeln vorliegt (Doyle, 2006), so hat sich doch ein breiter Konsens über die optimale Beschaffenheit von Regeln etabliert (Emmer et al., 2003; Little & Akin-Little, 2008; Malone & Tietjens, 2000; McGinnis, Frederick & Edwards, 1995; Rademacher, Callahan & Pederson-Seelye, 1998; Schönbacher, 2006). Es sollten altersangemessene Verhaltenserwartungen expliziert und positives Verhalten eingeübt werden. Regeln, die gemeinsam mit den Schülerinnen und Schülern erarbeitet werden und mit allgemeinen Normen und Schulregeln im Einklang stehen, sind effektiver als Regeln die diese Eigenschaften nicht aufweisen (Emmer et al., 1980; Evertson & Emmer, 1982; Malone & Tietjens, 2000). Dabei sollte eine gezielte Auswahl von Regeln für verschiedene Verhaltensbereiche getroffen werden (u.a. sozialer Umgang, Lernen, Kommunikation im Unterricht, Konfliktlösung), um das kontingente Durchsetzen von Regeln durch die Lehrkraft–auch dauerhaft–zu gewährleisten (Little & Akin-Little, 2008; McGinnis et al., 1995; Rademacher et al., 1998; Schönbacher, 2006; van Tartwijk et al., 2009). Im Verlaufe des Schuljahres werden die eingeführten Regeln durch die stete Präsenz der Lehrkraft und die Sanktionierung bei Regelverstoß kontinuierlich gesichert und gefestigt (Emmer et al., 1980; Evertson & Emmer, 1982). Klare und von den Schülerinnen und Schülern akzeptierte Regeln, sowie die konsequente Sanktionierung bei Regelverstößen fördern die internale Kontrolle der Schülerinnen und Schüler und regen selbstregulatives Verhalten an (Brophy, 1985; Landrum & Kaufmann, 2006); die individuell wahrgenommene Regelklarheit war in einer querschnittlichen Studie mit fast 400 Siebt- und Achtklässlern mit mehr Interesse der Schülerinnen und Schüler assoziiert (Kunter et al., 2007).

Störungsintervention

Störungen sind keine isolierten Entitäten, sondern Teil des ökologischen Systems (Doyle, 1986). Demnach ist die Wirksamkeit von Interventionen nicht nur von der (temporären) Beendigung des störenden Verhaltens abhängig, sondern auch vom Zeitpunkt und der Angemessenheit der Intervention (Kounin, 1970) sowie der gleichzeitigen Aufrechterhaltung des etablierten Handlungsvektors (Doyle, 1986; Ophardt & Thiel, 2008). Störungsintervention bezeichnet also das situationsangemessene, adaptive Reagieren auf störendes oder unangemessenes Schülerverhalten (Brophy, 1988, 1996). Hierbei steht Lehrkräften ein breites Repertoire von indirekten und direkten Strategien zur Verfügung (Larrivee, 2009; Marzano,

2003; Scott & Quiroz, 1997), die stufenweise implementiert werden sollten (vgl. Stufenmodell nach Fink et. al, 2007, zitiert nach Ophardt & Thiel, 2013, S. 95): Stärkung und Schutz des primären Handlungsvektors, Schwächung und schließlich Beendigung des konkurrierenden Handlungsvektors. Die Stärkung des primären Handlungsvektors erfolgt mit Hilfe instruktionaler Management-Strategien (vgl. Kapitel 1.2.2) sowie durch gezieltes Ignorieren unerwünschten Verhaltens und die Unterstützung der Selbstkontrolle der Schülerinnen und Schüler (Leflot, van Lier, Onghena & Colpin, 2010; Thiel, Richter & Ophardt, 2012). Zum Schutz des primären Handlungsvektors setzen Lehrkräfte zusätzlich indirekte und minimale Interventionen ein (z. B. zeigen von Empathie, Entzug von Aufmerksamkeit, bewusster Augenkontakt, van Tartwijk et al. 2009; Weinstein, 2007; Woolfolk & Brooks, 1985). Erst wenn diese Maßnahmen nicht ausreichen, wird der primäre Handlungsvektor für kurze Zeit durch eine direkte Intervention bzw. Zurechtweisung unterbrochen. Hierbei ist ein unauffälliges und rasches Intervenieren zentral, sodass der Unterrichtsfluss erhalten bleibt, bevor Störungen sich ausbreiten oder eskalieren können (Doyle, 2006; Evertson & Emmer, 1982; Witt, van der Heyden & Gilbertson, 2004). Zurechtweisungen sollten nicht nur der Störung selbst, sondern auch situationalen und Schülermerkmalen (z. B. Interessen, motivationale Orientierungen, kultureller Hintergrund) angepasst sein (LePage, Darling-Hammond, Akar, Gutierrez, Jenkins-Gunn & Rosebrock, 2005; Martin, 2004; Witt et al., 2004). Nur als Reaktion auf wiederholtes Störverhalten werden Sanktionen implementiert; potentielle Konflikte müssen gegebenenfalls ausgelagert werden (Ophardt & Thiel, 2008; Witt et al., 2004). Konstruktives Durchsetzungsvermögen und empathische Gesprächsführung erhöhen den Erfolg der Durchsetzung von Sanktionen (Emmer et al., 2003; Ophardt & Thiel, 2013) und sind besonders dann wirksam, wenn sie motivationsadäquat sind, dem Schüler bzw. der Schülerin Entscheidungsmöglichkeiten eröffnen und konsequent implementiert werden (vgl. Brophy, 1985; Emmer et al., 1980). Außerdem sollten alternative Verhaltensweisen und ein alternatives Verhältnis zur Lehrkraft angeboten sowie ein günstiges Attributionsmuster (eigenes Verhalten ist kontrollierbar und variabel) gefördert werden (Brophy, 1985; Witt et al., 2004). Dies gelingt vor allem durch eine aufgabenorientierte Unterrichtsführung, die die Wichtigkeit und den Wert von Unterricht und Lernen signalisiert (vgl. Doyle, 1983; Doyle & Carter, 1984; Emmer & Gerwels, 2006).

Monitoring

Angelehnt an die Konzepte der Allgegenwärtigkeit und Überlappung (Kounin, 1970) hat sich heute der Begriff des Monitorings etabliert (Clausen, 2002; Doyle, 1984). Monitoring

beschreibt die Fähigkeit von Lehrkräften, die Aktivitäten jedes einzelnen Schülers und jeder einzelnen Schülerin aber auch der Gruppe als Ganzes im Blick zu behalten und gleichzeitig das Handlungsprogramm zu etablieren. Dies passiert durch „Abscannen“ des Raumes (Doyle, 1984), die gezielte Wahl eines geeigneten Standortes im Klassenzimmer (vgl. Thiel et al., 2012) sowie das Demonstrieren situationaler Bewusstheit (*situational awareness*, Doyle, 2006; Endsley, 1995), z. B. durch Vermeidung von Objekt- und Zeitfehlern bei Zurechtweisungen. Monitoring betrifft damit die Simultaneität des Lehrerhandelns und hat eine diagnostische Funktion. Berliner zeigte, dass zwischen Experten- und Novizen-Lehrkräften deutliche Unterschiede in der kategorialen Wahrnehmung von Unterrichtsereignissen vorliegen: Während Experten über Konzepte typischer Unterrichtsereignisse verfügen und einen ganzheitlichen Begriff von der Klasse haben, sehen Anfänger eher isolierte Ereignisse und einzelne Schülerinnen und Schüler. Diese Unterschiede konnten auch in einem experimentellen Laborsetting nachgewiesen werden (Copeland, 1987). Die Relevanz situationaler Bewusstheit wies bereits Kounin (1970) nach: Allgegenwärtige Lehrkräfte begingen seltener Zeitfehler und identifizierten frühzeitig unbeschäftigte, abgelenkte oder unaufmerksame Schülerinnen und Schüler. Sie konnten rechtzeitig und ohne Handlungsdruck entscheiden, ob und wie sie darauf reagieren (vgl. auch Doyle, 2006; Thiel et al., 2012). Demnach bezieht sich Monitoring nicht ausschließlich auf die Verhaltenssteuerung von Schülerinnen und Schülern, sondern dient primär dem Einholen von Informationen darüber, ob diese dem Handlungsvektor folgen bzw. folgen können–hat also in der Folge auch ein Funktion für die nachfolgende Dimension von Klassenmanagement.

1.2.3. Dimensionen des Klassenmanagements: Management der Instruktion

Klassenmanagement zielt neben dem Verhaltensmanagement vor allem auf die Aktivitätssteuerung und Orchestrierung des Unterrichts ab (Bromme, 1997; Brophy, 2006; vgl. Momentum und Schwung bei Kounin, 1976); es ist eng mit der Entwicklung des Stoffes verknüpft und wird nicht von der Lehrkraft aus gedacht, sondern bezieht sich auf das Lernen des Schülerinnen und Schüler (Ophardt & Thiel, 2008; Weinstein, 1999). Klassenmanagement hat demnach eine unmittelbare Stützfunktion für die Instruktion (Doyle, 2006; Evertson & Harris, 1992; Leinhardt et al., 1987; Mayr, 2006; Ophardt & Thiel, 2013). Der Einsatz instruktionaler Klassenmanagementstrategien dient dazu, relativ kontinuierlich akademische Aufgaben (Lerngelegenheiten) zur Verfügung zu stellen, die von den Schülerinnen und Schülern auch tatsächlich als aktive Lernzeit genutzt werden (Brophy, 1998; Doyle & Carter, 1984; Haag & Streber, 2012; Helmke, 2007a; Weinstein, 2007). Im Sinne eines Angebots-Nutzungs-Modells

(Helmke, 2009) können ein effizientes *Zeitmanagement* und ein *klares Handlungsprogramm* die tatsächliche Instruktionszeit–das Angebot im Unterricht–positiv beeinflussen. Gut eingeschliffene und einfach abrufbare *Prozeduren* (Arlin, 1979; Kounin & Gump, 1974; Leinhardt et al., 1987; Woolfolk & Brooks, 1985) sowie eine an den Schülerinnen und Schülern orientierte *Gruppenmobilisierung* betreffen demgegenüber die Nutzungsseite von Lerngelegenheiten (Stallings, 1980; Weinstein, 2007). Diese genannten Strategien (Prozeduren, Gruppenmobilisierung, Zeitmanagement und Klarheit des Handlungsprogramms) werden nachfolgend beschrieben.

Prozeduren

Prozeduren sind Interaktionsskripte, die „das aufeinander bezogene Handeln von Lehrkräften und Schülern [vorschreiben] und umfassen mehrere aufeinanderfolgende Einheiten, deren Durchführung durch ein Signal ausgelöst wird“ (Ophardt & Thiel, 2013, S. 78). Prozeduren müssen mit den Schülerinnen und Schülern gezielt eingeübt, stabilisiert und routinisiert werden (Emmer et al., 1980; Wong & Wong, 2004). Daher findet sich häufig auch eine synonyme Verwendung mit den Begriff Routinen und Rituale. Effektive Routinen werden durch entsprechende Signalgebung der Lehrkraft relativ automatisch bzw. selbstständig von den Schülerinnen und Schülern ausgeführt, ohne den Unterrichtsfluss zu unterbrechen (z. B. beim Verteilen von Arbeitsmaterialien oder dem Zusammenfinden bei Gruppenarbeit, vgl. Arlin, 1979; Kounin & Gump, 1974; Leinhardt et al., 1987). Sie sind also ein effizientes Mittel, um den Handlungsvektor zu etablieren und zu stabilisieren (Doyle, 2006; McGarity & Butts, 1984; Nash, 2009) und dadurch die aktive Lernzeit zu erhöhen (Simonsen, Fairbanks, Briesch, Myers & Sugai, 2008). Es gibt unterschiedliche Vorschläge, Prozeduren zu klassifizieren. Leinhardt und Kollegen (1987) unterscheiden drei Gruppen von Prozeduren: *management* (betrifft nicht-akademische Interaktionen, z. B. Umgang mit zu spät kommenden Schülerinnen und Schülern, Verfahren der Anwesenheitskontrolle), *instructional support* (dient der Herstellung von Lehr-Lern-Situationen, z. B. Austeilen von Arbeitsblättern, Besorgen relevanter Geräte) und *teacher-student exchange* (koordiniert das interaktive Verhalten, z. B. Schülerinnen und Schüler melden sich beim Klassengespräch oder bei Nachfragen; häufig aktivitätsspezifisch). Da vor allem sensible Phasen wie Beginn und Ende des Unterrichts sowie Übergänge besondere Anforderungen an die effektive Orchestrierung der Lehrkraft stellen, sind Prozeduren hier am häufigsten etabliert (ebd.).

Gruppenmobilisierung

Aufbauend auf Kounins Forschungsarbeiten (1970, 1976) zum Gruppenfokus wurde wiederholt bestätigt, dass Strategien der Gruppenmobilisierung die Aktivierung von Schülerinnen und Schülern fördern und zu mehr Beteiligung und aufgabenbezogenem Verhalten, sowie zu weniger Störverhalten führen (Doyle, 2006).

Gruppenmobilisierung meint, dass die Lehrkraft die Klasse als Ganzes fokussiert und gleichzeitig den Radius der am Unterricht beteiligten Lernenden maximiert, indem sie jeden einzelnen Schüler aktiv in den Unterricht einbezieht sowie kontinuierlich akademische Aufgaben zu Verfügung stellt (Doyle, 1986; Kounin, 1970). Der aktive Einbezug erfolgt durch Strategien zur Spannungserzeugung (z. B. wenn Schülerinnen und Schüler nicht wissen, wer als nächstes aufgerufen wird; Intensität mit welcher die Lehrkraft zuhört), durch Techniken des *impression management* (verbale, nonverbale und paraverbale Signale, Thiel et al., 2012, vgl. Mummendey & Bolten, 1985) sowie durch das von Kounin (1970) formulierte Rechenschaftsprinzip (z. B. Anzahl der Schülerinnen und Schüler, die etwas Überprüfbares leisten müssen; Umfang, in welchem die Lehrkraft die Gruppenmitglieder Rechenschaft über ihre Leistungen ablegen lässt; Kontrolle des Arbeitsverhaltens). Monitoring und Gruppenmobilisierung sind daher unmittelbar miteinander interagierende Strategien des Klassenmanagement.

Zeitmanagement

Obwohl in der deutschen Unterrichtsforschung häufiger der Begriff der *Zeitnutzung* Verwendung findet (Baumert & Kunter, 2006; Ditton & Merz, 2000; Helmke, 2009; Neuenchwander, 2006a; Schönbächler, 2008) wird hier der Begriff des *Zeitmanagements* aufgegriffen (vgl. Bromme, 1997; Helmke, 2007a; Stallings, 1980; Weinstein, 2007), der auch die vorausschauende, planerische Komponente (z. B. durch Unterrichtsvorbereitung, situationale Anpassungen) des Lehrerhandels mit umfasst (vgl. Weinert, 1996, der eine adaptive Planungs- und eine adaptive Handlungskompetenz unterscheidet, Kapitel 1.2.1). Eine ungünstige Zeitnutzung betrifft vor allem die Quantität von Unterricht, also die Differenz von nominaler und tatsächlich gewährter Unterrichtszeit (Helmke, 2009), die dadurch entsteht, dass keine oder nicht für alle Schülerinnen und Schüler akademische Aufgaben zur Verfügung gestellt werden (z. B. unpünktlicher Stundenbeginn, Leerläufe, thematisches Abschweifen der Lehrkraft im Klassengespräch). Das aktive Zeitmanagement umfasst darüber hinaus qualitative unterrichtliche Aspekte wie Pacing (für die Schülerinnen und Schüler angemessene Geschwindigkeit der Stoffdarbietung und Wechsel von Aktivitäten, vgl. Gump, 1967; Kounin &

Gump, 1974), Materialorganisation (notwendige Materialien stehen bei Bedarf gut erreichbar zur Verfügung, Leinhardt et al., 1987; Schönbachler, 2008) und die Herstellung eines reibungslosen und schwungvollen Unterrichtsverlaufs durch den Abruf eingeführter Verhaltensroutinen (Administrativa werden simultan zu akademischer Beschäftigung der Schülerinnen und Schüler platziert, Übergänge sind durch Prozeduren gestützt und werden eindeutig markiert, vgl. Arlin, 1979; Coddington & Smyth, 2008; Thiel et al., 2012). Dass Experten die zur Verfügung stehende Unterrichtszeit gezielt koordinieren, zeigte der Vergleich von Experten und Novizen (Thiel et al., 2012) bei Deutsch- und Mathematiklehrkräften–Experten nahmen sich bei der Beendigung von Aktivitäten deutlich mehr Zeit als Novizen, überkompensierten diesen erhöhten Zeitbedarf jedoch deutlich bei der Etablierung eines neuen Handlungsvektors (vgl. auch Bromme, 1992). Doyle (1984) untersuchte Englisch-Lehrer und zeigte, dass effektivere Klassenmanager durchweg eher in der Lage waren, ihr geplantes Handlungsprogramm in der vorgegebenen Unterrichtszeit vollständig zu implementieren.

Klarheit des Handlungsprogramms

Klarheit und Strukturiertheit werden im Rahmen der Prozess-Produktforschung und entsprechenden Modellen und Messinstrumenten zur Unterrichtsqualität nicht dem Klassenmanagement zugeordnet (Baumert, Gruehn, Heyn, Keller & Schnabel, 1997; Ditton & Merz, 2000; Helmke, 2009; Slavin, 1997). Klarheit umfasst nach Helmke (2007b) vier Aspekte: die Hörbarkeit, die sprachliche Prägnanz, die inhaltliche Kohärenz und die fachliche Korrektheit. Strukturiertheit hat demgegenüber zwei Bedeutungen: Einerseits betrifft Strukturiertheit Merkmale des Informationsangebots (Maßnahmen zum Aufbau einer gut organisierten Wissensbasis), andererseits sind didaktische Merkmale gemeint (der Unterricht wird so geplant und sequenziert, dass ein erklärtes Ziel erreicht wird, ebd.). Mit einer ökologischen Perspektive auf Klassenmanagement sind die Konzepte anders konnotiert und orientieren sich an der Herstellung sozialer Ordnung, für die ein klares Handlungsprogramm eine zentrale proaktive Funktion erfüllt (Doyle, 1984; Kounin & Gump, 1974). Klarheit gilt daher als Teil des instruktionsbezogenen Klassenmanagements (vgl. Emmer et al., 1980; Merkens, 2010). Ein Handlungsprogramm gilt als klar, wenn sowohl der Stundenverlauf im Ganzen (geordneter Stundenaufbau, sequentielle Abfolge, curriculare Strukturierung) als auch einzelne Aufgaben und Instruktionen gut strukturiert und für die Schülerinnen und Schüler gut nachvollziehbar sind (Mayr, 2006; Thiel et al., 2012). Es werden nicht nur akademische, sondern auch soziale Verhaltenserwartungen definiert, so dass der Handlungsvektor zügig etabliert werden kann (Doyle, 1983; Thiel et al., 2012). Der nach Helmke (2007b) definierte Aspekt der Klarheit

wird damit nur mittelbar operationalisiert; es werden ausschließlich jene Unklarheiten mit einbezogen, die dazu führen, dass das Handlungsprogramm intransparent ist und oder unzureichend etabliert wurde. Bereits in einer frühen Studie von Arlin (1979) mit Lehramtsstudierenden wurde aufgezeigt, dass gut identifizierbare Übergänge und aktiv orchestrierte Klassenereignisse zu mehr *involvement* der Schülerinnen und Schüler führten (vgl. auch Doyle, 1984). Auch für Strukturierungshilfen (Mayr, 2006; Smith, 1985) und von den Schülerinnen und Schülern verstandene Instruktionen (van de Grift, 2007) konnten Effekte auf kognitive, aber auch motivationale Schülermerkmale nachgewiesen werden (Chilcoat, 1989; Mayr, 2006).

1.2.4. Dimensionen des Klassenmanagements: Management der Lehrer-Schüler-Beziehung

Studien zu schülerzentrierten Forschungsansätzen von Klassenmanagement zeigen, dass ein gut gesteuerter Unterricht nicht ausschließlich auf dem Commitment der Schülerinnen und Schüler und der externalen Kontrolle der Lehrkraft beruhen kann (Brophy, 1985; Hastie et al., 2007; Kunter et al., 2007; McCaslin & Good, 1992; van Tartwijk et al., 2009). Besonders gilt dies für ältere Schülerinnen und Schüler (Mayr, 2006). Starke Verhaltenskontrolle war in einer Studie mit über 3000 singapurischer Schülerinnen und Schülern positiv mit unerwünschtem Schülerverhalten assoziiert (Nie & Lau, 2009) und führt eher zu Reaktanz (Lewis, Romi, Qui & Katz; 2005). Auch Lehrkräfte empfinden die Balance zwischen Kontrolle und Einfühlungsvermögen als zentrale und zum Teil belastende Herausforderung (McLaughlin, 1991; Scarlett, Ponte & Singh, 2009; Weinstein, 1998). In humanistisch geprägten und personzentrierten Ansätzen bildet daher die Gestaltung der Lehrer-Schüler-Beziehung einen zentralen Angelpunkt für einen wirkungsvollen Umgang mit Problemverhalten und eine schülerorientierte Lernumgebung (Bergin & Bergin, 1999; Freiberg, 1999; Gordon, 1974; Rogers & Freiberg, 1994; Tausch & Tausch, 1998; Weinstein, 1999). Dazu gehören ein auf Lernen orientiertes *Arbeitsbündnis* sowie der konstruktive *Umgang der Lehrkraft mit Konflikten unter Schülern*. Mit Bezug auf die Selbstbestimmungstheorie für die Ausbildung von Motivation (Ryan & Deci, 2000; vgl. Kunter et al., 2007; Opdenakker, Maulana & den Brok, 2012; van Tartwijk et al., 2009; Wubbels & Brekelmans, 2005) werden Lehrkräfte dem Bedürfnis ihrer Schülerinnen und Schüler nach sozialer Eingebundenheit, Autonomie- und Kompetenzerleben gerecht, wenn es ihnen gelingt, deren Selbstständigkeit (Heckt, 2005) und Selbstkontrolle (Bergin & Bergin, 1999) bzw. self-guidance (Brophy, 1985) zu fördern (vgl. auch Freiberg, 1999; Hart, 2010; Mayr, 2006; Opdenakker et al., 2012; Seidel, 2009; Ophardt & Thiel,

2013). Dies ist wichtig, um Steuerungsverantwortung mit den Schülerinnen und Schülern zu teilen sowie ihre persönliche Entwicklung zu unterstützen (Freiberg, 1983).

Arbeitsbündnis

Der Begriff des Arbeitsbündnisses (*working alliance*) ist der Forschung zur therapeutischen Beziehung entliehen und wird seit geraumer Zeit–mit entsprechender Anpassung–auch auf eine Subkomponente des Lehrer-Schüler-Verhältnisses angewandt, der sich auf Aspekte des Führungsstils und des Klassenklimas bezieht (Larose, Chaloux, Monaghan & Tarabulsy, 2010; Rogers, 2009; Toste, Heath & Dallaire, 2010; Ursano, Kartheiser & Ursano, 2007). Angelehnt an die klinisch-psychologische Definition von Bordin (1979) konstituiert sich das Arbeitsbündnis von Lehrkraft und Schülerinnen und Schülern durch gemeinsame Ziele, festgelegte Aufgaben zur Erreichung dieser Ziele und eine gegenseitige Verpflichtung und Vertrauen in die Erreichung dieser Ziele („agreement on goals, assignment of task or a series of tasks, and the development of bonds“, S. 253; vgl. auch Eigenmanns Konzept der *engagierten Gegenseitigkeit*, 2009). Ein Arbeitsbündnis herzustellen meint demnach, den Schülerinnen und Schülern verantwortlich, fürsorglich und unterstützend gegenüber zu treten und gleichzeitig hohe akademische und soziale Erwartungen an diese zu stellen (Freiberg, 1999; Ophardt, 2006; Pianta & Hamre, 2006; Seidel, 2009; van Tartwijk et al., 2009). Klassenmanagement kann somit nur dann die Instruktion stützen, wenn auch anspruchsvolle Lernziele und ein lernförderliches Klima in der Klasse etabliert werden. Die Erweiterung des Konzepts Klassenmanagement um den Aspekt des Arbeitsbündnisses ist daher notwendig, um Klassenmanagement hinreichend fassen zu können (vgl. Ophardt & Thiel, 2013; Weinstein, 1999). Sowohl das individuelle, als auch das kollektive Arbeitsbündnis (Dubs, 1995; Oevermann, 1996; Rogers, 2009) fördern die Motivation, die sozio-emotionale Entwicklung sowie akademische Leistungen von Schülerinnen und Schülern (Eder & Bergmann, 2004; Larose et al., 2010; Nie & Lau, 2009; Sergiovanni, 1994; Slavin, Hurley & Chamberlain, 2003; Toste et al., 2010). Ein gut funktionierendes Arbeitsbündnis stärkt auch die Selbstwirksamkeitsüberzeugungen von Lehrkräften sowie die Freude am Beruf (Spilt, Helma, Koomen & Thijs, 2011).

Umgang mit Konflikten unter Schülern

Konflikte spiegeln widerstreitende Interessen, Wertvorstellungen oder Ziele der Interaktionsteilnehmer wider und werden von den Konfliktparteien als Einschränkung ihrer subjektiven Handlungsfreiheit aufgefasst (Ophardt & Thiel, 2013). Konflikte treten häufig vor und nach dem Unterricht auf, seltener während des Unterrichtsgeschehens (van Tartwijk et al., 2009). Im Rahmen des Klassenmanagements sind Konflikte dann relevant, wenn sie sich ne-

gativ auf das Lernklima und die soziale Teilhabe der Schülerinnen und Schüler auswirken und damit ein Eingreifen der Lehrkraft erforderlich machen (Dubs, 1995; Emmer et al., 2003; LePage et al., 2005; Oser, 1997a). Konflikte werden sowohl in der Metakommunikation im Unterricht als auch außerhalb dessen als Konfliktgespräche mit Hilfe von Techniken konstruktiver Konfliktlösung thematisiert und bearbeitet (Johnson & Johnson, 2006; Walker, Ramsey & Gresham, 2004). Dabei werden vier Schritte durchlaufen: Klärung von Gesprächsregeln und Zieldefinition, Identifikation des Problems, Entwicklung von Lösungsmöglichkeiten und Diskussion von Lösungsalternativen sowie die Herstellung eines Einverständnisses über die Erprobung der Lösung (Neubauer, Gampe & Knapp, 1992; Wertheim, Love, Peck & Littlefield, 1998). Auch weitere Personen (Eltern, Schulleiter, Konfliktschlichter) können dabei einbezogen werden. Die Konfliktlösung sollte fair und konstruktiv erfolgen und ist stets an der Stabilisierung eines langfristigen und lerndienlichen Arbeitsbündnisses orientiert (Brophy, 1996; Ophardt & Thiel, 2013; Walker et al., 2004). Oser (1997a) gehört im deutschsprachigen Raum zu den ersten, der die gezielte Auslagerung von Konflikten zur Bewältigung von Disziplinproblemen und Schülerrisiken als Standard formulierte und indirekt Strategien des Klassenmanagements zuordnete (vgl. Kapitel 1.1.3).

1.3. Erfassung von Kompetenzen im Klassenmanagement

Trotz des zentralen Stellenwerts für schulisches Lernen und die professionelle Expertise von Lehrkräften ist eine wissenschaftlich fundierte Erfassung von Kompetenzen im Klassenmanagement derzeit nur eingeschränkt möglich. Nachfolgend werden zunächst potentielle methodische Zugänge dargestellt, die eine valide Erfassung von Klassenmanagement erlauben (Kapitel 1.3.1), ein Überblick über die bislang verfügbaren Verfahren und Instrumente zur Erfassung von Klassenmanagement gegeben (Kapitel 1.3.2) sowie die zu berücksichtigenden Schritte bei einer empirisch fundierten Fragebogenkonstruktion—orientiert auf das vorliegende Forschungsanliegen—dargestellt (Kapitel 1.3.3).

1.3.1. Methodische Zugänge

Der Bedarf an proximalen, also direkten Indikatoren für Lehrerkompetenzen mit hoher psychometrischer Güte gilt derzeit als großes Forschungsdesiderat (Baumert & Kunter, 2006; Frey, 2005; Klieme, 2004; Oser, Curcio & Düggele, 2007; Terhart, 2007), auch und insbesondere für den Bereich Klassenmanagement (Blömeke et al., 2008a; Kemna, 2012; Voss et al., 2011). Obwohl Klassenmanagement im Vergleich zu anderen Prozessmerkmalen des Unterrichts als gut beobachtbar gilt, werden häufig eher distale Indikatoren zur Erfassung der Effektivität von Klassenmanagement erhoben (vgl. Baumert & Kunter, 2006; Brophy, 2006),

z. B. das Ausmaß aufgabenbezogenen Schülerverhaltens (Bromfield, 2006; Leflot et al., 2010), auf Klassenmanagement bezogene Lehrerselbstwirksamkeit (Brouwers & Tomic, 2000) oder Überzeugungen (Martin, 2004; Martin & Yin, 1999). Allerdings handelt es sich hierbei eher um *Ergebnisse* (Kirkpatrick, 1998) bzw. assoziierte Merkmale des Klassenmanagements: Es können nur indirekt Rückschlüsse auf die tatsächlich vorhandenen Kompetenzen gezogen werden. Auch ist darin implizit die Annahme enthalten, Klassenmanagement sei ein-dimensional zu konzipieren. Während diese verkürzte Sichtweise zum Beispiel zur Programmevaluation oder im Rahmen breit angelegter Studien durchaus zielführend sein kann, sind Aussagen über das Zusammenwirken oder die prognostische Validität einzelner Kompetenzfacetten daraus nicht abzuleiten.

Fragebogen- und Testverfahren

Zu den direkten Erhebungsmethoden zählen Fragebögen (Kompetenzselbsteinschätzung, Schülerbefragung, Beobachterrating) und das durch Tests erfasste Wissen und Können von Lehrkräften (Kunter & Klusmann, 2010). Vor allem im Bereich der Testentwicklung wurden in jüngerer Zeit vermehrt Anstrengungen unternommen, um Alternativen zu den populären Selbsteinschätzungen zu entwickeln (König & Blömeke, 2009). Bezüglich Klassenmanagement sind hier von Bischoff und Kollegen (2005), Gold, Förster & Holodynski (2013) sowie in den Projekten COACTIV (Voss et al., 2011) und TEDS-M (Blömeke et al., 2008a; König, 2012; König & Blömeke, 2010) anspruchsvoll entwickelte Instrumente entstanden, die über verschiedene geschlossene und auch offene Formate (Multiple-Choice-Aufgaben, Vignetentests) Wissen und Können im Klassenmanagement zu erfassen versuchen. Dabei wird vor allem auf den von Oser und Kollegen (2007) vorgeschlagenen advokatorischen Zugang zurückgegriffen, bei dem Professionelle über andere Professionelle urteilen und aus diesem Urteil auf die eigene Kompetenz geschlossen wird (vgl. auch *situational judgement tests*, Weekly & Ployhart, 2006).

Trotz zahlreicher Vorteile diagnostischer Testverfahren, u. a. die hohe Objektivität durch geringe Fälschbarkeit, die Möglichkeit der Normierung sowie die situationsunabhängige Erfassung gibt es erhebliche Nachteile, die den Einsatz der bislang entwickelten Tests zum Klassenmanagement–zumindest noch–einschränken: Die angesprochenen Verfahren befinden sich zum Teil noch im Entwicklungsstadium (z. B. weil noch keine Normwerte vorliegen), die Reliabilität ist–wie die Autoren z. T. selbst anmerken–bei allen genannten Testverfahren an der unteren Grenze (vgl. auch Frey, 2006) bzw. bleibt unerwähnt (Bischoff et al., 2005). Der bisherige Nutzen für die empirische Forschung ist damit noch begrenzt (Blömeke, 2007).

Weiterhin ist der Einsatz von Kompetenztests–je nach Auswertungsform–in der Regel nur mit großen personellen und finanziellen Ressourcen realisierbar. Gegen einen Einsatz im Bereich der *Kompetenzentwicklung* und Veränderungsmessung sprechen insbesondere die potentiell eingeschränkte Sensitivität dieser Verfahren (u. a. aufgrund ihrer Ungenauigkeit) und nicht kontrollierbare Konfundierungseffekte, beispielsweise durch sprachliche Kompetenzen (ebd., vgl. auch Kemna, 2012; Lüders, 2012). Auch Testübung schränkt den Nutzen bei wiederholter Messung ein, da diese mit einer wahren Wissens- oder Kompetenzveränderung konfundieren kann.

Der Fokus der vorliegenden Arbeit liegt daher auf den ökonomischeren und–unter bestimmten Bedingungen hoch sensitiven und validen–Fragebogenverfahren (oft auch *Ratings* genannt, vgl. Langer & Schulz von Thun, 1974). Die Unterrichtsforschung greift hierbei im Wesentlichen auf drei Datenquellen zurück: Schülerinnen und Schüler, die Adressaten des Unterrichts, Beobachter, die den „Blick von außen“ repräsentieren und schließlich die unterrichtenden Lehrkräfte selbst (vgl. Clausen, 2002). Die Übereinstimmung verschiedener Verfahren kann sich in Abhängigkeit von unterschiedlichsten Moderatoren wesentlich voneinander unterscheiden (Elliot & Busse, 2004). Clausen (2002) identifizierte die drei Merkmale *Beobachtbarkeit*, *fachlich-didaktischer Anspruch* und *Evaluationsgehalt* der einzuschätzenden Merkmale als zentrale Vermittler von Übereinstimmung.⁹ Aus seinen Ergebnissen zur Perspektivität von Unterrichtsqualität, die auf Multi-Trait-Multi-Method-Analysen beruhen, zog er jedoch nicht den Schluss, dass damit die Nützlichkeit oder psychometrische Güte dieser Instrumente per se eingeschränkt sei (vgl. auch Elliot & Busse, 2004), sondern wies jeder der genannten Perspektiven eine *spezifische* Validität zu. Wagner (2008) mahnt jedoch an, dass spezifische Verzerrungstendenzen nicht unreflektiert als valide Varianzquellen betrachtet werden dürfen. Die multiperspektivistische bzw. multimethodale¹⁰ Erfassung (im Sinne der Triangulation) unterrichtlicher Prozesse gilt daher als zentrales Gütekriterium in der Unterrichtsforschung (Mühlig & Petermann, 2006).

Erfassung von Klassenmanagement durch die Befragung von Lehrkräften

Standardisierte Selbsteinschätzungen (auch der subjektiv eingeschätzte *Kompetenzerwerb*) gelten als der ökonomischste Zugang zum Feld und werden daher häufig im Rahmen der Evaluation von Trainings oder von Studienabschnitten verwendet (Braun & Hannover,

⁹ Übereinstimmung ist dann wahrscheinlich, wenn das Zielkriterium beobachtbar ist (potentielle Herausforderung für Beobachter), geringe fachlich-didaktische Ansprüche aufweist (potentielle Herausforderung für Schülerinnen und Schüler) sowie einen geringen evaluativen Gehalt hat (potentielle Herausforderung für Lehrkräfte).

¹⁰ *Perspektive* und *Methode* werden in der vorliegenden Arbeit synonym verwendet (vgl. Schermelleh-Engel & Schweizer, 2007)

2008; Frey, 2006; König & Blömeke, 2009; Kunter & Klusmann, 2010; vgl. z. B. Blüthmann, Ophardt, Thiel & Felsberger, 2011; Nuangchalerm, 2009; Thiel, Blüthmann & Watermann, 2012). Gerade im Rahmen der universitären Ausbildung können häufig keine Verhaltensbeobachtungen oder Schülerbefragungen durchgeführt werden (Paechter, Skliris & Macher, 2011). Selbsteinschätzungen wird eine hohe Augenscheinvalidität zugesprochen, sie gelten als handlungsleitend und sind damit relevant für die Prognose zukünftigen Verhaltens (Cramer, 2010; Eckert, Dunn, Guiney & Coddling, 2000; Fernandez-Ballesteros, 2004; Lucas & Baird, 2006; Mabe & West, 1982). Dennoch ist die Validität von Selbsteinschätzungen eher begrenzt: Es zeigen sich nur kleine bis mittlere Zusammenhänge mit anderen Beurteilungen oder objektiven Maßen (Freund & Kasten, 2012; Mabe & West, 1982); Konfundierungen mit Drittvariablen sind wahrscheinlich und führen sowohl zu Über- als auch zu Unterschätzungen (z. B. evaluativer Gehalt des Zielkriteriums, verzerrte Antworten durch Impression Management oder Verhaltenserwartungen wie Lehrerselbstwirksamkeit, vgl. Fernandez-Ballesteros, 2004; Kunter & Klusmann, 2010). So können Selbsteinschätzungen von Wissen und Können zwar prinzipiell eine diagnostische Funktion übernehmen (König, 2012; Nolle, 2004; Terhart, 2007), es werden jedoch relativierende Bezeichnungen wie „Kompetenzselbstkonzept“ (Maag Merki & Grob, 2005), „Selbstkonzept beruflicher Kompetenz“ (Kaufhold, 2006), „subjektive Kompetenzeinschätzung“ (Gröschner, Schmitt & Seidel, 2013) oder „Kompetenzerwartung“ (Cramer, 2010) gewählt.

Erfassung von Klassenmanagement durch die Befragung der Schülerinnen und Schülern

Die Befragung von Schülerinnen und Schülern ist seit den 90er Jahren als reliable und valide Datenquelle in der Unterrichtsforschung etabliert (Aleamoni, 1999; Fraser, 1991; Greenwald, 1997; de Jong & Westerhof, 2001). Schülerinnen und Schüler gelten als „natürliche Experten“ des Unterrichts (Kämpfe, 2009); ihr Urteil hat insbesondere für die Lehrenden einen hohen empirischen Gehalt und ist daher ein zentraler Bestandteil der systematischen Evaluation von Unterricht (Gärtner, 2011; Horster & Rolff, 2001).

Zentrale Güteindikatoren von Schülerbefragungen sind die prädiktive Validität für die kognitive, affektive und motivationale Entwicklung von Schülerinnen und Schülern (Clausen, 2002; Gruehn, 2000) sowie die hohe Reliabilität aggregierter Schülerurteile, die die geteilte bzw. kollektive Wahrnehmung abbilden (Cronbach, 1976; Gruehn, 2000; Kenny, 2004; Lüdtke, Trautwein, Kunter & Baumert, 2006). Insbesondere jene Bereiche und Prozesse des Unterrichts können Schülerinnen und Schüler valide beurteilen, die wenig didaktisch-pädagogisches Wissen erfordern (Clausen, 2002). Die Einschätzungen von Schülerinnen und

Schülern basieren auf einer sehr großen Verhaltensstichprobe, sodass Lehrkräfte intraindividuell über verschiedene Situationen oder Fächer hinweg als auch interindividuell im Vergleich mit anderen Lehrkräften beurteilt werden können (Ditton, 2000; Gärtner, 2010).

Zu den Schwächen von Schülerbefragungen zählen das hohe Verzerrungspotenzial (z. B. durch Halo-Effekte, Akquieszenz oder sozial erwünschte Antworten) und die eher globale Sichtweise auf Unterricht (Aleamoni, 1999; Baumert, Kunter, Brunner, Krauss, Blum & Neubrand, 2004; Greenwald, 1997; Wagner, 2008). Unklar bleibt oft auch, welcher Bewertungsmaßstab zu Grunde gelegt wurde und auf welchen Zeitraum sich die Einschätzungen beziehen (Clausen, 2002; Baumert et al., 2004).

Erfassung von Klassenmanagement durch standardisierte Beobachtung

Die standardisierte Unterrichtsbeobachtung galt lange Zeit als anerkannteste Datenquelle in der empirischen Unterrichtsforschung, insbesondere im Bereich der Kompetenzdiagnostik (Beck, 1987; Clausen, 2002; Desimone; 2009; Evertson & Burry, 1989). Im Vergleich zu Schülerbeurteilungen und Selbstauskünften von Lehrkräften wird Beobachtern ein unbefangener, objektiver Blick auf unterrichtliche Prozesse zugeschrieben; dies gilt insbesondere für Merkmale mit evaluativem Gehalt und hohen Ansprüchen an das pädagogisch-didaktische Verständnis.

Die Verhaltensstichprobe, auf die sich eine Beobachtung bezieht, ist in der Regel kürzer als die von Schülerinnen und Schülern und der Lehrkraft und damit mehr durch situative Einflüsse geprägt (u. a. Uhrzeit am Tag, Zeitpunkt im Schuljahr, Anwesenheit bzw. Abwesenheit bestimmter Schülerinnen und Schüler; vgl. Beck, 1987; Marshall & Green, 1977). Auch die Beobachtungssituation an sich birgt die Gefahr einer positiven wie negativen Verzerrung „echten“ Unterrichts (z. B. positiv: eine intensivere Vorbereitung der Lehrkraft, negativ: Nervosität der Lehrkraft). Voraussetzung für eine zuverlässige Beobachtung ist demnach eine ausreichende Basisrate des zu beobachtenden Verhaltens, sowie eine auf konkreten (verhaltensbezogenen) und kontextspezifischen (im Sinne konditionaler Festlegungen, z. B. mit Blick auf vorherrschende Aktivitätsstrukturen, das Unterrichtsfach oder das Alter der Schülerinnen und Schüler) Indikatoren gestütztes Beurteilungssystem (Bortz & Döring, 2006; Langer & Schulz von Thun, 1974). Eine besondere Herausforderung für Beobachter ist dabei, dass gelungenes Klassenmanagement sich häufig durch das Unterlassen von Handlungen äußert bzw. implizit bleibt (z. B. wird es in einem gut geführten Klassenzimmer möglicherweise nur selten zu Konflikten oder massivem Störverhalten kommen, vgl. Brophy, 1999). Um die psychometrische Güte von Beobachterurteilen zu gewährleisten sind intensive Schu-

lungen unabdingbar. Potentielle Verzerrungen in den Bereichen Wahrnehmung, Beobachtung und Beurteilung (Grewe & Wentura, 1997; Hoyt & Kerns, 1999) können hierbei thematisiert und in der Beurteilungssituation reflektiert werden. Die Durchführung von Beobachtungen ist im Vergleich zu anderen standardisierten Befragungen relativ aufwendig und kostenintensiv.

1.3.2. Vorliegende Fragebogenverfahren zur Erfassung von Klassenmanagement

Die Erfassung von Klassenmanagement erfolgt häufig durch theoretisch und/oder empirisch relativ wenig fundierte Selbst- und Beobachtungsinstrumente, z. B. in Form sogenannter *Checklists* (Gordon, 2002; MacSuga & Simonsen, 2011; McGarity & Butts, 1984; Mir Pozo, 2003; van Tassel-Baska, van Tassel-Baska, Quek & Feng, 2006; Webster Stratton, Reid & Hammond, 2001, vgl. Brookhart, 2004). Hierbei werden vor allem Disziplinschwierigkeiten und der Umgang mit Schülerverhalten fokussiert (vgl. auch Copeland, 1987; Gordon, 2002; Helmke, Schneider & Weinert, 1986). Die elaboriertesten Instrumente entstanden in den 80er Jahren in der Arbeitsgruppe um Evertson (Evertson, 1985; Evertson & Burry, 1989) und im Zuge breiter angelegter Verfahren zur Unterrichtsqualität (Pianta & Hamre, 2009; van de Grift, 2007). Schülerbefragungen bezogen sich bis in die 90er Jahre fast ausschließlich auf die Bereiche Unterrichts- und Schulklima (Eder, 1998; Fend, 1977; Fraser, 1980; Fraser, Anderson & Walberg, 1982; Moos & Trickett, 1974; Oswald, Pfeiffer, Ritter-Berlach & Tanzer, 1989; Saldern & von Littig, 1987; Wahlberg, 1966); einige dieser Instrumente bildeten die Itemgrundlage der aktuell verfügbaren Instrumente und Skalen (z. B. Eder, 1998 oder Moos & Trickett, 1974, vgl. Baumert et al., 1997; Mayr et al., 2005; Schönbächler, 2008). Auch heute werden klimaorientierte Instrumente häufig als Indikatoren für Klassenmanagement eingesetzt (vgl. den Brok, Brekelmans & Wubbels, 2006; Wubbels & Levy, 1993).

Im deutschen Sprachraum wird Klassenmanagement überwiegend als Subaspekt („Klassenführung“ und/oder „Zeitnutzung“) im Rahmen von Unterrichtsqualität mit erfasst, vorrangig durch Lehrer- und Schülerfragebögen (Baumert et al., 1997; Ditton & Merz, 2001; Helmke, Hosenfeld, Schrader & Wagner, 2002; Mayr, 1998), aber auch durch Ratings (Baer et al., 2009; Clausen, 2002; Helmke, Helmke, Schrader & Wagner, 2007; Rakoczy & Pauli, 2006; Weinert et al., 1990), die vor allem für den Einsatz an videografierten Unterrichtsmit-schnitten konzipiert sind. Insbesondere die in COACTIV (Baumert et al., 2009), DESI (*Deutsch-Englisch-Schülerleistungen-International*, Klieme & Pauli, 2006) und BIJU (*Bildungsverläufe und psychosoziale Entwicklung im Jugendalter*, Baumert, Gruehn, Heyn, Köller & Schnabel, 1997; Gruehn, 2000) verwendeten Operationalisierungen verweisen auf eine theoretisch umfassendere Konzeptualisierung von Klassenmanagement; neben Disziplin und

Zeitnutzungsaspekten werden auch störungspräventive Strategien, wie Regelarbeit und Monitoring, berücksichtigt.

Nur wenige deutschsprachige Verfahren adressieren ausschließlich Klassenmanagement (Lotz, Berner, Gabriel, Post, Faust & Lipowsky, 2011; Mayr et al., 2005; Neuenschwander et al., 2003; Schönbächler, 2004a, 2004b; Wannack, Herger, Gruber & Barblan, 2009). Die erste mehrdimensionale Konzeptualisierung erfolgte mit dem *Linzer Diagnosebogen zur Klassenführung* (LDK, für Lehrer und Schüler, Mayr et al., 2005), der aus dem *Linzer Diagnosebogen zum Schul- und Klassenklima* hervorging (Eder, 1998). Er enthält drei Dimensionen mit 21 pädagogischen Handlungsstrategien (vgl. Kapitel 1.2.1 bzw. Fußnote Nr.7): *Gestaltung des Unterrichts*, *Förderung der Beziehung* und *Kontrolle des Verhaltens*. Erstmals wurde damit die verhaltensorientierte Sichtweise mit der eher systemisch-ökologischen Gestaltung von Lernumgebungen sowie der Berücksichtigung der Lehrer-Schüler-Beziehungen zusammengeführt und die von Kounin (1970, 1976) extrahierten Merkmale erfolgreicher Klassenmanager (Allgegenwärtigkeit, Beschäftigung der Schülerinnen und Schüler als Teil des Gruppenfokus) in ein empirisches Relativ überführt. Auch die Klarheit des Handlungsprogramms wird hier über die Skalen *Strukturierter Unterricht* und *Klare Arbeitsanweisungen* abgebildet. Der Abstraktionsgrad der operationalisierten Merkmale ist jedoch sehr unterschiedlich und Klassenmanagement durch den Einbezug von Fachkompetenz sehr weit gefasst.

Schönbächler (2004a, 2004b) übertitelte ihre auf dem LDK basierenden Instrumente (für Schüler und Lehrer) erstmalig mit dem Begriff Klassenmanagement und erweiterte in der Folge den LDK u.a. um die Skalen *Regeln*, *Unterrichtsfluss*, *Materialorganisation*, *Zeitnutzung* und *Schülerpartizipation*. Leider wurde die faktorielle Validität der Instrumente nur ausschnittsweise geprüft; die Augenscheinvalidität einiger Skalen-Item Zuordnungen ist vor dem Hintergrund der in dieser Arbeit skizzierten Aspekte von Klassenmanagement nicht immer gegeben.

Schließlich liegen noch Selbsteinschätzungsskalen von Wannack, Herger, Gruber und Barblan (2012, mit den Skalen *Regeln*, *Organisation*, *Aufmerksamkeit* und *Gestaltung des Klassenzimmers*) und Neuenschwander et al. (2003, mit den Skalen *Regeln setzen*, *Flexibilität*, *effektive Zeitnutzung*, *Bildungsorientierung*, *Kommunikationskompetenz*, *Diagnosekompetenz* und *Lehrerethos*) vor. Die psychometrische Prüfung der Instrumente ist bislang nur unzureichend beschrieben bzw. nicht zufriedenstellend.

Es existiert lediglich ein deutschsprachiges, mehrdimensionales Rating zur Klassenführung, das kürzlich in der PERLE-Studie für den Einsatz an videografiertem Unterricht entwi-

ckelt wurde (Lotz et al., 2011; Lotz & Gabriel, 2012); das Instrument wurde jedoch noch nicht publiziert (vgl. Lotz, Gabriel & Lipowsky, 2013).

1.3.3. Prinzipien der Fragebogenkonstruktion

Während einige Aspekte von Klassenmanagement bereits gut abgebildet werden können, so sind vorliegende Operationalisierungen zum Teil noch unbefriedigend oder liegen schlicht nicht vor. Anhand der Phasen der Testkonstruktion nach Lienert und Raatz (1998) werden Konstruktionsprinzipien zur multidimensionalen und multimethodalen Erfassung von Klassenmanagement strukturiert (vgl. auch Bühner, 2004; Meyer, 2004):¹¹

- 1) Planungsstadium
- 2) Testentwurf und Aufgabenkonstruktion
- 3) Itemanalyse
- 4) Verwertung der Analysedaten
- 5) Empirisches Kontrollstadium
- 6) Testeichung/Testnormierung

Ist ein Bedarf für neue Instrumente identifiziert, wird im *Planungsstadium* der Geltungsbereich (z. B. Klassenstufe, Schulfächer, Schularten) der Instrumente festgelegt (vgl. Gütekriterien nach Gollwitzer & Jäger, 2009) sowie theoretisch fundiert und unter Berücksichtigung der spezifischen Validität der verschiedenen Methoden (Clausen, 2002) eine Auswahl der abzubildenden Dimensionen getroffen, um das Zielkonstrukt hinreichend zu erfassen. Wichtig ist, einander ausschließende Dimensionen zu bilden, um eine ausreichende Differenzierbarkeit zwischen den Dimensionen zu gewährleisten.

Im Rahmen der *Aufgabenkonstruktion* werden vorhandene Items sondiert, ggf. angepasst und bei Bedarf neue Items formuliert. In der Tradition kumulativer Forschung ist die Verwertung und potentielle Optimierung existierender Items anzustreben (Kirchoff, Kuhnt & Lipp, 2008). Unter angemessener Berücksichtigung der vier kognitiven Stadien der Itembeantwortung (Verstehen, Informationen abrufen, Urteil festlegen, Antwortabgabe, vgl. Jankisz & Moosbrugger, 2007) können zentrale Grundsteine für die psychometrische Güte des zu entwickelten Instruments gelegt werden. Hierzu gehören verhaltensnah und eindeutig formulierte Items, die Verwendung zielgruppenadäquater Sprache, die Berücksichtigung der voraussichtlichen Diskriminationsfähigkeit des Items, die Auswahl des angemessenen Inferenz-

¹¹ Die konkrete Ausgestaltung der Phasen ist nicht erschöpfend sondern fokussiert die für diese Arbeit relevanten Prinzipien. Da die Fragebogenkonstruktion deduktiv, also auf Grundlage der rationalen Konstruktionsstrategie erfolgte, orientiert sich die Ausdifferenzierung der Phasen an diesem Vorgehen.

niveaus,¹² eine methodenspezifische Anzahl der Antwortoptionen, klare und informative Instruktionen (z. B. für die Beurteilung zu berücksichtigendes Zeitintervall; Zusicherung der Anonymität), Invertierung von Items sowie ungerade Antwortoptionen zur Vermeidung von Antworttendenzen (z. B. Akquieszenz, Tendenz zur Mitte, vgl. Killus, 2006; Krosnick, 1999; Langer & Schulz von Thun, 1974; Rammstedt, 2006; Wagner, 2008).

In der Phase der *Itemanalyse* erfolgt eine erste empirische Prüfung des entwickelten Instruments. Dafür werden statistische Maße und Kennwerte festgelegt, um die Güte einzelner Items bestimmen zu können. Nicht nur klassische Kennwerte wie Trennschärfe und Schwierigkeit, sondern auch die Verständlichkeit einzelner Items, der Instruktion sowie der Beantwortungsoptionen sollten hierbei überprüft werden. In konfirmativen Faktorenanalysen–wenn notwendig, unter Berücksichtigung von *nested structures*, der Mehrebenenstruktur–werden die Konsistenz der postulierten Faktoren und die Interkorrelation der Faktoren überprüft (Köller, 2009). Bei mehreren Beurteilern können gezielt Items ausgewählt werden, die zu hohen Übereinstimmung führen (anhand von Übereinstimmungsindizes) oder viel Varianz zwischen den Klassen aufklären (ICC[1], Lüdtke et al., 2006; Wagner, 2008).

Die *Verwertung der Analysedaten* impliziert die Überarbeitung oder die Eliminierung von Items anhand definierter Kriterien. Das Entfernen von Items muss stets im Hinblick auf den Erhalt der inhaltlichen Validität erfolgen (Hartig, Frey & Jude, 2007).

Erst dann, im *empirischen Kontrollstadium*, liegt das (ggf. noch vorläufige) Instrument in seiner finalen Version vor. Im Rahmen von Objektivitäts-, Reliabilitäts- und Validitätsanalysen wird die Nützlichkeit des Instruments überprüft. Die Bestimmung der Reliabilität kann sich dabei nicht nur an Cronbachs Alpha zur Quantifizierung der internen Konsistenz orientieren (vgl. Schmitt, 1996), sondern sollte auch Maße der Interraterreliabilität und Stabilität erfassen. Validitätsprüfungen sollten die inhaltliche, faktorielle, diskriminante, konvergente und wenn möglich prognostische Validität des entwickelten Instrumentes umfassen, auch mit Hilfe der Multi-Trait-Multi-Method Validierung (Courvoisier, Nussbeck, Eid, Geiser & Cole, 2008; Messick, 1995). Die Phasen 1) bis 5) werden im Rahmen von Pilotierungsstudien unter

¹² Die Operationalisierung von zu bewertenden Unterrichtsmerkmalen erfolgt auf einem Kontinuum von hoch- bis niedrig-inferenten Beobachtungskategorien (Babad, 1996; Chavéz Chavéz, 1984; Clausen, Reusser & Klieme, 2003; Langer & Schulz von Thun, 2007; Mishra, 1981). Während niedrig-inferente Beobachtung eher spezifische, beobachtbare Verhaltensweisen kodiert, die kaum schlussfolgernde Kognitionen erfordert (z. B. Anzahl der Zurechtweisungen, Häufigkeit von Strukturierungshinweisen), wird in hoch-inferenten Beurteilungen auf abstrakte Sachverhalte oder allgemeine Verhaltenstendenzen geschlossen (z. B. Qualität von Zurechtweisungen, Güte von Strukturierungshinweisen). Kompetenzen werden üblicherweise durch höher-inferente Instrumente erfasst, da hierbei–gegenüber niedrig-inferenten Verfahren–von einer höheren Stabilität (im Sinne replizierbarer Ergebnisse über die Zeit) und Validität der Beurteilungen ausgegangen wird (Clausen, 2002; Lotz, Gabriel & Lipowsky, 2013).

Umständen mehrere Male wiederholt, bis ein psychometrisch hochwertiges Instrument vorliegt.

Schließlich werden im Rahmen der *Testeichung* Verteilungs- und Gruppenhomogenitäts-Analysen durchgeführt. Vorrangiges Ziel ist die Erstellung von Normtabellen, damit absolute Werte in Relation zu einer relevanten Vergleichspopulation gesetzt werden können. Zur Testeichung sind große und repräsentative Stichproben notwendig.

1.4. Entwicklung von Kompetenzen im Klassenmanagement

Die Lehrerbildung umfasst sowohl die Erstausbildung (wissenschaftliche Erstausbildung an Universität und Fachhochschule sowie praktische Erstausbildung im Vorbereitungsdienst bzw. Referendariat) als auch das Lernen im Beruf (kontinuierliche Fort- und Weiterbildung, vgl. KMK, 2004a). Die von der KMK (2004a) festgehaltenen Ausbildungsstandards mit dem Ziel einer „möglichst hohen professionellen Qualität“ (S. 6) angehender Lehrkräfte beinhalten nicht nur den Erwerb professionellen Wissens, sondern auch „berufsbezogenes Reflexionsvermögen, eine eigenständige Urteilsfähigkeit sowie die Erprobung und konkrete Einübung eines breiten Handlungsrepertoires (inklusive der Routinisierung mancher beruflicher Handlungsformen)“ (S. 6). Auch die Weiterentwicklung der eigenen Kompetenzen wird als Standard definiert (Terhart, 2002a, vgl. auch Hammerness et al., 2005), so dass am Ende der Ausbildung zwar „die Fähigkeit zur Bewältigung der beruflichen Aufgaben gesichert sein [soll]; das Ziel ist aber der weiterhin lernfähige und lernbereite Lehrer.“ (S. 7)

Diesen Vorgaben steht in Deutschland ein erheblicher Mangel an systematischer Aus- und Weiterbildung im Klassenmanagement gegenüber (Helmke, 2007a; Kunter & Baumert, 2011; Ophardt & Thiel, 2008; Schneider & Bodensohn, 2010). Noch immer gehört Klassenmanagement zu den Bereichen, in dem Studierende sich am schlechtesten auf ihre spätere Berufsausübung vorbereitet fühlen (Meister & Jenks, 2000; Merrett & Wheldall, 1993; Pigge & Marso, 1997)–die Ausbildung sei zu allgemein, zu theoretisch und zu sehr auf einfache Probleme fokussiert. Befragungen von Ausbildern und Vertretern der Schulaufsicht sowie Curriculumsanalysen unterstützen diese Einschätzung (Jones, 2006). Gleichzeitig werden Aspekte des Klassenmanagements als besondere Herausforderung beim Unterrichten erlebt, vor allem der Umgang mit Unterrichtsstörungen und Konflikten–nicht nur von Berufsanfängern (Admiraal et al., 2000; Brouwers & Tomic, 2000; Friedman, 2006; Gallagher, 2009; Meister & Melnick, 2003; Veenman, 1984; Wheldall & Merrett, 1988). Geringe Kompetenzen und Selbstwirksamkeitsüberzeugungen im Klassenmanagement sind assoziiert mit erhöhtem Stresserleben und Burnout (Brouwers & Tomic, 2000; Geving, 2007; Hastings & Bham,

2003; Schwarzer & Hallum, 2008) sowie einem frühen Ausscheiden aus dem Beruf (Ingersoll, 2002; Jones, 2006), hingegen begünstigen Kompetenzen in Klassenführung eine erfolgreiche Bewältigung beim Berufseinstieg (Klusmann, Kunter, Voss & Baumert, 2012).

Im folgenden Kapitel wird der Frage nachgegangen, wie die Entwicklung von Kompetenzen im Klassenmanagement im Rahmen der Erstausbildung sowie der Weiterentwicklung im Beruf systematisch unterstützt werden kann. Dafür werden zunächst die spezifischen wissenspsychologischen Aspekte von Kompetenzen im Klassenmanagement sowie ausgewählte Modelle zur Kompetenzgenese vorgestellt (Kapitel 1.4.1). Daran anknüpfend werden Lerngelegenheiten systematisiert, die den Entwicklungsprozess in verschiedenen Entwicklungsphasen fördern können (Kapitel 1.4.2). Ein Überblick über Ausbildungskonzepte und Trainings, die die Entwicklung von Klassenmanagement zum Gegenstand haben, schließt das Kapitel ab (Kapitel 1.4.3).

1.4.1. Wissenstypen, mentale Repräsentation und Phasen der Kompetenzgenese

Die in Kapitel 1.2.1 dargestellten Anforderungen des Klassenmanagements (u. a. Multidimensionalität, Simultaneität und Unmittelbarkeit) werden durch das Zusammenspiel von Wissen, Können und Orientierungen bearbeitet bzw. bewältigt (vgl. Kapitel 1.1.1). Derzeitige Auffassungen und Befunde über die mentale Repräsentation sowie die Genese von Wissen und Können stammen vor allem aus der kognitionspsychologischen Expertenforschung und ihrer Anwendung auf professionelle Berufe (Baumert & Kunter, 2006; Bromme, 1992). In diversen theoretischen Ansätzen wurden unterschiedliche Konzepte und Begriffe zur Bezeichnung von Wissenstypen entwickelt, darüber hinaus ist ihre Verwendung keineswegs einheitlich (vgl. Fenstermacher, 1994; Grossman, 1995; Munby, Russell & Martin, 2001; Renkl, 2009; Schraw, 2006). Im Folgenden wird daher versucht eine Auswahl und Integration der relevanten Konzepte vorzunehmen. Allen theoretischen und empirischen Ansätzen gemeinsam ist die Annahme, dass es *unterschiedliche* Wissenstypen gibt, die spezifische Eigenschaften aufweisen. Dazu gehört zunächst explizierbares Faktenwissen—deklaratives Wissen (unterteilt in formales und episodisches Wissen, Schraw, 2006) bzw. *formal knowledge* (Fenstermacher, 1994), von dem man annimmt, dass es propositional repräsentiert ist (in semantischen Netzwerken oder mentalen Modellen, Brophy, 1988; Schraw, 2006; Shulman, 1986). Dazu gehört in Bezug auf Klassenmanagement explizierbares Wissen darüber, wie Regeln eingeführt werden, welche Prozeduren Übergänge stützen können oder wie Zurechtweisungen sukzessive gesteigert werden.

Gegenüber Fach- und Fachdidaktischem Wissen beruhen Kompetenzen im Klassenmanagement jedoch nur zu verhältnismäßig geringen Anteilen auf deklarativem Wissen (Stough, 2006; Weinert et al., 1990; Westerman, 1991). Vielmehr handelt es sich um „hochgradig prozeduralisierte Wissensbestände“ (Weinert et al., 1990, S. 190) – also prozedurales Wissen nach Schraw (2006). Prozedurales Wissen wird in der Handlungspraxis aufgebaut und weiterentwickelt (Emmer et al., 2003; Weinert et al., 1990) und im Langzeitgedächtnis in Form von Skripts, Heuristiken und (seltener) auch Algorithmen gespeichert (Schraw, 2006). Im Handlungsvollzug sind diese prozeduralisierten Wissensbestände an die professionellen Überzeugungen, Werte und Motive von Lehrkräften gekoppelt (Bromme & Tilema, 1995; Gholami & Husu, 2010), welche u. a. die Wahrnehmung vom unterrichtlichen Geschehen beeinflussen (Brophy, 1996; Martin, 2004; Martin et al., 1998) und bei der Aneignung neuen Wissens häufig eine Filterfunktion übernehmen (Hammerness et al., 2005; Jones, 2006). Für dieses Zusammenspiel aus deklarativem und prozeduralem Wissen, Überzeugungen und Motiven (die in Kapitel 1.1.1 beschriebene professionelle Handlungskompetenz) führte Fenstermacher (1994) den Begriff des *practical knowledge* ein (vgl. auch Gholami & Husu, 2010; van Tartwijk et al., 2009). Dieses ist handlungsorientiert, erfahrungsbasiert und in spezifische Kontexte eingebettet; im Handlungsvollzug bleibt es eher intuitiv und implizit (Berliner, 2004; Bromme, 1992; Brophy, 1988; Leinhardt, Young & Merriman, 1995). Erworbenes *practical knowledge* ist somit nicht ohne weiteres auf andere Kontexte übertragbar.¹³

Brophy (1988) benennt neben deklarativem und prozeduralem einen dritten Wissenstypus: konditionales Wissen. Konditionales Wissen ermöglicht, begründet aus einem breiten Repertoire von Handlungsoptionen adaptiv und kontextsensitiv auszuwählen. Damit können Strategien differenziell eingesetzt werden, so dass z. B. Zurechtweisungen der aktuell vorherrschenden Aktivitätsstruktur, dem einzelnen Schüler bzw. der einzelnen Schülerin sowie der Stärke des Handlungsvektors angepasst sind. Konditionales Wissen ist somit den Konzepten des strategischen (Baartman & de Bruijn, 2011; Baumert & Kunter, 2006; Shulman, 1987) bzw. metakognitiven Wissens ähnlich (Hammerness et al., 2005; Renkl, 2009; Schraw, 2006).

Neben der Frage der Eigenschaften unterschiedlicher Wissenstypen ist für ein theoretisches Verständnis von Kompetenzentwicklung jedoch auch relevant, wie sich die spezifische Qualität von Expertenwissen manifestiert. So zeigen Ergebnisse der Expertenforschung, dass

¹³ Daher ist die Unterscheidung von deklarativem vs. prozeduralem Wissen (Schraw, 2006) nicht gleichzusetzen mit der Unterscheidung von formal knowledge und practical knowledge (Fenstermacher, 1994). Während sich deklaratives Wissen und prozedurales Wissen qualitativ hinsichtlich ihrer mentalen Repräsentation unterscheiden, unterscheiden sich formales und praktisches Wissen insbesondere darin, wie es erworben wird. So enthält praktisches Wissen neben prozeduralen auch deklarative Wissensbestände, dieses wird aber – gegenüber formalem Wissen – durch praktische Erfahrung erworben, ebenso wie prozedurales Wissen. Der Erwerb formalen Wissens wird demgegenüber durch systematische Lerngelegenheiten initiiert (Fenstermacher, 1994).

Experten im Vergleich zu Novizen über ein umfangreicheres sachlich richtiges Wissen verfügen, das tiefer und reichhaltiger mental repräsentiert, sehr gut vernetzt und hierarchisch organisiert ist (Berliner, 1994, 2001; Bromme, 1992). Es gelingt ihnen, verschiedene Wissensdomänen bzw. Kompetenzbereiche (z. B. Fachwissen und erziehungswissenschaftliches Wissen) zu integrieren und sie sind eher in der Lage, ihr Handeln flexibel und variantenreich auf verschiedene Anforderungssituationen zu adaptieren.

Damit stellt sich als nächstes die Frage, wie Novizen zu Experten werden bzw. werden können (vgl. dazu im Folgenden Ophardt & Thiel, 2013). Grundsätzlich wird kein linearer, sondern ein kurvilinearere Zusammenhang zwischen Dauer der Berufstätigkeit und der Kompetenzentwicklung angenommen (Berliner, 1994; Weinert et al., 1990), also ein stärkerer Kompetenzzuwachs in der Ausbildungs- und Berufseingangsphase. Die Entwicklung professioneller Kompetenz ist abhängig von den persönlichen Voraussetzungen von Lehrkräften (z. B. Berufswahlmotive oder Enthusiasmus), dem Angebot und der Nutzung von Lerngelegenheiten sowie den institutionellen Rahmenbedingungen (vgl. Modell der Determinanten und Konsequenzen der professionellen Kompetenz von Lehrkräften, Kunter, Kleickmann, Klusmann & Richter, 2011 sowie Messner & Reusser, 2000). Nach Berliner (1994) gibt es jedoch mindestens drei, im Idealfall fünf charakteristische Stufen, die Lehrkräfte im Verlauf ihrer beruflichen Entwicklung durchlaufen. In seinem an Dreyfus und Dreyfus (1986) orientierten Stufenmodell zur Expertiseentwicklung differenziert er folgende Phasen: (1) *novice*, (2) *advanced beginner*, (3) *competent*, (4) *proficient* und (5) *expert*.

1. Im Novizen-Stadium (*novice level*) erwerben die Lernenden zunächst verallgemeinerbare und im Idealfall evidenzbasierte Konzepte, Regeln und Prinzipien, die noch ohne kontextuellen oder situativen Bezug sind (z. B. Merkmale effektiven Lobs), also deklaratives Wissen. Diese noch kontextfreien Prinzipien werden eher rational und relativ unflexibel umgesetzt. Typischerweise sind Studierende sowie Lehrkräfte in ihrem ersten Berufsjahr in diesem Stadium.
2. Fortgeschrittene Anfänger (*advanced beginner level*) beginnen, praktische Erfahrungen mit Konzept- und Regelwissen zu verknüpfen, episodisches und fallspezifisches Wissen sowie konditionales bzw. strategisches Wissen aufzubauen (vgl. auch Shulman, 1986). Diese Integration erfolgt jedoch eher implizit und bleibt unreflektiert; die Auswahl von Handlungsalternativen ist noch selten zielgerichtet und aktiv gesteuert. Typischerweise sind Lehrkräfte im zweiten und dritten Berufsjahr in diesem Stadium.

3. Erst im Kompetenz-Stadium (competent level) beginnen Lehrkräfte Handlungsalternativen bewusst auszuwählen, Prioritäten zu setzen und Abläufe zu planen. Außerdem können sie Wichtiges von Unwichtigem unterscheiden und es gelingt ihnen, das Handlungsprogramm an den Kontext und die Bedürfnisse der Schülerinnen und Schüler zu adaptieren. Dieses Stadium wird nach Berliner ab dem dritten Berufsjahr, jedoch nicht von allen Lehrkräften erreicht.
4. Flüssig, schnell, flexibel und intuitiver wird das Handeln erst auf der vierten Stufe (proficient level), nach ungefähr fünf Jahren Berufserfahrung. Es bildet sich eine holistische Wahrnehmung heraus, so dass Ähnlichkeiten zwischen Situationen erkannt und damit auch Entwicklungen des Unterrichts besser vorhergesagt werden können. Dennoch entscheiden Lehrkräfte auf dieser Stufe noch immer analytisch und bewusst, wie sie sich verhalten.
5. Erst im letzten Stadium spricht Berliner von Experten (expert level). Experten erfassen die zentralen Merkmale intuitiv und reagieren unbewusst und ohne vorhergehende Analysen („arational“, S. 8). Das Handeln erscheint flüssig und unangestrengt. Nur wenn Probleme auftauchen handeln Experten analytisch, rational und bewusst.

Bereits das Kompetenz-Stadium erreichen nicht alle Lehrkräfte und nur sehr wenige erreichen das Stadium des Experten. Die Entwicklung von Expertise erfolgt also nicht automatisch mit zunehmender Berufserfahrung (Richardson & Placier, 2001), sondern in aktiver und reflektierter Auseinandersetzung mit dem eigenen Beruf, dem beruflichen Können und der professionellen Identität (vgl. Berliners Konzept der *adaptive expertise*, 2004, Ericssons Konzept der *deliberate practice*, 2003 sowie Schöns Konzept des *reflective practitioner*, 1983).

Die von Berliner beschriebenen Phasen deuten auf eine zunehmende Integration einzelner Wissenskomponenten hin (vgl. Baer et al., 2009). Wie diese Integration gelingt, also Kompetenzen durch spezifische Lerngelegenheiten entwickelt werden können, wird in der vorliegenden Arbeit heuristisch anhand des Stufenmodells von Schraw (2006, „stages of skill acquisition“, S. 246) ausgeführt. Schraw geht von drei allgemeineren Phasen aus, die noch expliziter an wissenspsychologischen Merkmalen orientiert sind:

1. Akkumulation von Wissen (Erwerb von formalem, deklarativem Wissen)
2. Integration von Wissen (Organisation des Wissens in prozeduralisierten Schemata und Skripts)
3. Automatisierung und Tuning (Automatisierung von Routinen und Anpassung der Routinen an neue Anforderungen).

Beim Erwerb deklarativen Wissens werden zunächst relativ isolierte Wissenssegmente gespeichert, die erst im zweiten Stadium in größere konzeptuelle *chunks* (Schemata und Skripts) oder noch größere Einheiten für komplexere Prozesse transformiert werden (Schraw, 2006). Dementsprechend kann die benötigte Zeit bis zur vollständigen Integration in Abhängigkeit des zu Erlernenden (z. B. Auswahl einer Regel zum Arbeitsverhalten vs. Etablierung eines Regelsystems) sehr unterschiedlich sein und auch mehrere Jahre in Anspruch nehmen. Mit zunehmender Praxiserfahrung können diese Schemata flüssiger, effektiver und schneller abgerufen und in konkrete Handlungen umgesetzt werden, wenngleich selbst sehr fortgeschrittene Lehrkräfte aufgrund ständiger Veränderungen der situativen und kontextuellen Rahmenbedingungen nie ausschließlich intuitiv und automatisiert handeln (vgl. auch Ericson, 2003).

1.4.2. Lerngelegenheiten zur Entwicklung von Kompetenzen im Klassenmanagement

Die formale Lehreraus- und Weiterbildung hat zum Ziel, dass angehende und Junglehrkräfte früher (z. B. vom novice level zum advanced beginner level) und praktizierende Lehrkräfte mit größerer Wahrscheinlichkeit eine höhere Kompetenzstufe (z. B. proficient oder expert level) erreichen sowie im Verlaufe der Berufsausübung neue Entwicklungen und wissenschaftliche Erkenntnisse in ihrer beruflichen Tätigkeit berücksichtigen und nutzen können (Kunter et al., 2011; Müller, 2010). Da Klassenmanagement wenig explizierbares Handlungswissen umfasst, gilt es gegenüber Fach- und Fachdidaktischem Wissen jedoch als schwerer lehrbar (Ophardt & Thiel, 2013; Reinke, Sprick & Knight, 2009). Zentral für die Ausbildung von Kompetenzen im Klassenmanagement sind daher angemessene (d. h. kompetenzspezifische und zielgruppenadäquate) Lerngelegenheiten, insbesondere praxisbezogene Lehr-Lernarrangements (Grossman, 1995; LePage et al., 2005). Nachfolgend werden, orientiert am dreistufigen Modell der Kompetenzentwicklung nach Schraw (2006, vgl. Kapitel 1.4.1), potentiell erfolgreiche Lerngelegenheiten vorgestellt, die den Erwerb von Kompetenzen im Klassenmanagement in spezifischer Weise unterstützen können. Dabei wird der Entwicklungsprozess auf jeder Stufe nach Erfahrungsstufe bzw. Zielgruppen (Studierende vs. Lehrkräfte) differenziert.

Akkumulation von Wissen

Beim Erwerb deklarativen Wissens—sowohl bei erfahrenen, als auch bei angehenden Lehrkräften—sind zunächst zwei Aspekte zu berücksichtigen. Erstens sollte die Vermittlung deklarativen Wissens auf einer wissenschaftlich fundierten Wissensbasis (*knowledge base*) beruhen (Brophy, 1988; Munby et al., 2001; Wang et al., 1993), über die ein Konsens besteht.

Diese ist nicht immer gegeben und auch wissenschaftlichen Moden unterworfen: Der Umgang mit störendem Schülerverhalten wird zunehmend schülerorientierter gedacht und bezieht sich nicht mehr vorrangig auf die lehrerseitige Kontrolle des Schülerverhaltens. Zudem kann gerade im Bereich Klassenmanagement–insbesondere auf molarer Ebene (wie viele Füllwörter darf ein Arbeitsauftrag enthalten, um verständlich zu sein; sind „Wir“-Regeln besser als „Ich“-Regeln)–kaum experimentell geprüft werden, welche Strategien die größte Wirksamkeit aufweisen, da sich nur schwer standardisierte Untersuchungsbedingungen herstellen lassen, die gleichzeitig authentische Praxis abbilden (Brophy, 2006). Im Bereich des Klassenmanagements ist es daher zweckmäßiger, leitende Prinzipien statt vermeintlich richtiger Lösungen zu vermitteln (z. B. durch ein Konstrukt wie den Handlungsvektor oder die proaktive Funktion von Monitoring). Dies ermöglicht eine heuristische und regelgeleitete, nicht regelgebundene Praxis (Messner & Reusser, 2000).

Zweitens muss berücksichtigt werden, dass nicht nur erfahrene, sondern auch angehende Lehrkräfte bereits über unterrichtsbezogenes Wissen und Überzeugungen verfügen (Hammerness et al., 2005; Korthagen, Loughran & Russell, 2006; Richardson & Placier, 2001). Konzeptuelle Veränderungen und eine Erweiterung der Wissensbasis werden demnach nur dann im Unterricht handlungsleitend, wenn Vorwissen und domänenspezifische Überzeugungen berücksichtigt und explizit thematisiert werden (Brophy, 1988; Korthagen et al., 2006, Larrivee, 2006).

Formen der direkten Instruktion und des problembasierten Lernens eignen sich in besonderem Maße, empirische Befunde aufzuarbeiten sowie theoretische Konzepte zu veranschaulichen (im Sinne externer Steuerung nach Glaser, 1996, zitiert nach Ophardt & Thiel, 2013, S. 132, vgl. auch Brophy, 1988; Ophardt & Thiel, 2013). Dabei können schriftliche oder videografierte Fallbeispiele eingesetzt werden, um Standardsituationen vergleichend darzustellen oder *best practice* zu demonstrieren (Borko, Jacobs, Eiteljorg & Pittman, 2008; Krammer & Reusser, 2005). Gleichzeitig ist die Analyse von Fallvignetten geeignet, die Beobachtung und Beurteilung von Unterrichtsbeispielen zu trainieren (Choi & Lee, 2009; Gold et al., 2013; Seidel, Blomberg & Renkl, 2013). Im Rahmen der Diskussion verschiedener, kontrastiver Fälle können stellvertretende Devianzerfahrungen gesammelt, die wichtig sind, um Überzeugungen zu reflektieren und zu verändern (Bauer & Gruber, 2007; Hammerness et al., 2005). Während bei Studierenden im Vordergrund steht, eine breite Wissensbasis anzulegen, wird bei Lehrkräften neues Wissens direkt mit vorhandenem Wissen in Beziehung gesetzt (Ophardt & Thiel, 2013).

Integration von Wissen

Bromme und Tilema (1995) betonen, dass zunehmende Professionalisierung nicht bedeutet, das durch Erfahrung erworbene Wissen durch theoretisches Wissen zu ersetzen, sondern beides miteinander zu „vereinen“ (fusing theory and experience, S. 266), neues Wissen also in die eigene Erfahrung zu integrieren (vgl. auch Leinhardt et al., 1995). Es müssen daher explizit Verbindungen zwischen deklarativem, konditionalem und prozeduralem Wissen sowie zwischen den verschiedenen Wissensdomänen hergestellt werden. Dies gelingt durch eine intensive Auseinandersetzung mit konkreten praktischen Anforderungen, z. B. durch in ein Lernsetting eingebettete Fallarbeit und Simulationsübungen (wie Rollenspiele oder Microteaching, vgl. Darling-Hammond, Hammerness, Grossman, Rust & Shulman, 2005; Jones, 2006; Klinzig, 2002; Lüders, 2010) sowie bei Studierenden im Rahmen von intensiv begleiteten Unterrichtspraktika, z. B. durch Fachspezifisch-Pädagogisches Coaching (Staub, 2004; West & Staub, 2003).

Sowohl Studierenden als auch Lehrkräften helfen Simulationsübungen, im Rahmen komplexitätsreduzierter, nachgestellter Szenen alternative Strategien zielgerichtet und bewusst zu erproben oder in einer Fremdperspektive zu erleben. Während dies bei Lehramtsanwärtern vor allem dazu dient, die Wissensintegration anzubahnen, unterstützen solche Übungen bei erfahrenen Lehrkräften die Explikation vorhandenen Wissens und damit die Integration in bereits vorhandene Schemata. Da es erfahrenen Lehrkräften oft schwer fällt, eine andere als die eigene Perspektive einzunehmen (Schraw, 2006), kann dies durch das Einnehmen anderer Rollen (i. d. R. Schülerinnen und Schüler, Beobachter) ebenfalls geübt werden. Schließlich unterstützen solche Übungen Feedbackgeben und -nehmen zu erproben sowie eine Feedbackkultur zu etablieren, die auch im Rahmen kollegialer Hospitationen eine zentrale Funktion erfüllt (Fukink, Trienekens & Kramer, 2011; Keuffer, 2010).

Demgegenüber dient die (vor allem videobasierte) Analyse und Reflexion authentischer Fälle besonders dazu, die Wahrnehmung von Tiefenstrukturen im Unterricht zu üben (Studierende) bzw. weiterzuentwickeln (Lehrkräfte), (neues) Fallwissen aufzubauen, bestehendes und erworbenes Wissen aufeinander zu beziehen sowie durch die Entwicklung von Alternativen ein breites Handlungsrepertoire zu erwerben (Studierende) bzw. ein bereits vorhandenes zu adaptieren (Lehrkräfte, vgl. Borko et al., 2005; Goldman, Pea, Barron & Denny, 2007; Krammer & Reusser, 2005; Seidel et al., 2013). Erfolgreiche Konzepte hierzu liegen bereits vor (Goeze, Hetfleisch & Schrader, 2013; Han, Eom & Shin, 2013).

Den genannten Lerngelegenheiten gemeinsam ist, dass sie die die Verknüpfung von propositionalem, fallbezogenem und strategischem Wissen unterstützen (Krammer, Razka,

Klieme, Lipowsky, Pauli & Reusser, 2006) und sowohl Studierenden als auch Lehrkräften ermöglicht, unterrichtliches Handeln systematisch zu reflektieren (Adler, 1991; Hatton & Smith, 1995; Korthagen et al., 2006; Schön, 1983). Reflexion hat eine zentrale Bedeutung für die kontinuierliche, problemorientierte und erfolgreiche Weiterentwicklung und ist entscheidende Voraussetzung für die nächste Stufe, Automatisierung und Tuning.

Automatisierung und Tuning

Diese höchste Stufe ist gleichzeitig jene, die durch außerunterrichtliche Lernarrangements am schwierigsten unterstützt werden kann und für Studierende noch kein Entwicklungsziel darstellt (vgl. Terhart 2002a). Die besondere Herausforderung für Lehrkräfte besteht darin, einen situations- und kontextadäquaten Transfer (Baartman & Bruijn, 2011; Messner & Reusser, 2000; Renkl, 2009) aus den vorangegangenen Stufen zu bewerkstelligen, die erlernten Routinen an neue Anforderungen anzupassen, gleichzeitig aber auch intuitiver zu handeln. Dieser Prozess kann insbesondere durch längerfristige und praxisbegleitende Lerngelegenheiten unterstützt werden, wie professionelle bzw. kollaborative Lerngemeinschaften (Vescio, Ross & Adams, 2008), Qualitätszirkel (Lovett & Gilmore, 2003), Video Clubs (van Es & Sherin, 2006, 2010) oder eine Kombination daraus (Gärtner, 2008). Dabei steht vor allem der eigene (häufig videografierte) Unterricht im Zentrum der Analyse und Reflexion, die im kollegialen Austausch stattfindet und unterschiedlich stark durch Moderatoren oder Coaches strukturiert und begleitet wird. Erst wenn die professionelle Kompetenz reflexiv eingeholt werden kann hat sie die für erfolgreiches Unterrichten notwendige Flexibilität und Entwicklungsoffenheit (Terhart, 2007).

1.4.3. Trainings zur Entwicklung von Kompetenzen im Klassenmanagement

Nachdem ein Überblick über potentielle Lerngelegenheiten gegeben wurde, die für verschiedene Phasen der Kompetenzentwicklung besonders geeignet sind, werden nun vorhandene Trainings und Ausbildungsformen dargestellt, die spezifisch auf Klassenmanagement fokussiert sind.

Entsprechend der Darstellung der Kompetenzentwicklung nach Schraw (2006) sind dabei für Studierende vor allem die Akkumulation und Integration von Wissen zentrale Entwicklungsziele; möglicherweise finden erste Automatisierungsprozesse im Rahmen von Praktikumsphasen statt. Demgegenüber liegt der Fokus bei erfahrenen Lehrkräften vor allem auf der dritten Stufe: Automatisierung und Anpassung an neue Anforderungen, sowie Entwicklung adaptiver Kompetenzen.

Ausbildung an der Universität und im Referendariat

Die deutsche Lehrerausbildung wurde in den letzten Jahren zunehmend kritisiert, vor allem der defizitäre Praxisbezug (Havers & Toepell, 2002; Korthagen et al., 2006), der Mangel an partizipativen und forschungsbasierten Lehrmethoden sowie der Erwerb reflexionsbezogener Kompetenzen (Felbrich, Müller & Blömeke, 2008). So resümiert Oser (1997b), dass für viele Standards keine Zugänge in der Ausbildung bzw. kein einheitliches Curriculum institutioneller Lerngelegenheiten vorliegen (vgl. auch Messner & Reusser, 2000; Terhart, 2012). Dies gilt auch für den Bereich Klassenmanagement (Helmke, 2007a; Kunter & Baumert, 2011; Ophardt & Thiel, 2008; Schneider & Bodensohn, 2010).

Havers (1998, 2006) entwickelte im deutschsprachigen Raum als erster ein spezifisch auf die Bedürfnisse von Studierenden kurz vor Abschluss ihres Studiums abgestimmtes Training, das *Münchener Lehrertraining*. Dieses ist umfassend angelegt und bezieht sich nicht nur auf den Umgang mit problematischem Schülerverhalten sondern auch auf die Elternarbeit und Techniken zur Stressreduktion. Im Rahmen eines einwöchigen Workshops kommen dabei Rollenspiele, Videoanalysen dieser Rollenspiele sowie sogenannten Demonstrationsexperimente zum Einsatz; die Akzeptanz des Trainings war hoch und die ehemaligen Studierenden berichteten positive Effekte auf ihr Lehrerverhalten kurz nach dem Einstieg in den Beruf (Havers, 2006). Ein semesterbegleitendes Training für Studierende des Grundschullehramts von Gold und Kollegen (2013) fokussiert die professionelle Wahrnehmung anhand fremden Unterrichtsmaterials im Hinblick auf klassenführungsrelevante Unterrichtsereignisse. In einem experimentellen Design konnten Effekte auf die Wahrnehmung von Strukturierung und Gruppenfokus nachgewiesen werden. Auch das Fachspezifisch-Pädagogische Coaching (Staub, 2004), welches an Praxisphasen im Rahmen des Studiums gebunden ist, hat–indirekt über die Schulung fachdidaktischer Kompetenzen–positive Effekte auf die Prävention von Unterrichtsstörungen gezeigt (Wettstein, 2010).

Fort- und Weiterbildungsangebote

Damit Lehrkräfte mit Eintritt in das Berufsleben berufsbegleitend Lernen und ihre Praxis kontinuierlich weiterentwickeln können sind sie auf Fort- und Weiterbildungsangebote angewiesen. Die Qualität dieser Angebote wurde in den letzten Jahren zunehmend kritisiert, insbesondere deren geringe Wirksamkeit, z. B. aufgrund reiner Wissensvermittlung und mangelnder Bedarfsorientierung (Lipowsky, 2010; Messner & Reusser, 2000; Wettstein, 2010) sowie einem Mangel an belastbaren Wirksamkeitsnachweisen (Lipowsky, 2010; Müller, 2010).

Insbesondere in den USA existiert bereits eine große Bandbreite von Trainings und Schulprogrammen für Lehrkräfte, die direkt oder indirekt Klassenmanagement zum Gegenstand haben (einen Überblick geben u.a. Emmer & Aussiker, 1989; Freiberg & Lapointe, 2006; Hardin, 2008; Weinstein, 1999). Zu den prominentesten und empirisch am besten untersuchten gehören COMP (*Classroom Organization and Management Program*, Evertson & Harris, 1999), CMCD (*Consistency Management & Cooperative Discipline*, Freiberg, Stein & Huang, 1995), *Incredible Years* (Menting, Orobio de Castro & Matthys, 2013; Webster-Stratton, Reid & Hammond, 2001) und Three Cs (*Cooperative Community, Constructive Conflict Resolution and Civic Values*, Johnson & Johnson, 2000). Darüber hinaus existieren auf spezifische Aspekte fokussierte Trainings, unter anderem zur Störungskontrolle, zum Umgang mit Konflikten, zur Bildung positiver Beziehungen (vgl. LePage et al., 2005) und zum effektiven Einsatz von Lob (Kubany, Sloggett & Ogata, 1974) sowie spezielle Coachingansätze (Reinke et al., 2009; Roelofs, Veenman & Raemaekers, 1994). Diese Angebote unterscheiden sich dabei vor allem hinsichtlich der implementierten Lerngelegenheiten, des zeitlichen Umfangs (wenige Stunden bis mehrere Schuljahre), der strukturellen Einbettung (Schulkontext oder außerhalb der Schule) und des Wirksamkeitsnachweises (vgl. Freiberg & Lapointe, 2006).

In Deutschland sind Fortbildungen in der Regel durch staatliche Fortbildungsinstitute organisiert und erfolgen berufsbegleitend (Richter, 2011); darüber hinausgehende externe Angebote sind nur zum Teil verfügbar und werden nicht immer formal anerkannt (z. B. in Form von Stundenreduktion, ebd.). Bislang werden fast keine Trainings für den Bereich Klassenmanagement angeboten, die systematisch dokumentiert oder deren Wirksamkeit empirisch nachgewiesen ist (vgl. Havers, 2010). Die Ausnahme bildet das KTM (*Konstanzer Trainingsmodell*, Tennstädt, 1987) in dem der Umgang mit problematischem Schülerverhalten fokussiert wird. In acht aufeinander folgenden Schritten reflektieren die Lehrkräfte in Gruppen oder Teams und unter Anleitung eigene Unterrichtssituationen mit dem Ziel, subjektive Theorien zu modifizieren und die Bewältigung von Störverhalten und aggressiven Handlungen der Schülerinnen und Schüler zu verringern. Es konnten Effekte auf das Störverhalten, das Lehrerverhalten und die Wirksamkeitsüberzeugungen von Lehrkräften identifiziert werden (Dann & Humpert, 2002).

1.5. Empirische Überprüfung von Kompetenzentwicklung

Nicht nur die Initiierung von Veränderung stellt eine anspruchsvolle Herausforderung dar. Es muss auch ein valider Nachweis gelingen, dass Veränderung auf eine spezifische In-

tervention zurückzuführen ist (vgl. Wilson & Lipsey, 2001). Es geht dabei zum einen um die Prüfung von Kausalhypothesen, die Schlussfolgerungen über Ursache-Wirkungszusammenhänge erlauben (interne Validität), um die Verallgemeinerbarkeit gefundener Ergebnisse (externe Validität) aber auch um die Zulässigkeit und Güte der eingesetzten diagnostischen und statistischen Verfahren (statistische Validität). Die genannten drei Validitätsaspekte sind essentiell für den Wirksamkeitsnachweis von Interventionsmaßnahmen und bedingen sich gegenseitig.

Es soll daher in diesem letzten einleitenden Kapitel zunächst darum gehen, wie interne und externe Validität durch konkrete versuchsplanerische Maßnahmen forschungsstrategisch optimiert werden können (Kapitel 1.5.1) und wie die statistische Validität durch angemessene evaluative Verfahren zur Auswertung der erhobener Daten sichergestellt werden kann (Kapitel 1.5.2).

1.5.1. Interne und externe Validität

Die interne und externe Validität stellen ein zentrales Gütemerkmal empirischer Untersuchungen dar (vgl. Aiken & West, 1990; Bortz & Döring, 2006; Gollwitzer & Jäger, 2009). Kausalinterpretationen sind dann zulässig, wenn die Veränderung in der abhängigen Variable auf die Variation der unabhängigen Variable zurückgeführt werden kann, ohne dass mögliche Alternativerklärungen existieren (Eid, 2003). Diese strenge Kausalitätsbedingung ist im Bereich der empirischen Sozialforschung quasi nie erfüllt (Steyer, 1992)–allerdings stehen diverse methodische und statistische Strategien zur Verfügung, um ein möglichst hohes Niveau interner Validität zu gewährleisten.

Bereits in der Konzeptualisierungsphase von Evaluationsstudien können dafür günstige Bedingungen geschaffen werden (McKnight & Sechrest, 2004; Rost, 2000): Neben theoretisch-konzeptuellen sind auch versuchsplanerische Überlegungen notwendig, die eine Falsifikation von Hypothesen gewährleisten, also den empirischen Gehalt dieser Annahmen betreffen (Popper, 1968).¹⁴ Für den Wirksamkeitsnachweis müssen darüber hinaus Untersuchungsdesigns konzipiert werden, die systematische Konfundierungen von Störvariablen mit der abhängigen Variablen entweder nicht zulassen oder diese zu kontrollieren erlauben. Dies erfordert die Implementation experimenteller Designs und ein aussagekräftiges Design zur Ver-

¹⁴ Dass ein Training etwas bewirkt, kann so gut wie nicht widerlegt werden, da es unwahrscheinlich ist, dass Teilnehmer nicht–wenn auch nur für kurze Zeit–z. B. Freude, Ärger oder Bestätigung empfinden. Es muss also spezifiziert werden, welche Veränderungen (die gleichzeitig praktische Relevanz aufweisen) konkret erwartet werden und Überlegungen dazu angestellt werden, wie wahrscheinlich es ist, dass *keine* dieser spezifizierten Veränderungen eintreten. Ist dies unwahrscheinlich, ist der empirische Gehalt der aufgestellten Hypothesen gering.

änderungsmessung (direkte, indirekte und quasi-indirekte Veränderungsmessung, Gollwitzer & Jäger, 2009).

Experimentelle Designs, bei denen Teilnehmende Interventions- und Kontrollbedingung(en) zugeordnet werden, gelten als Standard in der Trainingsevaluation (Bortz & Döring, 2006; Köller, 2009; Rost, 2000). Nur durch den Vergleich zu einer geeigneten Kontrollgruppe (Wartegruppe, Placebo- oder Alternativtraining) lassen sich identifizierte Veränderungen auf die Maßnahme und nicht auf andere, möglicherweise nicht kontrollierte Merkmale zurückführen (z. B. natürliche Reifung oder Variabilität des Merkmals). Die randomisierte Zuordnung zu den Versuchsbedingungen gewährleistet dabei, dass potentiell relevante Eingangsmerkmale *zufällig* auf die Untersuchungsgruppen verteilt werden.

Da Interventionsstudien im Bereich der Lehreraus- und Weiterbildung in der Regel als Feldstudien durchgeführt werden, finden sich fast ausschließlich *quasi*-experimentelle Versuchsbedingungen (Brandtstätter, 1990; Fries & Souvignier, 2009; McKnight & Sechrest, 2004), d. h. die Zuordnung zu den Versuchsbedingungen erfolgt nicht randomisiert, sondern nach Präferenz (zur Vermeidung selektiven Dropouts) oder durch bereits vorgegebene Gruppen (wie Kollegien an Schulen oder Seminare an der Universität). Potentiell relevante Eingangsbedingungen bzw. Störvariablen sind dann unter Umständen nicht zufällig oder nicht gleichverteilt und können systematisch mit den abhängigen Variablen konfundieren (z. B. ist ein geringes Expertise-Niveau mit mehr Wissenszuwachs oder eine hohe Teilnahme-Motivation mit hoher Akzeptanz des Treatments assoziiert, vgl. Brandtstätter, 1990; Wottawa & Thierau, 1998).

Experimentelle Designs werden im Rahmen direkter, indirekter oder quasi-indirekter Veränderungsmessung realisiert (Gollwitzer & Jäger, 2009; Lam, 2003). Die direkte Veränderungsmessung erfolgt durch eine subjektive Einschätzung der Veränderung (z. B. über *Goal-Attainment Scaling*), ist jedoch durch den hohen Aufforderungscharakter anfällig für Verzerrungen wie Konfirmationseffekte (Veränderung wird so wahrgenommen, wie man sie erwartet hätte) oder Kontrasteffekte (Abweichungen werden falsch gewichtet, vgl. Elliot & Busse, 2004; Lam, 2003). Sie dient daher selten der Operationalisierung von Veränderung und gilt eher als Indikator für die subjektiv empfundene Kompetenzentwicklung, die durchaus auch von Bedeutung sein kann (z. B. hinsichtlich eines veränderten Belastungserlebens von Lehrkräften). Die indirekte Veränderungsmessung wird als geeignetstes Verfahren zur Programmevaluation angesehen (Aiken & West, 1990; Dimitrov & Rumrill, 2003; Gollwitzer & Jäger, 2009). Abhängige Variablen werden über mehrere Messzeitpunkte und relativ zum Ausgangszustand miteinander verglichen. Die Differenzwerte wiederholter Messungen quan-

tifizieren dabei das Ausmaß der Veränderung und kontrollieren gleichzeitig potentielle Unterschiede in den Ausgangsbedingungen.¹⁵ Obwohl indirekte Designs gegenüber direkten weniger anfällig für Verzerrungen oder Verfälschungen sind, kann auch die Vorher-Nachher Differenz unter Umständen nicht nur Effekte des Programms, sondern auch andere, von echter Veränderung unabhängige Komponenten abbilden (z. B. durch Testübung oder einen veränderten Bewertungsmaßstab). Gründe hierfür können methodisch und auch statistisch sein (wie selektive Dropouts oder einen Regressionseffekt, vgl. Köller, 2009; McKnight & Sechrest, 2004).¹⁶ Daher wurde als dritter Zugang die quasi-indirekte Veränderungsmessung entwickelt. Hierbei wird der Ausgangszustand retrospektiv, also nach der Teilnahme an der Intervention, erfasst (Aiken & West, 1990; Hill, 2005; Pratt, McGuigan & Katzev, 2000). Dieses Vorgehen kann dann indiziert sein, wenn Effekte der Prämessung auf das Treatment erwartet werden, beispielsweise eine gesteigerte Sensitivität gegenüber Fortbildungsinhalten (vgl. Aiken & West, 1990; Kim & Willson, 2010); häufiger jedoch geben erhebungslogische Gründe Anlass dazu, zum Beispiel ein fehlender Zugang zur Prämessung. Die Kritik an dem Verfahren betrifft auch hier potentiell unkontrollierbare Konfundierungen, da intendierte wie nicht intendierte Verzerrungen der Retrognose möglich sind (Henry, Moffitt, Caspi, Langley & Silva, 1994; Lam, 2003).

Die interne Validität ist eine notwendige, jedoch keine hinreichende Bedingung für die externe Validität von Untersuchungsergebnissen (Bortz & Döring, 2006; McKnight & Sechrest, 2004). Studienergebnisse können nur dann verallgemeinert werden, wenn auch bei Variation äquivalenter Stichproben (z. B. Lehrkräfte anderer Schulformen, Schulfächer oder Berufserfahrung), zulässiger statistischer Methoden (wie latente vs. manifeste Modelle) oder adäquater diagnostischer Instrumente (wie Tests und Ratings) die gleichen Ergebnisse resultieren.

1.5.2. Statistische Validität

Um auch statisch valide Ergebnisse zu erhalten (vgl. Hussy & Jain, 2002) sind eine sorgfältige Auswahl der eingesetzten Messinstrumente sowie eine begründete und auf Angemessenheit überprüfte Auswahl der gewählten statistischen Verfahren notwendig (Eid, 2003; Wottawa & Thierau, 1998). Die zu verwendenden Messinstrumente müssen dabei nicht nur den psychometrischen Gütekriterien genügen (vgl. Kapitel 1.3.3) sondern sollten so miteinan-

¹⁵ Können echte experimentelle Designs durch randomisierte Zuordnung realisiert werden ist die Kontrolle von Eingangsmerkmalen streng genommen nicht notwendig und ein Gruppenvergleich der abhängigen Variablen nach Teilnahme am Treatment ausreichend (Aiken & West, 1990).

¹⁶ Die Reliabilität von Differenzwerten ist dabei in der Regel geringer als die Reliabilität der einzelnen Messungen (vgl. Eid, 2003)

der kombiniert werden, dass multiple Evaluationskriterien angelegt werden können (multikriteriale Veränderungsmessung).¹⁷ Nach Kirkpatrick's (1998) Zielhierarchie sind dabei vier Ebenen bedeutsam (mit Beispielen für die Evaluation von Lehrerbildung, vgl. auch Desimone, 2009; Lipowsky, 2010):

1. *Reaktion*: Akzeptanz des Trainings sowie einzelner Trainingsbausteine (Inhalte, Organisation, Praxisrelevanz, Qualifikation der Durchführenden, auch durch kontinuierliche Prozessevaluation während der Intervention)
2. *Lernen*: Veränderungen im Wissen der Teilnehmenden
3. *Verhalten*: Veränderungen im Verhalten der Teilnehmenden (Performanz)
4. *Ergebnisse*: Veränderungen bei den Schülerinnen und Schülern

Diese Ebenen werden von Kirkpatrick (1998) hierarchisch und in stochastischer Abhängigkeit organisiert (vgl. auch Guskey, 2003): Eine positive Reaktion auf eine Intervention garantiert keinen Lernzuwachs, macht ihn jedoch wahrscheinlicher; eine Veränderung auf Verhaltensebene ist wahrscheinlicher, wenn auch kognitive Veränderungen stattgefunden haben, was wiederum eher zu Veränderungen bei den assoziierten Ergebnissen führt (vgl. Kapitel 1.4). Kirkpatrick's Zielhierarchie ermöglicht damit, das Ausmaß des Transfers der Trainingsinhalte zu quantifizieren (über die Zeit, Situationen und Anforderungen, vgl. Patry & Perrez, 2000).

Der Auswahl geeigneter Messinstrumente folgt die Auswahl geeigneter statistischer Verfahren. Grundlegend ist die Prüfung der Voraussetzungen für die Anwendung eines Verfahrens (Bortz & Döring, 2006). Störvariablen können ex post facto auch statistisch kontrolliert werden, sofern sie erhoben worden sind und die Teststärke bzw. Strichprobengröße ausreichend sind. Damit wird zwar idealiter die interne Validität erhöht, die Ergebnisse sind jedoch möglicherweise nicht oder nur noch eingeschränkt generalisierbar.

Als statistische Verfahren der Wahl werden insbesondere für die Veränderungsmessung zunehmend latente, also messfehlerbereinigte Verfahren diskutiert (Bollen & Curran, 2006; Duncan, Duncan & Strycker, 2006; Eid, 2003; Little, Bovaird & Slegers, 2005; Steyer, Schmitt & Eid, 1999), die eine größere Teststärke aufweisen und damit klinisch bedeutsame Effekte zuverlässiger identifizieren. Latente Modellierungen und Strukturgleichungsmodelle ermöglichen gegenüber den traditionell auf manifesten Variablen beruhenden varianz- und regressionsanalytischen Verfahren nicht nur die konfirmatorische Untersuchung komplexerer Wirkzusammenhänge, sondern auch die Überprüfung der Messinvarianz über die Zeit (Meredith, 1993; Vandenberg & Lance, 2000; Widaman & Reise, 1997). Dies ist vor allem für die

¹⁷ Die Auswahl erfolgt dabei nicht wahllos sondern auf der Grundlage eines wissenschaftlich fundierten Wirkungsmodells (Gollwitzer & Jäger, 2009; Mittag & Hager, 2000; Patry & Perrez, 2000).

Veränderungsmessung bedeutsam (Eid, 2003): Werden Prä- und Postwerte miteinander verglichen, muss geprüft werden, ob die jeweiligen Messungen zeitinvariant sind, also die Bedeutung des Konstrukts stabil und damit ein Vergleich verschiedener Messungen überhaupt zulässig ist. In der Terminologie der klassischen Testtheorie betrifft dies die „wahre Veränderung“ (Steyer, Eid & Schwenkmezger, 1997). Ändert sich die Bedeutung des Konstrukts, liegt ein *reponse shift* vor, der bei Nichtaufdeckung fälschlicherweise der wahren Veränderung zugeschrieben wird (Howard, 1980; Schwartz & Sprangers, 1999; Vandenberg & Lance, 2000). Mit Hilfe neuerer Verfahren ist die Identifikation und statistische Kontrolle solcher Verzerrungen jedoch zunehmend möglich (Gandhi, Ried, Huang, Kimberlin & Kauf, 2013; Oort, 2005). Neben methodischen Berührungängsten tragen sicherlich auch die oft zu geringen Fallzahlen dazu bei, dass diesem Desiderat bislang selten–mit Ausnahme methodisch orientierter Publikationen oder in *large-scale assessments* großer Schulleistungstudien–nachgekommen wird bzw. nachgekommen werden kann (Vandenberg & Lance, 2000).

1.6. Zusammenfassung und Ableitung der Forschungsfragen

Kompetenzen im Klassenmanagement haben prädiktive Validität für den effektiven Umgang mit Schülerverhalten, die schülergerechte Vermittlung von Lerninhalten und das berufliche Wohlbefinden von Lehrkräften. Es wurde dargestellt, dass sehr unterschiedliche Theorietraditionen Klassenmanagement untersucht und begrifflich gefasst haben. So lag bislang im deutschsprachigen Raum kein kohärentes Modell zur Systematisierung relevanter Klassenmanagementstrategien vor, das den aktuellen Forschungsstand zum Klassenmanagement angemessen widerspiegelt.

Da Kompetenzen im Klassenmanagement nur teilweise propositional repräsentiert sind erschließen sich diese vor allem in bestimmten Anforderungssituationen, auf der Ebene der Performanz. Zur fragebogengestützten *Erfassung* dieser Kompetenzen muss darüber hinaus ein Verständnis davon entwickelt werden, wie Lehrerhandeln qualitativ bewertet wird–im Sinne von Standards. Die Fundierung dieser Standards wurde in Kapitel 1.2 erbracht; es wurde jedoch auch aufgezeigt, dass die mehrdimensionale und psychometrisch hochwertige Operationalisierung von Klassenmanagement in den vorgestellten Fragebogenverfahren nur unzureichend umgesetzt werden konnte (Kapitel 1.3). Vielmehr muss konstatiert werden, dass sich die meisten Instrumente durch die zum Teil mangelhafte theoretische Fundierung und fehlende Pilotierungsphasen mitunter noch im Stadium der Itemanalyse befinden. Dementsprechend wurde dargelegt, welche Schritte der Fragebogenkonstruktion bei der Entwicklung neuer In-

strumente durchlaufen werden müssen, damit sie der Diagnose von Lehrerkompetenzen dienlich sind.

Die Analyse der *professionellen Entwicklung* von Kompetenzen orientiert sich–theoretisch–an der Erreichung von Kompetenzniveaus sowie–empirisch–an der erfolgreichen Bewältigung beruflicher Anforderungen (Kapitel 1.4). Auf der Grundlage eines Überblicks über Merkmalen wirksamer Lerngelegenheiten und Lernarrangements (Trainings) zur Entwicklung von Kompetenzen im Klassenmanagement wurde herausgearbeitet, dass noch zu wenig wirksame Angebote vorliegen, insbesondere im Bereich der Fort- und Weiterbildung.

Schließlich wurden die *methodischen Herausforderungen* bei der Erbringung von Wirksamkeitsnachweisen skizziert, in dessen Zentrum die Analyse von Ursache-Wirkungs-Prinzipien steht (Kapitel 1.5). Da die Wirksamkeitsprüfung von Trainings in Aus- und Weiterbildung üblicherweise im Feld ohne echte experimentelle Designs, mit kleinen Stichproben und begrenzten finanziellen wie personellen Ressourcen erfolgt, können die methodischen Anforderungen an die Evaluation von Interventionen oft nur eingeschränkt realisiert werden. Aus den genannten Gründen kommen häufig Selbsteinschätzungsverfahren zum Einsatz, die die interne Validität von Prä-Post Designs auf besondere Weise einschränken. Dies ist dann der Fall, wenn die über Selbstbericht erfasste abhängige Variable vor und nach der Intervention nicht Vergleichbares misst, z. B. aufgrund von Response Shift.

Mit Bezug auf die eingangs formulierten Fragestellungen konzentrieren sich die im Rahmen der vorliegenden Arbeit durchgeführten Studien damit noch auf drei Aspekte:

- Wie können Kompetenzen im Klassenmanagement erfasst werden (Teilstudie 1)?
- Wie können Kompetenzen im Klassenmanagement entwickelt werden (Teilstudie 2)?
- Wie können Kompetenzveränderungen valide nachgewiesen werden (Teilstudie 3)?

1.6.1. Teilstudie 1: Erfassung von Klassenmanagement

In Teilstudie 1 (Multidimensionale Erfassung von Kompetenzen im Klassenmanagement. Konstruktion und Validierung eines Beobachter- und eines Schülerfragebogens für die Sekundarstufe) werden zwei neu entwickelte Instrumente–ein Fragebogen für Schülerinnen und Schüler, ein Fragebogen für Beobachter–zur Erfassung von Kompetenzen im Klassenmanagement vorgestellt. Maßgeblich für die Neuentwicklung waren der Mangel an *multidimensionalen* Instrumenten zur Erfassung von Klassenmanagement, der zum Teil fehlende Nachweis über die *psychometrische Güte* des Instruments, sowie der Bedarf einer *perspektivenspezifischen* Operationalisierung von Klassenmanagement. Ausgehend von der in Kapitel 1.2.1 dargestellten Taxonomie von Strategien des Klassenmanagement wurden durch systematische

Literaturrecherche geeignete Items aus bereits entwickelten Instrumenten identifiziert sowie neue Items gebildet, um die Skalen hinreichend abzubilden. Das Manuskript stellt den Konstruktionsprozess der Instrumententwicklung sowie die Maße der Reliabilität und Validität der entwickelten Instrumente dar. Die Erprobung der Instrumente erfolgte in der Haupterhebung mit 88 Lehrkräften und 1811 Schülerinnen und Schülern an Berliner Oberschulen. Es wurden spezifische Hypothesen zu konvergenten und diskriminanten Zusammenhängen von Schüler- und Lehrermerkmalen mit Strategien des Klassenmanagements postuliert und geprüft.

1.6.2. Teilstudie 2: Training Teachers' Classroom Management

Teilstudie 2 (Training Inservice Teachers' Competencies in Classroom Management. A Quasi-Experimental Study with Teachers of Secondary Schools) stellt die Entwicklung und empirische Prüfung eines neu entwickelten Klassenmanagement-Trainings für Sekundarschullehrkräfte dar. Es wird theoretisch argumentiert, welche Aspekte des Klassenmanagements dabei thematisiert werden und welche Lerngelegenheiten dazu dienen, um die Kompetenzentwicklung anzuregen. Mit Hilfe eines multikriterialen, quasi-experimentellen Designs wurde im Prä-Post Vergleich die Interventionsgruppe mit einer Kontrollgruppe, die ebenfalls ein Training durchlief, verglichen. Es nahmen 19 Lehrkräfte in der Interventions- und 18 Lehrkräfte in der Kontrollgruppe teil. Untersucht wurden Effekte auf das Wissen der Lehrkräfte, auf das Können der Lehrkräfte sowie auf die Aufmerksamkeit und Beteiligung der Schülerinnen und Schüler. Aufgrund der Inhalte des Vergleichstrainings wurde angenommen, dass in beiden Gruppen das Wissen, jedoch nur in der Interventionsgruppe auch das Können und damit Veränderungen im Schülerverhalten nach dem Training identifizierbar sind. Zusätzlich zur Programmevaluation wurde eine Prozessevaluation durchgeführt, um die Akzeptanz der Trainingsinhalte zu quantifizieren und Aussagen über einzelne Lerngelegenheiten des Trainings treffen zu können.

1.6.3. Teilstudie 3: Response Shift in Training Evaluation

In Teilstudie 3 (Evaluating Response Shift in Training Evaluation. Comparing the Retrospective Pre-Test with a Structural Equation Modeling Approach in a Classroom Management Training Program) wird das Phänomen des Response Shift und dessen Bedeutung für Interventionsstudien mit Prä-Post Design durch den Vergleich von zwei statistischen Verfahren untersucht. In einem quasi-experimentellen Design wurden die Bedingungen für einen potentiellen Response Shift systematisch variiert, indem zwei unterschiedlich intensive Klassenmanagement-Trainings mit insgesamt 146 Studierenden des Lehramts durchgeführt wur-

den. Vor und nach dem Training wurde das Wissen über Klassenmanagement per Selbsteinschätzung erhoben, zusätzlich wurde eine retrospektive Befragung implementiert. Zum einen wurde über die sukzessive Prüfung der Messinvarianz geprüft, ob und wenn ja welche Arten von Response Shift vorliegen (latente, faktoranalytische Prüfung). Dem wurde die Identifikation von Response Shift durch den Abgleich von konventioneller (vor dem Training) und retrospektiver (rückblickend nach dem Training) Prämessung gegenübergestellt (manifeste, varianzanalytische Prüfung). Es wird angenommen, dass beide methodischen Vorgehensweisen zu den gleichen Aussagen über Response Shift führen, dass jedoch in den zu vergleichenden Gruppen unterschiedliche Arten von Response Shift vorliegen. Ergänzend wurde die Hypothese der Messinvarianz von retrospektiver und Postmessung überprüft.

1.7. Literaturverzeichnis

- Achtenhagen, F. & Baethge, M. (2008). Kompetenzdiagnostik als Large-Scale-Assessment im Bereich der beruflichen Aus- und Weiterbildung. *Zeitschrift für Erziehungswissenschaft, 10*, 51–70.
- Adler, S. (1991). The reflective practitioner and the curriculum of teacher education. *Journal of Education for Teaching, 17* (2), 139–150.
- Admiraal, W., Korthagen, F. A. J. & Wubbels, T. (2000). Effects of student teachers' coping behaviour. *British Journal of Educational Psychology, 70*(1), 333–352.
- Aiken, L. S. & West, S. G. (1990). Invalidity of true experiments: Self-report pretest bias. *Evaluation Review, 14*(4), 374–390.
- Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education, 13*(2), 153–166.
- Arlin, M. (1979). Teacher transitions can disrupt time flow classrooms. *American Educational Research Journal, 16*(1), 42–56.
- Arnold, F. (1908). *Text-book of school and class management, theory and practice*. New York, NY: Macmillan.
- Baartman, L. K., Bastiaens, T. J., Kirschner, P. A. & van der Vleuten, C. P. M. (2007). Evaluating assessment quality in competence-based education: A qualitative comparison of two frameworks. *Educational Research Review, 2*(2), 114–129.
- Baartman, L. K. & de Bruijn, E. (2011). Integrating knowledge, skills and attitudes: Conceptualizing learning processes towards vocational competence. *Educational Research Review, 6*(2), 125–134.
- Babad, E. (1996). How high is “high inference”? Within classroom differences in students' perceptions of classroom interaction. *The Journal of Classroom Interaction, 31*, 1–9.
- Baer, M., Guldemann, T., Kocher, M., Larcher, S., Wyss, C., Dörr, G. et al. (2009). Auf dem Weg zu Expertise beim Unterrichten: Erwerb von Lehrkompetenz im Lehrerinnen- und Lehrerstudium. *Unterrichtswissenschaft, 37*(2), 118–144.
- Bagley, W. C. (1907). *Classroom management. Its principles and technique*. New York, NY: Macmillan.
- Battistich, V., Solomon, D., Watson, M. & Schaps, E. (1997). Caring school communities. *Educational Psychologist, 32*(3), 137–151.
- Bauer, J., Gräsel, C. & Prenzel, M. (2012). Forschung zum Lehramtsstudium: Einführung in das Themenheft. *Unterrichtswissenschaft, 40*(2), 98–100.

- Bauer, J. & Gruber, H. (2007). Workplace changes and workplace learning: advantages of an educational micro perspective. *International Journal of Lifelong Education*, 26(6), 675–688.
- Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U. et al. (2009). *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Baumert, J., Gruehn, S., Heyn, S., Köller, O. & Schnabel, K.-U. (1997). *Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU). Dokumentation: Bd. 1. Skalen Längsschnitt I, Welle 1–4*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29–53). Münster: Waxmann.
- Baumert, J., Kunter, M., Brunner, M., Krauss, S., Blum, W. & Neubrand, M. (2004). Mathematikunterricht aus Sicht der PISA-Schülerinnen und -Schüler und ihrer Lehrkräfte. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, et al. (Hrsg.), *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 314–354). Münster: Waxmann .
- Beck, K. (1987). *Die empirischen Grundlagen der Unterrichtsforschung: Eine kritische Analyse der deskriptiven Leistungsfähigkeit von Beobachtungsmethoden*. Göttingen: Hogrefe.
- Bergin, C. & Bergin, D. A. (1999). Classroom discipline that promotes self-control. *Journal of Applied Developmental Psychology*, 20(2), 189–206.
- Berliner, D. C. (1983). Developing conceptions of classroom environments: Some light on the T in classroom studies of ATI. *Educational Psychologist*, 18(1), 1–13.
- Berliner, D. C. (1994). The wonder of exemplary performances. In J. N. Mangieri & C. C. Block (Hrsg.), *Creating powerful thinking in teachers and students. Diverse perspectives* (S. 1–37). Fort Worth, TX: Holt, Rinehart & Winston.
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35(5), 463–482.

- Berliner, D. C. (2004). Describing the behavior and documenting the accomplishments of expert teachers. *Bulletin of Science, Technology and Society*, 24(3), 200–212.
- Bischoff, S., Brühwiler, C. & Baer, M. (2005). Videotest zur Erfassung „adaptiver Lehrkompetenz“. *Beiträge zur Lehrerbildung*, 23(3), 382–397.
- Blomberg, G., Seidel, T. & Prenzel, M. (2011). Neue Entwicklungen in der Erfassung pädagogisch-psychologischer Kompetenzen von Lehrpersonen. *Unterrichtswissenschaft*, 39(2), 98–101.
- Blömeke, S. (2007). Qualitativ-quantitativ, induktiv-deduktiv, Prozess-Produkt, national-international: Zur Notwendigkeit multikriterialer und multiperspektivischer Zugänge in der Lehrerbildungsforschung. In M. Lüders & J. Wissinger (Hrsg.), *Forschung zur Lehrerbildung. Kompetenzentwicklung und Programmevaluation* (S. 13–37). Münster: Waxmann.
- Blömeke, S., Felbrich, A. & Müller, C. (2008a). Messung des erziehungswissenschaftlichen Wissens angehender Lehrkräfte. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare: Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (S. 171–193). Münster: Waxmann.
- Blömeke, S., Felbrich, A. & Müller, C. (2008b). Theoretischer Rahmen und Untersuchungsdesign. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare: Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (S. 15–48). Münster: Waxmann.
- Blömeke, S., Kaiser, G. & Lehmann, R. (Hrsg.). (2008). *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare: Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung*. Münster: Waxmann.
- Blüthmann, I., Ophardt, D., Thiel, F. & Felsberger, G. (2011). Wissenserwerb zum Thema Klassenmanagement im Lehramtsstudium. Individuelle und studienbezogene Einflussfaktoren. *Unterrichtswissenschaft*, 4, 290–309.
- Bodensohn, R. (2005). Die inflationäre Anwendung des Kompetenzbegriffs fordert die bildungstheoretische Reflexion heraus. In A. Frey (Hrsg.), *Kompetenzdiagnostik. Theorien und Methoden zur Erfassung und Bewertung von beruflichen Kompetenzen* (S. 136–149). Landau: Empirische Pädagogik.

- Bollen, K. A. & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: Wiley.
- Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice*, 16(3), 252–260.
- Borko, H., Jacobs, J., Eiteljorg, E. & Pittman, M. E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching and Teacher Education*, 24(2), 417–436.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Heidelberg: Springer.
- Brandtstädter, J. (1990). Evaluationsforschung: Probleme der wissenschaftlichen Bewertung von Interventions- und Reformprojekten. *Zeitschrift für Pädagogische Psychologie*, 4(4), 215–227.
- Bransford, J., Darling-Hammond, L. & LePage, P. (2005). Introduction. In L. Darling-Hammond & J. Bransford (Hrsg.), *Preparing teachers for a changing world. What teachers should learn and be able to do* (S. 1–39). San Francisco, CA: Jossey-Bass.
- Braun, E. & Hannover, B. (2008). Kompetenzmessung und Evaluation von Studienerfolg. In N. Jude, J. Hartig & E. Klieme (Hrsg.), *Kompetenzerfassung in pädagogischen Handlungsfeldern. Theorien, Konzepte und Methoden* (S. 153–160). Bonn: Bundesministerium für Bildung und Forschung.
- Bromfield, C. (2006). PGCE secondary trainee teachers & effective behaviour management: An evaluation and commentary. *Support for Learning*, 21(4), 188–193.
- Bromme, R. (1992). *Der Lehrer als Experte: Zur Psychologie des professionellen Wissens*. Bern: Huber
- Bromme, R. (1993). Können Lehrer Experten sein – können Experten Lehrer sein? In H. Bauersfeld & R. Bromme (Hrsg.), *Bildung und Aufklärung. Studien zur Rationalität des Lehrens und Lernens* (S. 33–58). Münster: Waxmann.
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (S. 177–212). Göttingen: Hogrefe.
- Bromme, R. & Tillema, H. (1995). Fusing experience and theory: The structure of professional knowledge. *Learning and Instruction*, 5(4), 261–267.
- Brookhart, S. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record*, 106(3), 429–458.

- Brophy, J. E. (1985). Classroom management as instruction: Socializing self-guidance in students. *Theory Into Practice*, 24(4), 233–240.
- Brophy, J. E. (1988). Educating teachers about managing classrooms and students. *Teaching and Teacher Education*, 4(1), 1–18.
- Brophy, J. E. (1996). *Teaching problem students*. New York, NY: Guilford Press.
- Brophy, J. E. (1998). Classroom management as socializing students into clearly articulated roles. *Journal of Classroom Interaction*, 33(1), 1–4.
- Brophy, J. E. (1999). Perspectives of classroom management. In H. J. Freiberg (Hrsg.), *Beyond behaviorism: Changing the classroom management paradigm* (S. 43–56). Boston: Allyn & Bacon.
- Brophy, J. E. (2006). History of research on classroom management. In C. M. Evertson & C. S. Weinstein (Hrsg.), *Handbook of classroom management. Research, practice and contemporary issues* (S. 17–43). Mahwah, NJ: Lawrence Erlbaum Associates.
- Brouwers, A. & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education*, 16(2), 239–253.
- Bühner, M. (2004). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson.
- Canter, L. & Canter, M. (1992). *Assertive discipline: Positive behavior management for today's classroom* (2. Aufl.). Santa Monica, CA: Lee Canter & Associates.
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, 64, 723–733.
- Chávez Chávez, R. (1984). The use of high-interference measures to study classroom climates: A review. *Review of Educational Research*, 54(2), 237–261.
- Chilcoat, G. W. (1989). Instructional behaviors for clearer presentations in the classroom. *Instructional Science*, 18(4), 289–314.
- Choi, I. & Lee, K. (2009). Designing and implementing a case-based learning environment for enhancing ill-structured problem solving: Classroom management problems for prospective teachers. *Educational Technology Research and Development*, 57(1), 99–129.
- Cipani, E. (2008). *Classroom management for all teachers: Plans for evidence-based practice* (3. Aufl.). Upper Saddle River, NJ: Pearson Merrill/Prentice Hall.
- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive?* Münster: Waxmann.
- Clausen, M., Reusser, K. & Klieme, E. (2003). Unterrichtsqualität auf der Basis hochinferenter Unterrichtsbeurteilungen. *Unterrichtswissenschaft*, 31(2), 122–141.

- Clunies-Ross, P., Little, E. & Kienhuis, M. (2008). Self-reported and actual use of proactive and reactive classroom management strategies and their relationship with teacher stress and student behaviour. *Educational Psychology*, 28(6), 693–710.
- Cochran-Smith, M. & Zeichner, K. (Hrsg.). (2005). *Studying teacher education: The report of the AERA Panel on research and teacher education*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Codding, R. S. & Smyth, C. A. (2008). Using performance feedback to decrease classroom transition time and examine collateral effects on academic engagement. *Journal of Educational and Psychological Consultation*, 18(4), 325–345.
- Copeland, W. D. (1987). Classroom management and student teachers' cognitive abilities: A relationship. *American Educational Research Journal*, 24(2), 219–236.
- Courvoisier, D. S., Nussbeck, F. W., Eid, M., Geiser, C. & Cole, D. A. (2008). Analyzing the convergent and discriminant validity of states and traits: Development and applications of multimethod latent state-trait models. *Psychological Assessment*, 20(3), 270–280.
- Cramer, C. (2010). Kompetenzerwartungen Lehramtsstudierender: Grenzen und Perspektiven selbsteingeschätzter Kompetenzen in der Lehrerbildungsforschung. In A. Gehrman, U. Hericks & M. Lüders. (Hrsg.), *Bildungsstandards und Kompetenzmodelle. Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht* (S. 85–97). Bad Heilbrunn: Klinkhardt.
- Cronbach, L. J. (1976). *Research on classrooms and schools: Formulations of questions, design and analysis*. Stanford, CA: Stanford Evaluation Consortium.
- Dann, H.-D. & Humpert, W. (2002). Das Konstanzer Trainingsmodell (KTM) – Grundlagen und neue Entwicklungen. *Zeitschrift für Pädagogik*, 48 (2), 215–226.
- Darling-Hammond, L. (2001). Standard setting in teaching: Changes in licensing, certification, and assessment. In V. Richardson (Hrsg.), *Handbook of research on teaching* (4. Aufl.). Washington, DC: AERA.
- Darling-Hammond, L., Hammerness, K., Grossman, P., Rust, F. & Shulman, L. S. (2005). The design of teacher education programs. In L. Darling-Hammond & J. Bransford (Hrsg.), *Preparing teachers for a changing world. What teachers should learn and be able to do* (S. 390–441). San Francisco, CA: Jossey-Bass.
- Darling-Hammond, L., Wise, A. E. & Klein, S. P. (1999). *A license to teach: Raising standards for teaching*. San Francisco: Jossey-Bass Publishers.
- De Jong, R. & Westerhof, K. J. (2001). The quality of student ratings of teacher behaviour. *Learning Environments Research*, 4(1), 51–85.

- Den Brok, P., Brekelmans, M. & Wubbels, T. (2006). Multilevel issues in research using students' perceptions of learning environments: The case of the Questionnaire on Teacher Interaction. *Learning Environments Research*, 9(3), 199–213.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181–199.
- Dimitrov, D. M. & Rumrill, P. D. (2003). Pretest-posttest designs and measurement of change. *Journal of Prevention, Assessment & Rehabilitation* 20(2), 159–165.
- Ditton, H. (2000). Qualitätskontrolle und Qualitätssicherung in Schule und Unterricht. Ein Überblick zum Stand der empirischen Forschung. In A. Helmke, W. Hornstein & E. Terhart (Hrsg.), *Qualität und Qualitätssicherung im Bildungsbereich* (S. 73–92). Weinheim: Beltz.
- Ditton, H. & Merz, D. (2000). *Qualität von Schule und Unterricht. Kurzbericht über erste Ergebnisse einer Untersuchung an bayerischen Schulen*. Eichstätt: Katholische Universität.
- Ditton, H. & Merz, D. (2001). *Fragebogen für Schülerinnen und Schüler*. Zugriff am 10.09.2013, http://www.quassu.net/Sch_fgb1.pdf
- Doyle, W. (1983). Academic work. *Review of Educational Research*, 53(2), 159–199.
- Doyle, W. (1984). How order is achieved in classrooms: An interim report. *Journal of Curriculum Studies*, 16(3), 259–277.
- Doyle, W. (1986). Classroom organization and management. In M. Wittrock (Hrsg.), *Handbook of research on teaching* (3. Aufl.). New York, NY: Macmillan.
- Doyle, W. (2006). Ecological approaches to classroom management. In C. M. Evertson & C. S. Weinstein (Hrsg.), *Handbook of classroom management. Research, practice and contemporary issues* (S. 97–125). Mahwah, NJ: Lawrence Erlbaum Associates.
- Doyle, W. & Carter, K. (1984). Academic tasks in classrooms. *Curriculum Inquiry*, 14, 129–149.
- Dreyfus, H. L. & Dreyfus, S. E. (1986). *Mind over machine. The power of human intuition and expertise in the era of the computer*. New York, NY: Free Press.
- Dubs, R. (1995). *Lehrerverhalten: Ein Beitrag zur Interaktion von Lehrenden und Lernenden im Unterricht*. Zürich: Verlag des Schweizerischen Kaufmännischen Verbandes.
- Duncan, T. E., Duncan, S. C. & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications* (2. Aufl.). Mahweh, NJ: Lawrence Erlbaum.

- Eckert, T. L., Dunn, E. K., Guiney, K. M. & Coddig, R. S. (2000). Self-reports: Theory and research in using rating scale measures. In E. S. Shapiro & T. R. Kratochwill (Hrsg.), *Behavioral assessment in schools. Theory, research, and clinical foundations* (2. Aufl., S. 288–322). New York, NY: Guilford Press.
- Edelmann, D. & Tippelt, R. (2007). Kompetenzentwicklung in der beruflichen Bildung und Weiterbildung. *Zeitschrift für Erziehungswissenschaft*, 10(8), 129–146.
- Eder, F. (1998). *Linzer Fragebogen zum Schul- und Klassenklima für die 8.-13. Klasse*. Göttingen: Hogrefe.
- Eder, F. & Bergmann, C. (2004). Der Einfluss von Interessen auf die Lehrer-Wahrnehmung von Schülerinnen und Schülern. *Empirische Pädagogik*, 18(4), 410–431.
- Eichhorn, C. (2008). *Classroom-Management: Wie Lehrer, Eltern und Schüler guten Unterricht gestalten*. Stuttgart: Klett-Cotta.
- Eid, M. (2003). Veränderungsmessung und Kausalanalyse. In M. Jerusalem & H. Weber (Hrsg.), *Psychologische Gesundheitsförderung – Diagnostik und Prävention* (S. 105–120). Göttingen: Hogrefe.
- Eigenmann, J. (2009). Engagierte Gegenseitigkeit: Klassenmanagement in schwierigen Unterrichtssituationen. *Pädagogik*, 61(2), 24–27.
- Elliot, S. & Busse, R. T. (2004). Assessment and evaluation of students' behavior and intervention outcomes. The utility of rating scale methods. In R. B. Rutherford (Hrsg.), *Handbook of research in emotional and behavioral disorders* (S. 123–142). New York, NY: Guilford Press.
- Emmer, E. T. & Aussiker, A. (1989). School and classroom discipline programs: How well do they work? In O. C. Moles (Hrsg.), *Strategies to reduce student misbehavior* (S. 105–142). Washington, DC: Office of Educational Research and Improvement.
- Emmer, E. T., Evertson, C. M. & Anderson, L. M. (1980). Effective classroom management at the beginning of the school year. *The Elementary School Journal*, 80(5), 219–231.
- Emmer, E. T., Evertson, C. M. & Worsham, M. E. (2003). *Classroom management for secondary teachers* (6. Aufl.). Boston: Allyn and Bacon.
- Emmer, E. T. & Gerwels, M. C. (2006). Classroom management in middle and high school classrooms. In C. M. Evertson & C. S. Weinstein (Hrsg.), *Handbook of classroom management. Research, practice and contemporary issues* (S. 407–437). Mahwah, NJ: Lawrence Erlbaum Associates.

- Emmer, E. T. & Stough, L. M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist*, 36(2), 103–112.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors and Ergonomics Society*, 37(1), 32–64.
- Eraut, M. (1994). *Developing professional knowledge and competence*. London, Washington, D.C: Falmer Press.
- Ericsson, K. A. (2003). The acquisition of expert performance as problem solving: Construction and modification of mediating mechanisms through deliberate practice. In J. E. Davidson & R. J. Sternberg (Hrsg.), *Problem solving* (S. 3–83). New York, NY: Cambridge University Press.
- Erpenbeck, J. & von Rosenstiel, L. (2007). Einführung. In J. Erpenbeck & von Rosenstiel, L. (Hrsg.), *Handbuch Kompetenzmessung. Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis* (2. Aufl., S. XVII–XLVI). Stuttgart: Schäffer-Poeschel.
- Ettlin, E. & Fuchs, M. (2007). Professionsstandards und Kompetenzentwicklung in der Lehrerinnen- und Lehrerbildung. In K. Möller, P. Hanke, C. Beinbrech, A. K. Hein, T. Kleickmann & R. Schages (Hrsg.), *Qualität von Grundschulunterricht entwickeln, erfassen und bewerten* (S. 115–118). Bonn: Verlag für Sozialwissenschaften.
- Evertson, C. M. (1985). Training teachers in classroom management: An experimental study in secondary school classrooms. *Journal of Educational Research*, 79(1), 51–58.
- Evertson, C. M. & Burry, J. A. (1989). Capturing classroom context: The observation system as lens for assessment. *Journal of Personnel Evaluation in Education*, 2(4), 297–320.
- Evertson, C. M. & Emmer, E. T. (1982). Effective management at the beginning of the school year in junior high classes. *Journal of Educational Psychology*, 74(4), 485–498.
- Evertson, C. M. & Harris, A. H. (1992). What we know about managing classrooms. *Educational Leadership*, 49(7), 74–78.
- Evertson, C. M. & Harris, A. H. (1999). Support for managing learning-centered classrooms: The classroom organization and management program. In H. J. Freiberg & J. E. Brophy (Hrsg.), *Beyond behaviorism. Changing the classroom management paradigm* (S. 59–74). Boston: Allyn and Bacon.
- Evertson, C. M. & Weinstein, C. S. (2006a). Classroom management as a field of inquiry. In C. M. Evertson & C. S. Weinstein (Hrsg.), *Handbook of classroom management. Re-*

- search, practice and contemporary issues* (S. 3–15). Mahwah, NJ: Lawrence Erlbaum Associates.
- Evertson, C. M. & Weinstein, C. S. (Hrsg.). (2006b). *Handbook of classroom management: Research, practice and contemporary issues*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Felbrich, A., Müller, C. & Blömeke, S. (2008). Lerngelegenheiten in der Lehrerbildung. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehenden Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und –referendare. Erste Ergebnisse zur Wirksamkeit der Lehrerbildung* (S. 327–362). Münster: Waxmann.
- Fend, H. (1977). *Schulklima. Soziale Einflußprozesse in der Schule*. Weinheim: Beltz.
- Fenstermacher, G. D. (1994). The knower and the known: The nature of knowledge in research on teaching. *Review of Research in Education*, 20, 3–56.
- Fernandez-Ballesteros, R. (2004). Self-report questionnaires. In S. N. Haynes (Hrsg.), *Comprehensive handbook of psychological assessment* (S. 197–221). Hoboken, NJ: John Wiley & Sons.
- Fleischer, J., Leutner, D. & Klieme, E. (2012). *Modellierung von Kompetenzen im Bereich der Bildung: Eine psychologische Perspektive* [Themenheft]. Psychologische Rundschau, 3.
- Fraser, B. J. (1980). *Criterion validity on an individualised classroom environment questionnaire*. Sydney, Australia: McQuaire University.
- Fraser, B. J. (1991). Two decades of classroom environment research. In H. J. Walberg (Hrsg.), *Educational environments: Evaluation, antecedents and consequences* (S. 3–27). Elmsford, NY: Pergamon.
- Fraser, B. J., Anderson, G. J. & Walberg, H. J. (1982). *Assessment of learning environments: Manual for Learning Environment Inventory (LEI) and My Class Inventory (MCI)*. Perth, Australia: Western Australian Institute for Technology.
- Freiberg, H. J. (1983). Consistency: The key to classroom management. *Journal of Education for Teaching*, 9(1), 1–15.
- Freiberg, H. J. (1999). Beyond behaviorism. In H. J. Freiberg & J. E. Brophy (Hrsg.), *Beyond behaviorism: Changing the classroom management paradigm* (S. 3–20). Boston: Allyn and Bacon.
- Freiberg, H. J. & Lapointe, J. M. (2006). Research-based programs for preventing and solving discipline problems. In C. M. Evertson & C. S. Weinstein (Hrsg.), *Handbook of class-*

- room management. Research, practice and contemporary issues* (S. 735–786). Mahwah, NJ: Lawrence Erlbaum Associates.
- Freiberg, H. J., Stein, T. A. & Huang, S. Y. (1995). Effects of a classroom management intervention on student achievement in inner-city elementary schools. *Educational Research and Evaluation*, 1(1), 36–66.
- Freund, P. A. & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin*, 138(2), 296–321.
- Frey, A. (2005). Kompetenzdiagnostik – Theorien und Methoden zur Erfassung und Bewertung von beruflichen Kompetenzen. In A. Frey (Hrsg.), *Kompetenzdiagnostik. Theorien und Methoden zur Erfassung und Bewertung von beruflichen Kompetenzen* (S. 1–5). Landau: Empirische Pädagogik.
- Frey, A. (2006). Methoden und Instrumente zur Diagnose beruflicher Kompetenzen von Lehrkräften: Eine erste Standortbestimmung zu bereits publizierten Instrumenten. *Zeitschrift für Pädagogik*, 51, 30–46.
- Friedman, I. A. (2006). Classroom management and teacher stress and burnout. In C. M. Evertson & C. S. Weinstein (Hrsg.), *Handbook of classroom management. Research, practice and contemporary issues* (S. 925–945). Mahwah, NJ: Lawrence Erlbaum Associates.
- Fries, S. & Souvignier, E. (2009). Training. In E. Wild & K. Möller (Hrsg.), *Pädagogische Psychologie* (S. 405–428). Heidelberg: Springer.
- Fukkink, R. G., Trienekens, N. & Kramer, L. J. C. (2011). Video feedback in education and training. Putting learning in the picture. *Educational Psychology Review*, 23(1), 45–63.
- Gallagher, T. L. (2009). Teacher candidates' questions within the context of an educational psychology course. *Teaching Educational Psychology*, 5(2), 50–60.
- Gandhi, P. K., Ried, L. D., Huang, I. C., Kimberlin, C. L. & Kauf, T. L. (2013). Assessment of response shift using two structural equation modeling techniques. *Quality of Life Research*, 22(3), 461–471.
- Gärtner, H. (2008). Implementation neuer Lehrpläne innerhalb eines videobasierten Qualitätszirkels – eine quasi-experimentelle Feldstudie. *Unterrichtswissenschaft*, 36(1), 37–54.
- Gärtner, H. (2010). Wie Schülerinnen und Schüler ihre Lernumwelt wahrnehmen: Ein Vergleich verschiedener Maße zur Übereinstimmung von Schülerwahrnehmungen. *Zeitschrift für Pädagogische Psychologie*, 24(2), 111–122.

- Gärtner, H. (2011). Selbstevaluation von Unterricht? Angebot und Nutzen des Selbstevaluationsportals (SEP). *BLuS*, 9, 19–25.
- Geving, A. M. (2007). Identifying the types of student and teacher behaviours associated with teacher stress. *Teaching and Teacher Education*, 23(5), 624–640.
- Gholami, K. & Husu, J. (2010). How do teachers reason about their practice? Representing the epistemic nature of teachers' practical knowledge. *Teaching and Teacher Education*, 26 (8), 1520–1529.
- Gillies, R. M. (2009). *Evidence-based teaching: Strategies that promote learning*. Rotterdam: Sense.
- Goeze, A., Hetfleisch, P. & Schrader, J. (2013). Wirkungen des Lernens mit Videofällen bei Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 16(1), 79–113.
- Gold, B., Förster, S. & Holodynski, M. (2013). Evaluation eines videobasierten Trainingsseminars zur Förderung der professionellen Wahrnehmung von Klassenführung im Grundschulunterricht. *Zeitschrift für Pädagogische Psychologie*, 27(3), 141–155.
- Gold, B. & Holodynski, M. (2011). Klassenführung. In E. Kiel & K. Zierer (Hrsg.), *Basiswissen Unterrichtsgestaltung: Unterrichtsgestaltung als Gegenstand der Praxis* (S. 133–151). Hohengehren: Schneider.
- Goldman, R., Pea, R., Barron, B. & Denny, S.J. (2007). *Video research in the learning sciences*. Mahwah: Lawrence Erlbaum.
- Gollwitzer, M. & Jäger, R. S. (2009). *Evaluation kompakt*. Weinheim: Beltz.
- Gordon, L. M. (2002, April). *A new self-report instrument for assessing teachers' classroom management intervention strategies*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Gordon, T. (1974). *T.E.T. Teacher Effectiveness Training*. New York, NY: Wiley.
- Gottfredson, D. C., Gottfredson, G. D. & Hybl, L. G. (1993). Managing adolescent behaviour. A multiyear, multischool study. *American Educational Research Journal*, 30(1), 179–215.
- Greenwald, A. G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, 52(11), 1182–1186.
- Grewe, W. & Wentura, D. (1997). *Wissenschaftliche Beobachtung. Eine Einführung* (2. Aufl.). Weinheim: Psychologie Verlags Union.
- Gröschner, A., Schmitt, C. & Seidel, T. (2013). Veränderung subjektiver Kompetenzeinschätzungen von Lehramtsstudierenden im Praxissemester. *Zeitschrift für Pädagogische Psychologie*, 27(1), 77–86.

- Grossman, P. L. (1995). Teachers' knowledge. In L. W. Anderson (Hrsg.), *Resources in education. International encyclopedia of teaching and teacher education* (2. Aufl., S. 20–24). Oxford, New York, NY: Pergamon.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen: Schüler als Quellen der Unterrichtsbeschreibung*. Münster: Waxmann.
- Gump, P. V. (1967). *The classroom behaviour setting: Its nature and relation to student behaviour (final report)*. Washington, DC: U.S. Office of Education, Bureau of Research.
- Guskey, T. R. (2003). *Evaluating professional development*. Thousand Oaks, CA: Corwin Press.
- Haag, L. & Streber, D. (2012). *Klassenführung: Erfolgreich unterrichten mit Classroom Management*. Weinheim: Beltz.
- Hammerness, K., Darling-Hammond, L., Bransford, J., Berliner, D. C., Cochran-Smith, M., McDonald, M. et al. (2005). How teachers learn and develop. In L. Darling-Hammond & J. Bransford (Hrsg.), *Preparing teachers for a changing world. What teachers should learn and be able to do* (S. 358–389). San Francisco, CA: Jossey-Bass.
- Han, I., Eom, M. & Shin, W. S. (2013). Multimedia case-based learning to enhance pre-service teachers' knowledge integration for teaching with technologies. *Teaching and Teacher Education*, 34, 122–129.
- Hardin, C. J. (2008). *Effective classroom management: Models and strategies for today's classrooms* (2. Aufl.). Upper Saddle River, NJ: Pearson/Merrill/Prentice Hall.
- Hart, R. (2010). Classroom behaviour management: Educational psychologists' views on effective practice. *Emotional and Behavioural Difficulties*, 15(4), 353–371.
- Hartig, J. (2008). Kompetenzen als Ergebnisse von Bildungsprozessen. In N. Jude, J. Hartig & E. Klieme (Hrsg.), *Kompetenzen in pädagogischen Handlungsfeldern. Theorien, Konzepte und Methoden* (S.13–24). Bonn: Bundesministerium für Bildung und Forschung.
- Hartig, J., Frey, A. & Jude, N. (2007). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 135–164). Heidelberg: Springer.
- Hastie, P. A., Sinelnikov, O. A. & Brock, S. J. (2007). Kounin revisited: Tentative postulates for an expanded examination of classroom ecologies. *Journal of Teaching in Physical Education*, 26(3), 298–309.
- Hastings, R. P. & Bham, M. S. (2003). The relationship between student behaviour patterns and teacher burnout. *School Psychology International*, 24 (1), 115–127.

- Hattie, J. (2009). *Visible learning: A synthesis of meta-analyses relating to achievement*. London: Routledge.
- Hatton, N. & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. *Teaching and Teacher Education*, 11(1), 33–49.
- Havers, N. (1998). Disziplinschwierigkeiten im Unterricht. Ein Trainingsseminar im Lehrstudium. *Die Deutsche Schule*, 90(3), 189–198.
- Havers, N. (2006). Lehrertraining für angehende Referendare. In B. Jürgens (Hrsg.), *Kompetente Lehrer ausbilden. Vernetzung von Universität und Schule in der Lehreraus- und -weiterbildung* (S. 81–87). Aachen: Shaker.
- Havers, N. (2010). Lässt sich effiziente Klassenführung lehren? Das Potential des Lehrertrainings. In J. Abel & G. Faust (Hrsg.), *Wirkt Lehrerbildung? Antworten aus der empirischen Forschung* (S. 283–290). Münster: Waxmann.
- Havers, N. & Toepell, S. (2002). Trainingsverfahren für die Lehrerbildung im deutschen Sprachraum. *Zeitschrift für Pädagogik*, 48(2), 174–193.
- Heckt, D. H. (2005). Classroom Management und... *Grundschule*, 37(2), 8–9.
- Helmke, A. (2007a). Aktive Lernzeit optimieren: Was wissen wir über effiziente Klassenführung? *Pädagogik*, 59(5), 44–49.
- Helmke, A. (2007b). Lernprozesse anregen und steuern: Was wissen wir über Klarheit und Strukturiertheit? *Pädagogik*, 59(6), 44–47.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität*. Seelze-Velber: Klett/Kallmeyer.
- Helmke, A., Helmke, T., Heyne, N., Hosenfeld, A., Schrader, F.-W. & Wagner, W. (2010). Effiziente Klassenführung als Schlüsselmerkmal der Unterrichtsqualität – ein Untersuchungsbeispiel aus der Grundschule. In C. Spiel (Hrsg.), *Bildungspsychologie* (S. 101–105). Göttingen: Hogrefe.
- Helmke, A., Helmke, T., Schrader, F.-W. & Wagner, W. (2007). *Der Ratingbogen der DESI-Videostudie*. Landau: Universität Koblenz-Landau.
- Helmke, A., Hosenfeld, I., Schrader, F.-W. & Wagner, W. (2002). Unterricht aus der Sicht der Beteiligten. In A. Helmke & R. S. Jäger (Hrsg.), *Das Projekt MARKUS – Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext* (S. 325–411). Landau: Verlag Empirische Pädagogik.
- Helmke, A., Schneider, W. & Weinert, F. E. (1986). Quality of instruction and classroom learning outcomes: The German contribution to the IEA classroom environment study. *Teaching and Teacher Education*, 2(1), 1–18.

- Henry, B., Moffitt, T. E., Caspi, A., Langley, J. & Silva, P. A. (1994). On the “remembrance of things past”: A longitudinal evaluation of the retrospective method. *Psychological Assessment*, 6(2), 92–101.
- Hill, L. G. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation*, 26(4), 501–517.
- Horster, L. & Rolff, H. (2001). *Unterrichtsentwicklung: Grundlagen, Praxis, Steuerungsprozesse*. Weinheim: Beltz.
- Howard, G. S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. *Evaluation Review*, 4(1), 93–106.
- Hoyt, W. & Kerns, M.-D. (1999). Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods*, 4(4), 403–424.
- Hussy, W. & Jain, A. (2002). *Experimentelle Hypothesenprüfung in der Psychologie*. Göttingen: Hogrefe.
- Ingersoll, R. M. (2002). The teacher shortage: A case of wrong diagnosis and wrong prescription. *NASSP Bulletin*, 86, 16–31.
- Jankisz, E. & Moosbrugger, H. (2007). Planung und Entwicklung von psychologischen Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 27–72). Heidelberg: Springer.
- Johnson, D. W. & Johnson, R. T. (2000). The Three Cs of reducing prejudice and discrimination. In S. Oskamp (Hrsg.), *Reducing prejudice and discrimination* (S. 239–268). Mahwah, NJ: Lawrence Erlbaum Associates.
- Johnson, D. W. & Johnson, R. T. (2006). Conflict resolution, peer mediation and peacemaking. In C. M. Evertson & C. S. Weinstein (Hrsg.), *Handbook of classroom management. Research, practice and contemporary issues* (S. 803–832). Mahwah, NJ: Lawrence Erlbaum Associates.
- Jones, V. F. (1996). Classroom management. In J. Silula (Hrsg.), *Handbook of research on teacher education* (2. Aufl., S. 503–521). New York, NY: Macmillan.
- Jones, V. F. (2006). How do teachers learn to be effective classroom managers? In C. M. Evertson & C. S. Weinstein (Hrsg.), *Handbook of classroom management. Research, practice and contemporary issues* (S. 887–907). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kämpfe, N. (2009). Schülerinnen und Schüler als Experten für Unterricht. *Die deutsche Schule*, 101(2), 149–163.

- Kaufhold, M. (2006). *Kompetenz und Kompetenzerfassung: Analyse und Beurteilung von Verfahren der Kompetenzerfassung*. Wiesbaden: Verlag für Sozialwissenschaften.
- Keller, G. (2008). *Disziplinmanagement in der Schulklasse: Unterrichtsstörungen vorbeugen, Unterrichtsstörungen bewältigen*. Bern: Huber.
- Kemna, P. W. (2012). *Messung pädagogischer Basiskompetenzen von Lehrerinnen und Lehrern: Entwicklung von Testinstrumenten: Vol. 571*. Münster: Waxmann.
- Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review*, 8(3), 265–280.
- Keuffer, J. (2010). Videofeedback – Ein Konzept für die Lehrerbildung mit Zukunft. In M. Trautmann (Hrsg.), *Unterrichtsentwicklung durch Videofeedback. Besser kommunizieren lernen* (S. 187–200). Göttingen: Vandenhoeck & Ruprecht.
- Killus, D. (2006). Vermittlung von Lernstrategien im Unterricht: Lässt sich das Lehrerhandeln mit einem Fragebogen authentisch erfassen? In S. Rahn, I. Mannes & M. Schratz (Hrsg.), *Schulpädagogische Forschung* (S. 67–80). Innsbruck: Studienverlag.
- Kim, E. S. & Willson, V. L. (2010). Evaluating pretest effects in pre-post studies. *Educational and Psychological Measurement*, 70(5), 744–759.
- Kirchhoff, S., Kuhnt, S. & Lipp, P. (2008). *Der Fragebogen: Datenbasis, Konstruktion und Auswertung* (4. Aufl.). Wiesbaden: Verlag für Sozialwissenschaften.
- Kirkpatrick, D. L. (1998). *Evaluating training programs. The four levels* (2. Aufl.). San Francisco, CA: Berrett-Koehler.
- Klieme, E. (2004). Was sind und wie misst man Kompetenzen? *Pädagogik*, 56(6), 10–13.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, P., Prenzel, M. et al. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Berlin: Bundesministerium für Bildung und Forschung (BMBF) .
- Klieme, E. & Hartig, J. (2007). Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. *Zeitschrift für Erziehungswissenschaft*, 10(8), 11–29.
- Klieme, E. & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik*, 52(6), 876–903.
- Klieme, E. & Pauli, C. R. D. (2006). *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“*. Frankfurt am Main: GPF.

- Klieme, E., Schümer, G. & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: Aufgabenkultur und Unterrichtsgestaltung. In Bundesministerium für Bildung und Forschung (Hrsg.), *TIMSS – Impulse für Schule und Unterricht* (S. 43–57). München: Biering.
- Klinzing, H.-G. (2002). Wie effektiv ist Microteaching? Ein Überblick über fünfunddreißig Jahre Forschung. *Zeitschrift für Pädagogik*, 48, 194–214.
- Klusmann, U., Kunter, M., Voss, T. & Baumert, J. (2012). Berufliche Beanspruchung angehender Lehrkräfte: Die Effekte von Persönlichkeit, pädagogischer Vorerfahrung und professioneller Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 26(4), 275–290.
- Köller, O. (2007). Die deutsche Schule im Lichte internationaler Schulleistungsuntersuchungen (TIMSS, PISA, IGLU, DESI). In H. J. Apel & W. Sacher (Hrsg.), *Studienbuch Schulpädagogik* (3. Aufl., S. 138–154). Bad Heilbrunn: Klinkhardt.
- Köller, O. (2009). Evaluation pädagogisch-psychologischer Maßnahmen. In E. Wild & K. Möller (Hrsg.), *Pädagogische Psychologie* (S. 333–352). Heidelberg: Springer.
- König, J. (2012). Die Entwicklung von pädagogischem Unterrichtswissen: Theoretischer Rahmen, Testinstrument, Skalierung und Ergebnisse. In J. König & A. Seifert (Hrsg.), *Lehramtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LER zur Wirksamkeit der erziehungswissenschaftlichen Lehrerausbildung*. (S. 141–182). Münster: Waxmann.
- König, J. & Blömeke, S. (2009). Pädagogisches Wissen von angehenden Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 12(3), 499–527.
- König, J. & Blömeke, S. (2010). *Pädagogisches Unterrichtswissen (PUW). Dokumentation der Kurzfassung des TEDS-M-Testinstruments zur Kompetenzmessung in der ersten Phase der Lehrerausbildung*. Berlin: Humboldt-Universität.
- Korthagen, F., Loughran, J. & Russell, T. (2006). Developing fundamental principles for teacher education programs and practices. *Teaching and Teacher Education*, 22(8), 1020–1041.
- Kounin, J. S. (1970). *Discipline and group management in classrooms*. New York, NY: Holt, Rinehart & Winston.
- Kounin, J. S. (1976). *Techniken der Klassenführung*. Bern: Huber.
- Kounin, J. S. & Gump, P. V. (1974). Signal systems of lesson settings and the task-related behavior of preschool children. *Journal of Educational Psychology*, 66(4), 554–562.
- Krammer, K. & Reusser, K. (2005). Unterrichtsvideos als Medium der Aus- und Weiterbildung von Lehrpersonen. *Beiträge zur Lehrerbildung*, 23(1), 35–50.

- Krammer, K., Ratzka, N., Klieme, E., Lipowsky, F., Pauli, C. & Reusser, K. (2006). Learning with classroom videos: Conception and first results of an online teacher-training program. *Zentralblatt für Didaktik der Mathematik*, 38, 422–432.
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50(1), 537–567.
- Kubany, E. S., Sloggett, B. B. & Ogata, R. F. (1974). Training teachers in classroom management skills. *American Journal of Community Psychology*, 2(2), 139–149.
- Kunter, M. & Baumert, J. (2011). Das COACTIV-Forschungsprogramm zur Untersuchung professioneller Kompetenz von Lehrkräften: Zusammenfassung und Diskussion. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 346–369). Münster: Waxmann.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Kunter, M., Baumert, J. & Köller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction*, 17(5), 494–509.
- Kunter, M., Brunner, M., Baumert, J., Klusmann, U., Krauss, S., Blum, W. et al. (2005). Der Mathematikunterricht der PISA-Schülerinnen und -Schüler. Schulformunterschiede in der Unterrichtsqualität. *Zeitschrift für Erziehungswissenschaft*, 8(4), 502–520.
- Kunter, M., Kleickmann, T., Klusmann, U. & Richter, D. (2011). Die Entwicklung professioneller Kompetenz von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 55–68). Münster: Waxmann.
- Kunter, M. & Klusmann, U. (2010). Kompetenzmessung bei Lehrkräften. *Unterrichtswissenschaft*, 38(1), 68–86.
- Lam, T. C. M. (2003). A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation*, 24(1), 65–80.
- Landrum, T. J. & Kauffman, J. M. (2006). Behavioral approaches to classroom management. In C. M. Evertson & C. S. Weinstein (Hrsg.), *Handbook of classroom management. Research, practice and contemporary issues* (S. 47–71). Mahwah, NJ: Lawrence Erlbaum Associates.
- Langer, I. & Schulz von Thun, F. (1974). *Messung komplexer Merkmale in Psychologie und Pädagogik. Ratingverfahren*. München: Reinhardt.

- Larose, S., Chaloux, N., Monaghan, D. & Tarabulsky, G. M. (2010). Working alliance as a moderator of the impact of mentoring relationships among academically at-risk students. *Journal of Applied Social Psychology, 40*(10), 2656–2686.
- Larrivee, B. (2006). The convergence of reflective practice and effective classroom management. In C. Evertson & C. Weinstein (Hrsg.), *Handbook of classroom management: Research, practice, and contemporary issues* (S. 983–1001). Mahwah, NJ: Lawrence Erlbaum.
- Larrivee, B. (2009). *Authentic classroom management: Creating a learning community and building reflective practice*. New Jersey: Pearson Education.
- Leflot, G., van Lier, P. A. C., Onghena, P. & Colpin, H. (2010). The role of teacher behavior management in the development of disruptive behaviors: An intervention study with the good behavior game. *Journal of Abnormal Child Psychology, 38*(6), 869–882.
- Leinhardt, G., Weidman, C. & Hammond, K. M. (1987). Introduction and integration of classroom routines by expert teachers. *Curriculum Inquiry, 17*(2), 135–176.
- Leinhardt, G., Young, K.M. & Merriman, J. (1995). Integrating professional knowledge: The theory of practice and the practice of theory. *Learning and Instruction, 5*, 401–408.
- LePage, P., Darling-Hammond, L., Akar, H., Gutierrez, C., Jenkins-Gunn, E. & Rosebrock, K. (2005). Classroom management. In L. Darling-Hammond and J. Bransford (Hrsg.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (S. 327–357). San Francisco, CA: Jossey-Bass.
- Lewis, R., Romi, S., Qui, X. & Katz, Y. J. (2005). Teachers' classroom discipline and student misbehavior in Australia, China and Israel. *Teaching and Teacher Education, 21*(6), 729–741.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz.
- Lipowsky, F. (2010). Lernen im Beruf. Empirische Befunde zur Wirksamkeit von Lehrerfortbildung. In F. H. Müller (Hrsg.), *Lehrerinnen und Lehrer Lernen* (S. 51–70). Münster: Waxmann.
- Little, S. G. & Akin-Little, A. (2008). Psychology's contributions to classroom management. *Psychology in the Schools, 45*(3), 227–234.
- Little, T. D., Bovaird, J. A. & Slegers, D. W. (2005). Methods for the analysis of change. In D. Mroczek & T. D. Little (Hrsg.), *Handbook of personality development Mahwah* (S. 181–211), NJ: Erlbaum.
- Lohmann, G. (2003). *Mit Schülern klarkommen: Professioneller Umgang mit Unterrichtsstörungen und Disziplinkonflikten*. Berlin: Cornelsen.

- Lotz, M., Berner, N. E., Gabriel, K., Post, S., Faust, G. & Lipowsky, F. (2011). Unterrichtsbeobachtung im Projekt PERLE. In D. Kucharz, T. Irion & B. Reinhoffer (Hrsg.), *Grundlegende Bildung ohne Brüche* (S. 183–194). Wiesbaden: Verlag für Sozialwissenschaften.
- Lotz, M. & Gabriel, K. (2012, März). *Niedrig und hoch inferente Verfahren zur Erfassung von Unterrichtsqualität in der Grundschule*. Vortrag im Rahmen des Symposiums „Was konstituiert Unterricht? Grenzgänge zwischen quantitativer und qualitativer Unterrichtsforschung“ (W. Helsper & E. Klieme) auf dem 23. Kongress der Deutschen Gesellschaft für Erziehungswissenschaft, Osnabrück, Deutschland.
- Lotz, M., Gabriel, K. & Lipowsky, F. (2013). Niedrig und hoch inferente Verfahren der Unterrichtsbeobachtung. Analysen zu deren gegenseitiger Validierung. *Zeitschrift für Pädagogik*, 3, 357–379.
- Lovett, S. & Gilmore, A. (2003). Teachers' learning journeys: The Quality Learning Circle as a model of professional development. *School Effectiveness and School Improvement*, 14(2), 189–211.
- Lucas, R. E. & Baird, B. M. (2006). Global self-assessment. In M. Eid & E. Diener (Hrsg.), *Handbook of multimethod measurement in psychology* (S. 29–59). Washington, DC: American Psychological Association.
- Lüders, M. (2010). Methodentraining in der Lehrerbildung und Lehrerfortbildung. Grundriss eines Programms für „effektives Unterrichten“. In F. H. Müller (Hrsg.), *Lehrerinnen und Lehrer Lernen* (S. 345–357). Münster: Waxmann.
- Lüders, M. (2012). „Pädagogisches Unterrichtswissen“ – eine Testkritik. *Zeitschrift für Erziehungswissenschaft*, 15(4), 775–791.
- Lüdtke, O., Trautwein U., Kunter M. & Baumert J. (2006). Analyse von Lernumwelten. Ansätze zur Bestimmung der Reliabilität und Übereinstimmung von Schülerwahrnehmungen. *Zeitschrift für Pädagogische Psychologie*, 20(1/2), 85–96.
- Maag Merki, K. & Grob, U. (2005). Überfachliche Kompetenzen: Zur Validierung eines Indikatorensystems. In A. Frey (Hrsg.), *Kompetenzdiagnostik. Theorien und Methoden zur Erfassung und Bewertung von beruflichen Kompetenzen* (S. 7–30). Landau: Empirische Pädagogik e.V.
- Mabe, P. A. & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, 67(3), 280–296.

- MacSuga, A. S. & Simonsen, B. (2011). Increasing teachers' use of evidence-based classroom management strategies through consultation: Overview and case studies. *Beyond Behavior*, 20(2), 4–12.
- Malone, B. G. & Tietjens, C. L. (2000). Re-examination of classroom rules: The need for clarity and specified behavior. *Special Services in the Schools*, 16(1-2), 159–170.
- Manning, M. L. & Bucher, K. T. (2007). *Classroom management: Models, applications, and cases* (2. Aufl.). Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Marshall, H. H. & Green, J. L. (1977). Stability of classroom variables as measured by a broad range observational system. *The Journal of Educational Research*, 70(6), 304–311.
- Martin, S. D. (2004). Finding balance: Impact of classroom management conceptions on developing teacher practice. *Teaching and Teacher Education*, 20(5), 405–422.
- Martin, N. K. & Yin, Z. (1999). Beliefs regarding classroom management style: Differences between urban and rural secondary level teachers. *Journal of Research in Rural Education*, 15(2), 101–105.
- Martin, N. K., Yin, Z. & Baldwin, B. (1998). Construct validation of the attitudes and beliefs on classroom control inventory. *Journal of Classroom Interaction*, 33(2), 6–15.
- Martin, N. K., Yin, Z. & Mayall, H. (2007). The Attitudes & Beliefs on Classroom Control inventory – revised and revisited: A continuation of construct validation. *Journal of Classroom Interaction*, 42(2), 11–20.
- Marzano, R. J. (2003). *What works in schools*. Alexandria, VA: ASCD.
- Marzano, R. J., Gaddy, B. B., Foseid, M. C., Foseid, M. R. & Marzano, J. S. (2005). *A handbook for classroom management that works*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Mayr, J. (1998). *Fragebögen zur Erkundung des Lehrens und Lernens an der Pädagogischen Akademie*. Linz: Pädagogische Akademie der Diözese.
- Mayr, J. (2006). Klassenführung auf der Sekundarstufe II: Strategien und Muster erfolgreichen Lehrerhandelns. *Schweizerische Zeitschrift für Bildungswissenschaften*, 28(2), 227–242.
- Mayr, J., Eder, F. & Fartacek, W. (2005). *Linzer Diagnosebogen zur Klassenführung (LDK)*. Zugriff am 10.09.2013 von <http://ius.uni-klu.ac.at/projekte/ldk/>.
- McCaslin, M. & Good, T. S. L. (1992). Compliant cognition: The misalliance of management and instructional goals in current school reform. *Educational Researcher*, 21(3), 4–17.

- McGarity, J. R. & Butts, D. P. (1984). The relationship among teacher classroom management behavior, student engagement and student achievement of middle and high school science students of varying aptitude. *Journal of Research in Science Teaching*, 21(1), 55–61.
- McGinnis, J. C., Frederick, B. P. & Edwards, R. (1995). Enhancing classroom management through proactive rules and procedures. *Psychology in the Schools*, 32, 220–224.
- McKnight, K. & Sechrest, L. (2004). Program evaluation. In S. N. Haynes (Hrsg.), *Comprehensive handbook of psychological assessment* (3. Aufl., S. 246–266). Hoboken, NJ: John Wiley & Sons.
- McLaughlin, H. J. (1991). Reconciling care and control: Authority in classroom relationships. *Journal of Teacher Education*, 42(3), 182–195.
- Meister, D. G. & Jenks, C. (2000). Making the transition from preservice to inservice teaching: Beginning teachers' reflections. *Action in Teacher Education*, 22(3), 1–11.
- Meister, D. G. & Melnick, S. A. (2003). National new teacher study: Beginning teachers' concerns. *Action in Teacher Education*, 24(4), 87–94.
- Melnick, S. A. & Meister, D. G. (2008). A comparison of beginning and experienced teachers concerns. *Educational Research Quarterly*, 31(3), 39–56.
- Menting, A. T., Orobio de Castro, B. & Matthys, W. (2013). Effectiveness of the Incredible Years parent training to modify disruptive and prosocial child behavior: A meta-analytic review. *Clinical Psychology Review*, 33(8), 901–913.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543.
- Merkens, H. (2010). *Unterricht: Eine Einführung*. Wiesbaden: Verlag für Sozialwissenschaften.
- Merrett, F. & Wheldall, K. (1993). How do teachers learn to manage classroom behaviour? A study of teachers' opinions about their initial training with special reference to classroom behaviour management. *Educational Studies*, 19(1), 91–106.
- Merritt, M. (1982). Distributing and directing attention in primary classrooms. In L. C. Wilkinson (Hrsg.), *Communicating in the classroom* (S. 223–244). New York, NY: Academic Press.
- Messick, S. (1995). Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist*, 9, 741–749.
- Messner, H. & Reusser K. (2000). Die berufliche Entwicklung von Lehrpersonen als lebenslanger Prozess. *Beiträge zur Lehrerbildung*, 18(2), 157–171.

- Meyer, H. K. (2004). *Theorie und Qualitätsbeurteilung psychometrischer Tests*. Stuttgart: Kohlhammer.
- Mir Pozo, M. L. (2003). An instrument for evaluating classroom management (QCME). *Electronic Journal of Research in Educational Psychology and Psychopedagogy*, 1(1), 67–94.
- Mishra, S. P. (1981). Teaching effectiveness as measured by high- and low-inference rating items. *Journal of Instructional Psychology*, 8(8), 20–24.
- Mitchell, K. J., Robinson, D. Z., Plake, B. S. & Knowles, K. T. (2001). *Testing teacher candidates: The role of licensure tests in improving teacher quality*. Washington, D.C: National Academy Press.
- Mittag, W. & Hager, W. (2000). Ein Rahmenkonzept zur Evaluation psychologischer Interventionsmaßnahmen. In W. Hager & J. L. Patry (Hrsg.), *Handbuch Evaluation psychologischer Interventionsmaßnahmen* (S. 102–128). Bern: Huber.
- Moos, R. H. & Trickett, E. J. (1974). *Classroom Environment Scale: Manual*. Palo Alto, CA: Consulting Psychologists Press.
- Mühlig, S. & Petermann, F. (2006). Grundprinzipien multimethodaler Diagnostik. In F. Petermann (Hrsg.), *Handbuch der Psychologischen Diagnostik* (S. 99–108). Göttingen: Hogrefe.
- Müller, F. H. (Hrsg.). (2010). *Lehrerinnen und Lehrer Lernen*. Münster: Waxmann.
- Mummendey, H. D. & Bolten, H.-G. (1985). Die Impression-Management-Theorie: Theorien der Sozialpsychologie. In D. Frey & M. Irle (Hrsg.), *Motivations- und Informationsverarbeitungstheorien* (S. 57–77). Bern: Huber.
- Munby, H., Russel, T. & Martin, A. (2001). Teachers' knowledge and how it develops. In V. Richardson (Hrsg.), *Handbook of research on teaching* (4. Aufl., S. 877–904). Washington, D.C: American Educational Research Association.
- Nash, R. (2009). *The active classroom: Practical strategies for involving students in the learning process*. Thousand Oaks, CA: Corwin Press.
- Neubauer, W., Gampe, H. & Knapp, R. (1992). *Konflikte in der Schule. Möglichkeiten und Grenzen kooperativer Entscheidungsfindung* (4. Aufl.). Neuwied: Luchterhand.
- Neuenschwander, M. P. (2006a). Editorial zum Themenheft Klassenführung – Konzepte und neue Forschungsbefunde. *Schweizerische Zeitschrift für Bildungswissenschaften*, 28(2), 189–197.
- Neuenschwander, M. P. (2006b). Überprüfung einer Typologie der Klassenführung. *Schweizerische Zeitschrift für Bildungswissenschaften*, 28(2), 243–257.

- Neuenschwander, M. P., Goltz, S., Balmer, T., Gasser, A., Hirt, U., Ryser, H. et al. (2003). *Eltern, Lehrpersonen und Schülerleistungen. Dokumentation des Schülerfragebogens*. Bern: Stelle für Forschung und Entwicklung.
- Nie, Y. & Lau, S. (2009). Complementary roles of care and behavioral control in classroom management: The self-determination theory perspective. *Contemporary Educational Psychology*, 34(3), 185–194.
- Nolle, A. (2004). *Evaluation der universitären Lehrerinnen- und Lehrerausbildung. Erhebung zur pädagogischen Kompetenz von Studierenden der Lehramtsstudiengänge*. München: Meidenbauer.
- Nolting, H.-P. (2009). *Störungen in der Schulklasse: Ein Leitfaden zur Vorbeugung und Konfliktlösung* (8. Aufl.). Weinheim: Beltz.
- Nuangchalerm, P. (2009). Implementing professional experiences to prepare preservice science teachers. *The Social Sciences*, 4(4), 388–391.
- Oelkers, J. & Oser, F. (2001). *Die Wirksamkeit der Lehrerbildungssysteme in der Schweiz*. Chur: Rüegger.
- Oevermann, U. (1996). Theoretische Skizze einer revidierten Theorie professionalisierten Handelns. In A. Combe & W. Helsper (Hrsg.), *Pädagogische Professionalität* (S. 70–182). Frankfurt am Main: Suhrkamp.
- Oliver, R. M., Wehby, J. H. & Reschly, D. J. (2011). *Teacher classroom management practices: Effects on disruptive or aggressive student behavior*. Campbell Systematic Reviews, Zugriff am 10.09.2013 von www.campbellcollaboration.org/lib/download/1189/
- Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of Life Research*, 14(3), 587–598.
- Opendakker, M.-C., Maulana, R. & den Brok, P. (2012). Teacher–student interpersonal relationships and academic motivation within one school year: Developmental changes and linkage. *School Effectiveness and School Improvement*, 23(1), 95–119.
- Ophardt, D. (2006). *Professionelle Orientierungen von Lehrerinnen und Lehrern unter den Bedingungen einer Infragestellung der Vermittlungsfunktion. Eine qualitativ-rekonstruktive Studie an einer Hauptschule im Reformprozess* (Dissertation, Freie Universität Berlin, Universitätsbibliothek), Zugriff von <http://www.diss.fu-berlin.de/2006/509>
- Ophardt, D. & Thiel, F. (2007). Klassenmanagement als professionelle Gestaltungsleistung. In D. Lemmermöhle, M. Rothnagel, S. Bögeholz, M. Hasselhorn & R. Watermann (Hrsg.), *Professionell Lehren. Erfolgreich Lernen* (S. 133–145). Münster: Waxmann.

- Ophardt, D. & Thiel, F. (2008). Klassenmanagement als Basisdimension der Unterrichtsqualität. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungsperspektiven und methodische Zugänge* (2. Aufl., S. 259–284). Wiesbaden: Verlag für Sozialwissenschaften.
- Ophardt, D. & Thiel, F. (2013). *Klassenmanagement – Ein Handbuch für Studium und Praxis*. Stuttgart: Kohlhammer.
- Oser, F. (1997a). Standards in der Lehrerbildung, Teil 1: Berufliche Kompetenzen, die hohen Qualitätsmerkmalen entsprechen. *Beiträge zur Lehrerbildung*, 15(1), 26–27.
- Oser, F. (1997b). Standards in der Lehrerbildung. Teil 2: Wie werden Standards in der schweizerischen Lehrerbildung erworben? Erste Empirische Ergebnisse. *Beiträge zur Lehrerbildung*, 15(2), 210–228.
- Oser, F. (2004). Standardbasierte Evaluation der Lehrerbildung. In P. Reinhold, G. Tulodziecki, J. Wildt & S. Blömeke (Hrsg.), *Handbuch Lehrerbildung* (S. 184–207). Bad Heilbrunn: Klinkhardt.
- Oser, F., Curcio, G.-P. & Düggeli, A. (2007). Kompetenzmessung in der Lehrerbildung als Notwendigkeit – Fragen und Zugänge. *Beiträge zur Lehrerbildung*, 25(1), 14–26.
- Oser, F., Düggeli, A. & Heinzer, S. (2010). Qualitätsmessung von Lehrpersonen-Kompetenzen: Ein neuer Ansatz. In J. Abel & G. Faust (Hrsg.), *Wirkt Lehrerbildung? Antworten aus der empirischen Forschung* (S. 133–153). Göttingen: Waxmann.
- Oswald, F., Pfeiffer, B., Ritter-Berlach, G. & Tanzer, N. (1989). *Schulklima: Die Wirkungen der persönlichen Beziehungen in der Schule*. Wien: Universitätsverlag.
- Paechter, M., Skliris, B. & Macher, D. (2011). Evaluation universitärer Lehre mittels Einschätzungen des subjektiven Kompetenzerwerbs. *Psychologie in Erziehung und Unterricht*, 58, 128–138.
- Patry, J.-L., Perrez, M. (2000). Theorie-Praxis-Probleme und die Evaluation von Interventionsprogrammen. In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Evaluation psychologischer Evaluationsmaßnahmen* (S. 19–40). Bern: Huber.
- Paulu, E. M. (1928). *Differentiated assignments in classroom management*. Oxford, England: Heath.
- Pianta, R. C. & Hamre, B. K. (2006). Classroom management and relationships between children and teachers: Implications for research and practice. In C. M. Evertson & C. S. Weinstein (Hrsg.), *Handbook of classroom management. Research, practice and contemporary issues* (S. 685–710). Mahwah, NJ: Lawrence Erlbaum Associates.

- Pianta, R. C. & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109–119.
- Pigge, F. L. & Marso, R. N. (1997). A seven year longitudinal multi-factor assessment of teaching concerns development through preparation and early years of teaching. *Teaching and Teacher Education*, 13(2), 225–235.
- Pollard, A. (1980). Teacher interests and changing situations of survival threat in primary school classrooms. In P. Woods (Hrsg.), *Teacher Strategies: Explorations in the Sociology of the School*. London: Croom Helm.
- Popper, K. R. (1968). *The logic of scientific discovery*. London: Hutchinson.
- Pratt, C. C., McGuigan, W. M. & Katzev, A. R. (2000). Measuring program outcomes: Using retrospective pretest methodology. *American Journal of Evaluation*, 21(3), 341–349.
- Prenzel, M., Gogolin, I. & Krüger, H.-H. (2007). Editorial zum Themenheft Kompetenzdiagnostik. *Zeitschrift für Erziehungswissenschaft*, 10(8), 5–8.
- Rademacher, J. A., Callahan, K. & Pederson-Seelye, V. A. (1998). How do your classroom rules measure up? Guidelines for developing an effective rule management routine. *Interventions in School and Clinic*, 33, 284–289.
- Rakoczy, K. & Pauli, C. (2006). Hoch inferentes Rating: Beurteilung der Qualität unterrichtlicher Prozesse. In E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lehrerverhalten und mathematisches Verständnis“* (S. 206–233). Frankfurt am Main: Gesellschaft zur Förderung Internationaler Pädagogischer Forschung.
- Rammstedt, B. (2006). Fragebogen. In F. Petermann (Hrsg.), *Handbuch der Psychologischen Diagnostik* (S. 109–117). Göttingen: Hogrefe.
- Reinke, W. M., Sprick, R. & Knight, J. (2009). Coaching classroom management. In J. Knight (Hrsg.), *Coaching. Approaches and perspectives* (S. 91–112). Thousand Oaks, CA: Corwin Press.
- Renkl, A. (2009). Wissenserwerb. In E. Wild & K. Möller (Hrsg.), *Pädagogische Psychologie* (S. 3–26). Heidelberg: Springer.
- Retelsdorf, J. & Südkamp, A. (2012). Professionalisierung von Lehrerinnen und Lehrern: Editorial zum Themenheft. *Zeitschrift für Pädagogische Psychologie*, 26(4), 227–231.
- Richardson, V. & Placier, P. (2001). Teacher change. In V. Richardson (Hrsg.), *Handbook of research on teaching* (4. Aufl., S. 905–947). Washington, D.C: American Educational Research Association.

- Richter, D. (2011). Lernen im Beruf. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 317–325). Münster: Waxmann.
- Roelofs, E., Veenman, S. & Raemaekers, J. (1994). Improving instruction and classroom management behaviour in mixed-age classrooms: Results of two improvement studies. *Educational Studies*, 20(1), 105–126.
- Rogers, B. & Kühn, M. (2013). *Classroom Management: Das Praxisbuch*. Weinheim: Beltz.
- Rogers, C. R. & Freiberg, H. J. (1994). *Freedom to learn* (3. Aufl.). Columbus, OH: Merrill.
- Rogers, D. (2009). The working alliance in teaching and learning: Theoretical clarity and research implications. *International Journal for the Scholarship of Teaching and Learning*, 3(2), 1–9.
- Rost, J. (2000). Allgemeine Standards für die Evaluationsforschung. In W. Hager & J. L. Patry (Hrsg.), *Handbuch Evaluation psychologischer Interventionsmaßnahmen* (S. 21–140). Bern: Huber.
- Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68–78.
- Saldern, M. & von Littig, K. E. (1987). *Landauer Skalen zum Sozialklima (LASSO 4-13): Beiheft mit Anleitung und Normtabellen*. Weinheim: Beltz.
- Scarlett, W. G., Ponte, I. C. & Singh, J. P. (2009). *Approaches to behavior and classroom management: Integrating discipline and care*. Thousand Oaks, CA: Sage Publications.
- Schermelleh-Engel, K. & Schweizer, K. (2007). Multitrait-Multimethod-Analysen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 325–342). Heidelberg: Springer.
- Schmidt, W. H., Blömeke, S. & Tatto, M. T. (2011). *Teacher education matters: A study of middle school mathematics teacher preparation in six countries*. New York, NY: Teachers College Press.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350–353.
- Schneider, C. & Bodensohn, R. (2010). Entwicklung beruflicher Handlungskompetenzen in der ersten Phase der Lehrerbildung. In J. Abel & G. Faust (Hrsg.), *Wirkt Lehrerbildung? Antworten aus der empirischen Forschung* (S. 227–234). Münster: Waxmann.
- Schön, D. A. (1983). *The reflective practitioner. How professionals think in action*. London: Temple Smith.

- Schönbächler, M.-T. (2004a). *Fragebogen zum Unterricht auf der Primarstufe. Fragebogen für Schülerinnen und Schüler*. Bern: Institut für Pädagogik und Schulpädagogik.
- Schönbächler, M.-T. (2004b). *Klassenmanagement auf der Primarstufe. Fragebogen für Lehrpersonen*. Bern: Institut für Pädagogik und Schulpädagogik
- Schönbächler, M.-T. (2006). Inhalte von Regeln und Klassenmanagement. *Schweizerische Zeitschrift für Bildungswissenschaften*, 28(2), 259–271.
- Schönbächler, M.-T. (2008). *Klassenmanagement: Situative Gegebenheiten und personale Faktoren in Lehrpersonen- und Schülerperspektive*. Bern: Haupt.
- Schraw, G. (2006). Knowledge: Structures and processes. In P. Alexander & P. Winne (Hrsg.), *Handbook of educational psychology* (2. Aufl., S. 245–264). San Diego, CA: Academic Press.
- Schwartz, C. E. & Sprangers, M. A. G. (1999). Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Social Science & Medicine*, 48(11), 1531–1548.
- Schwarzer, R. & Hallum, S. (2008). Perceived teacher self-efficacy as a predictor of job stress and burnout: Mediation analyses. *Applied Psychology*, 57(1), 152–171.
- Scott, S. & Quiroz, D. R. (1997). A meta-analysis of interventions to decrease disruptive classroom behavior in public education settings. *School Psychology Review*, 26(3), 333–368.
- Seidel, T. (2009). Klassenführung. In E. Wild & K. Möller (Hrsg.), *Pädagogische Psychologie* (S. 135–148). Heidelberg: Springer.
- Seidel, T., Blomberg, G. & Renkl, A. (2013). Instructional strategies for using video in teacher education. *Teaching and Teacher Education*, 34, 56–65.
- Seidel, T. & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454–499.
- Seifert, A., Hilligus, A. & Schaper, N. (2009). Entwicklung und psychometrische Überprüfung eines Messinstruments zur Erfassung pädagogischer Kompetenzen in der universitären Lehrerbildung. *Lehrerbildung auf dem Prüfstand*, 2(1), 82–103.
- Sekretariat der Ständigen Konferenz der Kulturminister der Länder in der Bundesrepublik Deutschland (2004a). Standards für die Lehrerbildung. Bericht der Arbeitsgruppe. Bonn: KMK.

- Sekretariat der Ständigen Konferenz der Kulturminister der Länder in der Bundesrepublik Deutschland (2004b). Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der KMK vom 16.12.2004. Bonn: KMK.
- Sergiovanni, T. J. (1994). *Building community in schools*. San Francisco, CA: Jossey-Bass.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23.
- Shulman, L. S. (2002). Making differences: A table of learning. *Change*, 34(6), 36–44.
- Simonsen, B., Fairbanks, S., Briesch, A., Myers, D. & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education and Treatment of Children*, 31(3), 351–380.
- Slavin, R. (1997). *Educational psychology* (5. Aufl.). Boston: Allyn and Bacon.
- Slavin, R. E., Hurley, E. A. & Chamberlain, A. (2003). Cooperative learning and achievement: Theory and research. In W. M. Reynolds & G. E. Miller (Hrsg.), *Handbook of psychology: Educational psychology*. New York, NY: Wiley.
- Smith, L. R. (1985). A low-inference indicator of lesson organization. *The Journal of Classroom Interaction*, 21(1), 25–30.
- Spilt, J., Koomen, H. & Thijs, J. (2011). Teacher wellbeing: The importance of teacher-student relationships. *Educational Psychology Review*, 23(4), 457–477.
- Stallings, J. (1980). Allocated academic learning time revisited, or beyond time on task. *Educational Researcher*, 9(11), 11–16.
- Staub, F. C. (2004). Fachspezifisch-Pädagogisches Coaching: Ein Beispiel zur Entwicklung von Lehrerfortbildung und Unterrichtskompetenz als Kooperation. *Zeitschrift für Erziehungswissenschaft*, 7(3), 113–141.
- Steyer, R. (1992). *Theorie kausaler Regressionsmodelle*. Stuttgart: Fischer.
- Steyer, R., Eid, M. & Schwenkmezger, P. (1997). Modeling true intraindividual change: True change as a latent variable. *Methods of Psychological Research - online*, 2(1), 21–33.
- Steyer, R., Schmitt, M. & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality*, 13(5), 389–408.
- Stough, L. (2006). The place of classroom management in the standards in teacher education. In C. M. Evertson & C. S. Weinstein (Hrsg.), *Handbook of classroom management. Research, practice, and contemporary issues* (S. 909–924). Mahwah, NJ: Lawrence Erlbaum.

- Taubert, R. T. (2007). *Classroom management: Sound theory and effective practice* (4. Aufl.). Westport: Praeger.
- Tausch, R. & Tausch, A.-M. (1998). *Erziehungs-Psychologie: Begegnung von Person zu Person* (11. Aufl.). Göttingen: Hogrefe.
- Tennstädt, K.-C. (1987). *Konstanzer Trainingsmodell (KTM). Theoretische Grundlagen, Beschreibung der Trainingsinhalte und erste empirische Überprüfung*. Bern: Huber.
- Terhart, E. (2002a). *Standards für die Lehrerbildung: Eine Expertise für die Kultusministerkonferenz*. Münster: ZKL.
- Terhart E. (2002b). Standards und Kompetenzen in der Lehrerbildung. In A. Hilligus & H.-D Rinkens (Hrsg.), *Standards und Kompetenzen – neue Qualität in der Lehrerausbildung?* (S. 29–42). Paderborn: Lit-Verlag.
- Terhart, E. (2007). Erfassung und Beurteilung der beruflichen Kompetenz von Lehrkräften. In M. Lüders & J. Wissinger (Hrsg.), *Forschung zur Lehrerbildung. Kompetenzentwicklung und Programmevaluation* (S. 37–62). Münster: Waxmann.
- Terhart, E. (2012). Wie wirkt Lehrerbildung? Forschungsprobleme und Gestaltungsfragen. *Zeitschrift für Bildungsforschung*, 2(1), 3–21.
- Thiel, F., Blüthmann, I. & Watermann, R. (2012). Konstruktion eines Fragebogens zur Erfassung der Lehrkompetenz (LeKo). In B. Berendt, H. Voss & J. Wildt (Hrsg.), *Neues Handbuch Hochschullehre* (S. 1–27). Berlin: Raabe.
- Thiel, F., Richter, S. G. & Ophardt, D. (2012). Steuerung von Übergängen im Unterricht. Eine Experten-Novizen-Studie zum Klassenmanagement. *Zeitschrift für Erziehungswissenschaft*, 4, 727–752.
- Toman, H. (2007). *Classroom-Management. Praxishilfen für das Classroom-Management*. Baltmannsweiler: Schneider.
- Toste, J. R., Heath, N. & Dallaire, L. (2010). Perceptions of classroom working alliance and student performance. *Journal of Educational Research* 56(4), 371–387.
- Ursano, A. M., Kartheiser, P. H. & Ursano, R. J. (2007). The teaching alliance: A perspective on the good teacher and effective learning. *Psychiatry*, 70, 187–194.
- Van de Grift, W. (2007). Quality of teaching in four European countries: A review of the literature and application of an assessment instrument. *Educational Research*, 49(2), 127–152.
- Van Es, E. A. & Sherin, M. G. (2006). How different video club designs support teachers in “learning to notice”. *Journal of Computing in Teacher Education*, 22 (4).

- Van Es, E. A. & Sherin, M. G. (2010). The influence of video clubs on teachers' thinking and practice. *Journal of Mathematics Teacher Education*, 13(2), 155–176.
- Van Tartwijk, J., den Brok, P., Veldman, I. & Wubbels, T. (2009). Teachers' practical knowledge about classroom management in multicultural classrooms. *Teaching and Teacher Education*, 25(3), 453–460.
- Van Tassel-Baska, J., van Tassel-Baska, J., Quek, C. & Feng, A. X. (2006). The development and use of a structured teacher observation scale to assess differentiated best practice. *Roeper Review: A Journal on Gifted Education*, 29(2), 84–92.
- Vandenberg, R. J. & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70.
- Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research*, 54(2), 143–178.
- Vescio, V., Ross, D. & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education*, 24(1), 80–91.
- Voss, T. & Kunter, M. (2011). Pädagogisch-psychologisches Wissen von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 193–214). Münster: Waxmann.
- Voss, T., Kunter, M. & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, 103(4), 952–969.
- Wagner, W. (2008). *Methodenprobleme bei der Analyse der Unterrichtswahrnehmung aus Schülersicht*. (Dissertation, Universität Koblenz-Landau), Zugriff von http://www.fachportal-paedagogik.de/fis_bildung/suche/fis_set.html?Fid=880436
- Wahlberg, H. J. (1966). *Classroom Climate Questionnaire*. Cambridge, MA: Harvard University.
- Walker, H. M., Ramsey, E. & Gresham, F. (2004). *Antisocial behaviour in school: Strategies and best practices* (2. Aufl.). Pacific Grove, CA: Brooks Cole.
- Wang, M. C., Haertel, G. D. & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63(3), 249–294.

- Wannack, E., Herger, K., Gruber, M. & Barblan, A. (2009). *Classroom Management in der Schuleingangsstufe. Dokumentation der Fragebogenstudie*. Zugriff am 10.09.2013 von <http://www.schuleingangsstufe.phbern.ch>
- Webster-Stratton, C., Reid, M. J. & Hammond, M. (2001). Preventing conduct problems, promoting social competence: A parent and teacher training partnership in head start. *Journal of Clinical Child Psychology*, 30(3), 283–302.
- Weekley, J. A. & Ployhart, R. E. (2006). *Situational judgment tests: Theory, measurement, and application*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Weinert, F. E. (1996). Lerntheorien und Instruktionsmodelle. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie.. Psychologie des Lernens und der Instruktion* (Bd. 2, S. 1–48). Göttingen: Hogrefe.
- Weinert, F. E. (1999). *Concepts of competence. Contribution within the OCED-Project Definition and Section of Competencies: Theoretical and conceptual foundations (DeSeCo)*. Neuchâtel: Bundesamt für Statistik.
- Weinert, F. E. (2001a). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Hrsg.), *Defining and selecting key competencies* (S. 45–65). Göttingen: Hogrefe.
- Weinert, F. E. (2001b). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessung in Schulen* (S. 17–32). Weinheim: Beltz.
- Weinert, F. E. & Schrader, F.-W. & Helmke, A. (1990). Unterrichtsexpertise – ein Konzept zur Verringerung der Kluft zwischen zwei theoretischen Paradigmen. In L.-M. Alisch, J. Baumert & K. Beck (Hrsg.), *Professionswissen und Professionalisierung* (S. 173–206). Braunschweig: Colmsee.
- Weinstein, C. S. (1991). The classroom as a social context for learning. *Annual Review of Psychology*, 42, 493–525.
- Weinstein, C. S. (1998). “I want to be nice, but I have to be mean”: Exploring prospective teachers’ conceptions of caring and order. *Teaching and Teacher Education*, 14(2), 153–163.
- Weinstein, C. S. (1999). Reflections on best practices and promising programs: Beyond assertive classroom discipline. In H. J. Freiberg & J. E. Brophy (Hrsg.), *Beyond behaviorism. Changing the classroom management paradigm* (S. 147–163). Boston: Allyn and Bacon.

- Weinstein, C. S. (2007). *Middle and secondary classroom management: Lessons from research and practice*. Boston: Mc Graw Hill.
- Wertheim, E., Love, A., Peck, C. & Littlefield, L. (1998). *Skills for resolving conflict: A cooperative problem solving approach*. Melbourne, Australia: Eruditions.
- West, L. & Staub, F. C. (2003). *Content-Focused Coaching. Transforming mathematics lessons*. Portsmouth, NH: Heinemann.
- Westerman, D.A. (1991). Expert and novice teacher decision making. *Journal of Teacher Education*, 42(4), 292–305.
- Wettstein, A. (2010). Lehrpersonen in schwierigen Unterrichtssituationen unterstützen: Ein pädagogisch-didaktisches Coaching zur Prävention von Unterrichtsstörungen. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete*, 79, 145–157.
- Wheldall, K., Merrett, F. (1988). Which classroom behaviours do primary school teachers say they find most troublesome? *Educational Review*, 40, 13–27.
- Widaman, K. F. & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In M. T. W. S. G. W. Kendall & J. Bryant (Hrsg.), *The science of prevention methodological advances from alcohol and substance abuse research* (S. 281–324). Washington, DC: American Psychological Association.
- Wilson, D. B. & Lipsey, M. W. (2001). The role of method in treatment effectiveness research: Evidence from meta-analysis. *Psychological Methods*, 6(4), 413–429.
- Witt, J. C., van der Heyden, A. M. & Gilbertson, D. (2004). Instruction and classroom management: Prevention and intervention research. In R. B. Rutherford (Hrsg.), *Handbook of research in emotional and behavioral disorders* (S. 426–445). New York, NY: Guilford Press.
- Wong, H. & Wong, R. (2004). *The first days of school. How to be an effective teacher*. Mountain View, CA: Wong.
- Woolfolk, A. E. & Brooks, D. M. (1985). Beyond words: The influence of teachers' nonverbal behaviors on students' perceptions and performances. *Elementary School Journal*, 85, 513–528.
- Wottawa, H. & Thierau, H. (1998). *Lehrbuch Evaluation* (2. Aufl.). Bern: Huber.
- Wubbels, T. & Brekelmans, M. (2005). Two decades of research on teacher-student relationships in class. *International Journal of Educational Research*, 43(1-2), 6–24.

Wubbels, T., den Brok, P., Veldman, I. & van Tartwijk, J. (2006). Teacher interpersonal competence for Dutch secondary multicultural classrooms. *Teachers and Teaching: Theory and Practice*, 12, 407–433.

Wubbels, T. & Levy, J. (1993). The Questionnaire on Teacher Interaction. In T. Wubbels & J. Levy (Hrsg.), *Do you know what you look like? Interpersonal relations in education* (S. 163–166). London: The Falmer Press.

2

Teilstudie 1: Multidimensionale Erfassung von Kompetenzen im Klassenmanagement. Konstruktion und Validierung eines Beobachter- und eines Schülerfragebogens für die Sekundarstufe.

Anmerkung: Dies ist die Version der Autorin von einer Arbeit die in der Zeitschrift für Pädagogische Psychologie veröffentlicht wurde (© by Hans Huber). Diese Artikelfassung entspricht nicht vollständig dem in der Zeitschrift veröffentlichten Artikel. Dies ist nicht die Originalversion des Artikels und kann daher nicht zur Zitierung herangezogen werden.

Piwowar, V. (2013). Multidimensionale Erfassung von Kompetenzen im Klassenmanagement: Konstruktion und Validierung eines Beobachter- und eines Schülerfragebogens für die Sekundarstufe. *Zeitschrift für Pädagogische Psychologie*, 27(4), 215-228.

[doi:10.1024/1010-0652/a000108](https://doi.org/10.1024/1010-0652/a000108)

Zusammenfassung

Klassenmanagement gilt als zentrales Merkmal guten Unterrichts. Psychometrisch zufriedenstellende Instrumente, die eine differenzierte Erfassung von Kompetenzen im Klassenmanagement ermöglichen, sind im deutschen Sprachraum bislang jedoch kaum etabliert. Die vorliegende Arbeit stellt einen Beobachter- und einen Schülerfragebogen zur mehrdimensionalen Erfassung von Klassenmanagement vor, die in einer Gelegenheitsstichprobe von 88 Sekundarschullehrkräften implementiert wurden. Sowohl hinsichtlich verschiedener Reliabilitätsmaße als auch in ersten Validierungsanalysen konnten gute bis sehr gute Testeigenschaften der Instrumente nachgewiesen werden; es zeigten sich Zusammenhänge mit dem Störverhalten und der Beteiligung von Schülern. Explorative Faktorenanalysen extrahierten fünf von neun theoretisch abgeleiteten Klassenmanagementstrategien im individuellen Schülerurteil, die übrigen vier Strategien fielen zu zwei Faktoren zusammen. In der geteilten Wahrnehmung wurden drei übergeordnete Dimensionen differenziert. In weiterführenden Studien sollten die Sensitivität und prädiktive Validität der Instrumente überprüft werden.

Schlagwörter: Klassenmanagement – Schüler – Beobachtung – Fragebogen – Multilevel EFA

Title: The Multidimensional Assessment of Classroom Management Competencies: Construction and Validation of an Observer Rating and a Student Rating for Secondary Teachers.

Abstract

Classroom management is a critical component of effective teaching. Psychometrically sound instruments that allow the differentiated assessment of the diverse aspects of classroom management are rare in Germany yet. This article introduces two newly developed questionnaires for a multidimensional assessment of classroom management: a student rating and an observer rating, which both were implemented in a sample of 88 secondary teachers. Reliability coefficients as well as analysis of validity revealed good to excellent psychometric properties; assessed classroom management competencies were associated with student misbehavior and student engagement. Exploratory factor analyses of individual student ratings confirmed five out of nine theoretically derived classroom management strategies. Aggregated student ratings still distinguished between three broader categories. Subsequent research should investigate the instruments' sensitivity and predictive validity.

Keywords: classroom management – secondary students – observation – questionnaire – evaluation – multilevel efa.

2. Multidimensionale Erfassung von Kompetenzen im Klassenmanagement. Konstruktion und Validierung eines Beobachter- und eines Schülerfragebogens für die Sekundarstufe.

2.1. Einleitung

Zentraler Bestandteil der Entwicklung und empirischen Erforschung von Lehrerkompetenzen ist deren systematische und theoriegeleitete Erfassung auf Basis objektiver, reliabler und valider Instrumente. Erst die differenzierte Diagnose des Könnens von Lehrkräften kann eine fundierte Analyse und Reflexion des Lehrerhandelns initiieren und zur Professionalisierung von Lehrkräften beitragen (Ophardt & Thiel, 2013).

Klassenmanagement gilt als ein Kernbereich professioneller Expertise von Lehrkräften und stellt eine Basisdimension qualitätsvollen Unterrichts dar (Brophy & Good, 1986; Kunter et al., 2005): Nicht nur schulisches und sozio-emotionales Lernen von Schülern (Seidel & Shavelson, 2007; Wang, Haertel & Walberg, 1993), sondern auch Selbstwirksamkeitserwartungen (Brouwers & Tomic, 2000) und das berufliche Wohlbefinden von Lehrkräften (Friedman, 2006) sind positiv mit Kompetenzen im Klassenmanagement assoziiert. Lehrkräfte selbst sehen im Bereich des Klassenmanagements den größten Entwicklungsbedarf, insbesondere am Beginn ihrer Karriere (Pigge & Marso, 1997).

Nachdem das Thema Klassenmanagement in der deutschsprachigen Unterrichtsforschung lange Zeit nur eine randständige Bedeutung hatte (Ophardt & Thiel, 2013), wird es mittlerweile als Teilbereich der Lehrerkompetenz im Rahmen groß angelegter Studien mit erhoben (z. B. in COACTIV, TIMSS oder TEDS-M). Die Erfassung verschiedener Facetten von Klassenmanagement ist mit den bisher entwickelten Instrumenten jedoch nur eingeschränkt möglich: Die Operationalisierung erfolgt in der Regel eindimensional im Sinne von Zeitnutzung und Störungsfreiheit oder beschränkt sich auf die von Kounin (1970) beschriebenen Strategien erfolgreichen Klassenmanagements. Dem aktuellen internationalen Forschungsstand (vgl. Evertson & Weinstein, 2006a) sowie der zunehmend auch im deutschen Sprachraum aufgegriffenen, umfassenderen Konzeption von Klassenmanagement (z. B. Mayr, 2006; Ophardt & Thiel, 2013 oder Schönbächler, 2008) kann damit nicht gerecht werden.

Ziel der vorliegenden Arbeit war die Entwicklung und Überprüfung eines deutschsprachigen Instrumentes zur multidimensionalen Beschreibung von Kompetenzen im Klassenmanagement von Lehrkräften der Sekundarstufe in Form standardisierter Fragebögen. Ein Vorteil von Fragebögen zur Erfassung von Unterrichtsprozessen liegt im Vergleich zu Kompe-

tenztests (z. B. Voss, Kunter & Baumert, 2011) darin, dass solche Fragebögen direkte Hinweise für die Unterrichtsentwicklung geben können (Clausen, 2002). Unter Rückgriff auf bereits existierende Instrumente sowie auf Basis empirischer Studien wurden theoriegeleitet ein Beobachter- und ein Schülerfragebogen entwickelt, die im vorliegenden Beitrag hinsichtlich ihrer Gütekriterien charakterisiert werden.¹

2.1.1. Kompetenzen des Klassenmanagements

Angelehnt an Shulman (1987) wird Klassenmanagement als Teil des pädagogisch-psychologischen Wissens von Lehrkräften verstanden. Zur Konzeptualisierung des Konstrukts werden unterschiedliche Forschungstraditionen integriert: das ökologische (Doyle, 2006), das behavioristische (Landrum & Kauffmann, 2006) und das handlungstheoretische Paradigma (Evertson & Harris, 1992). Klassenmanagement wird demnach definiert als “the actions teachers take to create an environment that supports and facilitates both academic and social-emotional learning” (Evertson & Weinstein, 2006b, S. 4). Orientiert an der im *Handbook of Classroom Management* für Lehrerüberzeugungen dargestellten Unterteilung in *behavior management*, *instructional management* und *people management* (Woolfolk Hoy & Weinstein, 2006, vgl. auch Martin & Yin, 1999) werden nachfolgend drei Dimensionen von Klassenmanagement postuliert: *Management von Schülerverhalten*, *Management der Instruktion*, *Management der Lehrer-Schüler-Beziehung*. Danach dient Klassenmanagement nicht nur der Kontrolle des Schülerverhaltens, sondern hat eine unmittelbare Stützfunktion für die Instruktion (Doyle, 2006; Ophardt & Thiel, 2013): Durch instruktionale Klassenmanagementstrategien können Lehrkräfte gezielt die Aufmerksamkeit und lernbezogenen Interaktionen der Schüler orientieren (Haag & Streber, 2012; Weinstein, 1999) und so direkt die aktive Lernzeit („time on task“) der Schüler erhöhen (Seidel, 2009). Sowohl verhaltens- als auch instruktionsbezogenes Klassenmanagement wird wesentlich beeinflusst von der funktionalen Qualität der Lehrer-Schüler-Beziehung (Freiberg, 1999; Mayr, 2006). Nur eine Lernumgebung, die durch gegenseitigen Respekt, Wertschätzung und Fürsorge gekennzeichnet ist, ermöglicht, dass Störungsprävention und -intervention über eine temporäre Disziplinierung der Schüler hinaus nachhaltig lernwirksam werden (Nie & Lau, 2009; van Tartwijk, den Brok, Veldman & Wubbels, 2009). Umgekehrt wird beim Einsatz instruktionaler und verhaltensorientierter Strategien die Beziehung von Lehrkraft und Schülern kontinuierlich neu kommuniziert und ausgehandelt, wie bei Zurechtweisungen oder der Regelarbeit.

¹ Aus Gründen der besseren Lesbarkeit sind nachfolgend mit Beobachtern immer Beobachterinnen und Beobachter sowie mit Schülern immer Schülerinnen und Schüler gemeint.

Die vorgetragenen Argumente sprechen dafür, in Anlehnung an die aktuellen internationalen Untersuchungen den in der deutschsprachigen Unterrichtsforschung bis auf die genannten Ausnahmen relativ eng gefassten Begriff des Klassenmanagements auszuweiten und die skizzierten drei Dimensionen in die Operationalisierung eines entsprechenden Konstrukts einzubeziehen. Auf der konkreten Handlungsebene werden reaktive, aber auch störungspräventive bzw. proaktive Strategien wirksam, um den intendierten Unterrichtsplan–das Handlungsprogramm (Doyle, 2006)–erfolgreich zu implementieren. Den Dimensionen werden in der vorliegenden Studie folgende Subdimensionen (bzw. Strategien, theoretisch und Skalen, empirisch) zugeordnet (nachfolgend in ihrer idealtypischen Ausprägung beschrieben):

A) *Management von Schülerverhalten*

- *Regeln*: Es existiert ein Regelsystem, auf das die Lehrkraft erfolgreich Bezug nehmen kann. Die Schüler kennen die vereinbarten Regeln und halten sich im Wesentlichen daran (Emmer, Evertson & Worsham, 2003).
- *Störungsintervention*: Die Lehrkraft wählt den richtigen Zeitpunkt für Interventionen zum Umgang mit Störungen, sie interveniert rasch und effizient ohne den Unterrichtsfluss zu unterbrechen. Zurechtweisungen sind der Störung angemessen und schonen das Arbeitsbündnis (Emmer et al., 2003; Kounin, 1970).
- *Monitoring*: Die Lehrkraft registriert stets alle Schüleraktivitäten, selbst wenn sie mit anderen Unterrichtsaktivitäten beschäftigt ist. So werden frühzeitig abgelenkte oder unaufmerksame Schüler identifiziert und die Lehrkraft kann rechtzeitig entscheiden, ob und wie sie darauf reagiert (Doyle, 2006; Kounin, 1970).

B) *Management der Instruktion*

- *Prozeduren*: Es gibt etablierte und einfach abrufbare Prozeduren (auch Routinen oder Rituale genannt), die nach entsprechender Signalgebung relativ automatisch bzw. selbstständig von den Schülern ausgeführt werden, ohne den Unterrichtsfluss zu unterbrechen (z. B. das Verteilen von Arbeitsmaterialien oder das Zusammenfinden bei Gruppenarbeit, Leinhardt, Weidman & Hammond, 1987).
- *Gruppenmobilisierung*: Die Lehrkraft fokussiert die Klasse als Ganzes und maximiert den Radius der am Unterricht beteiligten Lernenden, indem sie jeden einzelnen Schüler aktiv in den Unterricht einbezieht (Doyle, 2006; Kounin, 1970).
- *Zeitmanagement*: Unterrichtsphasen und -stunden sind so abgestimmt, dass die gestellten Aufgaben in der vorgegebenen Zeit erledigt werden können und die zur Verfügung stehende Unterrichtszeit effizient genutzt wird (Weinstein, 2007).

- *Klarheit des Handlungsprogramms*: Sowohl der Stundenverlauf im Ganzen als auch einzelne Aufgaben sind klar strukturiert und für die Schüler gut nachvollziehbar (Mayr, 2006), so dass das Handlungsprogramm kontinuierlich stabilisiert wird (Thiel, Richter & Ophardt, 2012).

C) Management der Lehrer-Schüler-Beziehung

- *Arbeitsbündnis*: Die Lehrkraft tritt ihren Schülern empathisch, fürsorglich und unterstützend gegenüber. Gleichzeitig werden hohe akademische und soziale Erwartungen an die Schüler expliziert. Nur so wird aus einer guten Lehrer-Schüler-Beziehung ein auf Lernen orientiertes Bündnis, das Leistung, Motivation und Selbstregulation von Schülern fördert (Nie & Lau, 2009; Oevermann, 1996).
- *Umgang mit Konflikten unter Schülern*: Die Lehrkraft beteiligt sich an der Lösung von Konflikten unter Schülern auf eine faire und konstruktive Weise (Ophardt & Thiel, 2013; Walker, Ramsey & Gresham, 2004). Die Konfliktlösung ist dabei immer an der Stabilisierung eines langfristigen und lerndienlichen Arbeitsbündnisses orientiert.

Die Zuordnung der Strategien zu den Dimensionen dient vorrangig der Systematisierung und ist nicht als disjunkt zu betrachten. So ist z. B. Monitoring auch Teil des instruktionalen Managements; Prozeduren steuern zugleich auch das Schülerverhalten. Auch hier gilt – wie für alle Teilaspekte der Unterrichts: Die einzelnen Handlungsstrategien (Mayr, 2006) bzw. Kompetenzen des Klassenmanagements sind zwar theoretisch klar abgrenzbare Konstrukte, in der Praxis sind sie aber häufig miteinander assoziiert: Ein starkes Arbeitsbündnis stützt die Störungsintervention; Prozeduren sind Teil eines effektiven Zeitmanagements.

Wie in ökologischen Studien dargestellt wurde, erfolgt die Orchestrierung dieser einzelnen Kompetenzen stets kontextabhängig (Doyle, 2006). Im Sekundarbereich steht daher weniger das Einüben von Verhaltensroutinen oder die Organisation und Vorbereitung von Materialien im Vordergrund, sondern vielmehr die Motivierung von Schülern, der soziale Umgang untereinander und die Bearbeitung von Konflikten (Weinstein, 2007). Insbesondere der Umgang mit Konflikten ist bislang jedoch selten operationalisiert worden, trotz seiner zentralen Bedeutung für Klassenmanagement.

2.1.2. Erfassung von Klassenmanagement

Fragebögen erlauben die proximale und direkte Erfassung professioneller Kompetenz (Kunter & Klusmann, 2010). Dabei stehen Lehrkräfte (Selbstbericht), sowie Schülerinnen und Schüler und externe Beobachterinnen und Beobachter (Fremdbericht) als Datenquellen zur

Verfügung. Wie Clausen (2002) herausgestellt hat, können sich die Einschätzungen dieser Perspektiven, abhängig von der Beobachtbarkeit, der Verständnisanforderungen und dem Evaluationsgehalt der einzuschätzenden Merkmale wesentlich voneinander unterscheiden. Durch Multi-Trait-Multi-Method-Analysen wies er nach, dass jede Perspektive eine spezifische Validität aufweist, der unterschiedliche methodische wie inhaltliche Stärken und Schwächen inhärent sind. Im vorliegenden Beitrag werden daher ein Beobachter- und ein Schülerfragebogen dargestellt.

Obwohl sich in der angloamerikanischen Forschung eine breite Auffassung von Klassenmanagement durchgesetzt hat, gibt es nur wenige Beobachtungsinstrumente, die explizit eine mehrdimensionale Erfassung vornehmen (z. B. Evertson, 1985). Zudem werden entsprechende Instrumente eher ad hoc und für eine konkrete Fragestellung entwickelt. Etablierte und psychometrisch anspruchsvolle Inventare zur Unterrichtsbeobachtung erfassen Klassenmanagement–international wie national–eher als Subaspekt der Unterrichtsqualität (z. B. Helmke, Helmke, Schrader & Wagner, 2007; Pianta & Hamre, 2009; Rakoczy & Pauli, 2006). Differenziertere Konzeptualisierungen wurden in Instrumenten der Projekte BIJU (Clausen, 2002) und PERLE (Lotz & Gabriel, 2012) für die Beurteilung videografierten Unterrichts zu Grunde gelegt.

Da die Beurteilung von Klassenmanagement–im Vergleich zu anderen Bereichen des Lehrerhandelns–wenig didaktisch-pädagogisches Wissen erfordert, gilt das Schülerurteil hier als besonders valide (Clausen, 2002). Vorliegende Schülerfragebögen bilden jedoch entweder das Klassen- bzw. Lernklima ab (z. B. Moos & Tricket, 1974; Saldern & Littig, 1987) oder fokussieren Disziplin und Zeitnutzungsaspekte (z. B. Baumert et al. 2009; Helmke et al., 2007). Mayr, Eder und Fartacek (2002) legten erstmals eine dreidimensionale Struktur mit Subfacetten zu Grunde (*Linzer Diagnosebogen zur Klassenführung: Gestaltung des Unterrichts, Förderung der Beziehung, Kontrolle des Verhaltens*), die von Schönbächler (2008) aufgegriffen und weiterentwickelt wurde. Die psychometrische Güte beider Instrumente ist jedoch noch nicht zufriedenstellend und die Konzeptualisierung zum Teil uneinheitlich (z. B. bezüglich des Abstraktionsgrades der Subskalen, vgl. Schönbächler, 2008). Auch Gruehn (2000) bildete in der Unterrichtsbeschreibung durch Schüler Klassenmanagement umfassender ab.

2.1.3. Ableitung des Untersuchungsgegenstandes

Ziel der vorliegenden Arbeit war die Entwicklung und empirische Prüfung eines Beobachter- (*B*) und eines Schülerfragebogens (*S*) für die Sekundarstufe, die die standardisierte

Beurteilung verschiedener Kompetenzen des Klassenmanagements (*KODEK*) anforderungsbezogen abbilden: *KODEK-B* und *KODEK-S*. Die Items wurden nicht gespiegelt, sondern perspektivenspezifisch formuliert.

Folgende Annahmen werden bezüglich der entwickelten Instrumente formuliert:

- *Geteilte Wahrnehmung*: Die von Schülern beurteilten Kompetenzen im Klassenmanagement erfassen eine geteilte Wahrnehmung der Lernumgebung: Es liegt sowohl hinreichende Übereinstimmung innerhalb von Klassen, als auch hinreichend Variabilität zwischen den Klassen vor.
- *Konvergenz von Schüler- und Beobachterbeurteilung*: Es liegen hinreichende Übereinstimmungen (mindestens mittlere Korrelationen) in der Beurteilung der verschiedenen Kompetenzen im Klassenmanagement zwischen Schülern und Beobachtern vor.
- *Zusammenhänge mit konvergenten Kriterien*: Die dargestellten Zusammenhänge von erfolgreichem Klassenmanagement und erhöhter Beteiligung sowie weniger unaufmerksamen und Störverhaltens (off-task Verhalten) lassen sich für alle operationalisierten Kompetenzen identifizieren (mindestens mittlere Korrelationen).
- *Zusammenhänge mit diskriminanten Kriterien*: Die Lehrerselbstwirksamkeit weist nicht mehr als mittlere positive, das Ausmaß der beruflichen Belastung von Lehrkräften nicht mehr als mittlere negative Zusammenhänge mit dem Klassenmanagement auf.
- *Internale Struktur des Schülerfragebogens*: Die abgeleiteten Kompetenzen des Klassenmanagements lassen sich für die Schülerdaten sowohl auf Individualebene als auch auf Klassenebene empirisch identifizieren.
- *Differenzierbarkeit der einzelnen Kompetenzen des Klassenmanagements*: Die operationalisierten Skalen weisen unterschiedlich starke Beziehungen mit den genannten Drittvariablen (off-task Verhalten, Schülerbeteiligung, berufliche Belastung, Lehrerselbstwirksamkeit) auf.

2.2. Methode

2.2.1. Entwicklung und Pilotierung der Fragebögen KODEK-B und KODEK-S

Auf Basis einer theoretischen und empirischen Merkmalsanalyse wurden relevante Kompetenzfacetten entsprechend ihrer Beobachtbarkeit und Auftretenshäufigkeit im Unterricht sowie unter Berücksichtigung hinreichender theoretischer Differenzierbarkeit abgeleitet und systematisiert, wie sie bereits in Kapitel 2.1.1 beschrieben wurden. Zusätzlich vervoll-

ständig ein allgemeiner Indikator die Beurteilung des Klassenmanagements, der nicht als Handlungsstrategie der Lehrperson konzeptualisiert ist, sondern einen allgemeinen Eindruck von der Klassensituation als Ganzem abbildet:

- *Störungsausmaß* (invers): Das Störungsausmaß betrifft den Umfang, die Intensität und die Häufigkeit von Störungen, die die Schüler vom Lernen abhalten und den Lehrer unter Umständen dazu zwingen, den Unterrichtsgegenstand außer Acht zu lassen.

Für die Fragebogenvorformen wurden relevante Items aus vorliegenden Instrumenten den genannten Skalen zugeordnet, ggf. übersetzt und angepasst sowie neue Items formuliert. Hierbei wurde darauf geachtet, dass Items intersubjektiv beobachtbares Verhalten repräsentieren und auch mit eingeschränkten Sprachkompetenzen valide von Schülern beurteilt werden können. Die Items werden auf einer Rating-Skala mit geschlossenem Antwortformat beurteilt und wurden so konstruiert, dass die Beurteilung unabhängig vom Fach und den vorkommenden Aktivitätsstrukturen der jeweiligen Stunde erfolgen kann.

Für den Beobachterfragebogen KODEK-B wurden hoch-inferente, durch Indikatoren spezifizierte Items gebildet. Skalen wurden nur dann durch mehrere Items repräsentiert, wenn das entsprechende Konstrukt durch verschiedene Facetten abgebildet werden sollte. Die Fragebogenvorform enthielt 22 Items, die auf einem sechsstufigen, bipolaren Antwortformat beurteilt werden (1 [*trifft überhaupt nicht zu*] bis 6 [*trifft absolut zu*]); eine vieldeutige Mittelkategorie wurde damit ausgeschlossen. Zusätzlich gibt es die Kategorie „nicht beobachtbar“ für den Fall, dass die Einschätzung aufgrund der gesehenen Unterrichtsstunde nicht möglich ist. Die Skala *Umgang mit Konflikten unter Schülern* wurde in KODEK-B aus zwei Gründen nicht operationalisiert: Zum einen aufgrund der potentiell geringen Basisrate von Konflikten im Unterricht (Konflikte treten häufig vor und nach dem Unterricht bzw. in Pausen auf) und zum anderen, weil auch die Konfliktlösung häufig außerhalb des Unterrichts stattfindet (z. B. in moderierten Konfliktgesprächen). KODEK-B wurde an 20 Unterrichtsvideos, die im Rahmen einer Experten-Novizen-Studie in den Klassenstufen 7 bis 9 im Deutsch- und Mathematikunterricht erstellt wurden (Thiel et al., 2012), durch zwei unabhängige Beobachter pilotiert. Die Videos wurden einmalig und ohne Unterbrechungen betrachtet (mit simultaner Anzeige einer Schüler- und einer Lehrerkamera), um der angedachten Anwendungssituation in vivo maximal zu entsprechen. Items mit niedriger Interrater-Reliabilität oder häufiger Nichtbeantwortung wurden entfernt, umformuliert oder die Indikatoren erweitert und ergänzt. Die Endform von KODEK-B enthält 18 Items, wobei jeweils ein bis drei Items eine Skala repräsentieren.

Für den Schülerfragebogen KODEK-S wurden Items mit niedrigerem Inferenzgrad konstruiert. Im Rahmen einer Expertendiskussion ($N = 8$) wurden die Items vorab auf ihre sprachliche Angemessenheit sowie die inhaltliche und eindeutige Zuordnung zu den Skalen geprüft. In die Fragebogenvorform mit insgesamt 103 Items gingen pro Skala 5 bis 10 Items ein, für die entsprechend der geringeren Differenzierungsfähigkeit von Schülern ein 4stufiges, bipolares Antwortformat gewählt wurde (1 [*stimmt gar nicht*] bis 4 [*stimmt genau*]). Anschließend erfolgte die Pilotierung in drei Phasen an (1) Stichproben mit zwei Hauptschulklassen² ($n = 42$), (2) neun Realschulklassen ($n = 208$) sowie (3) mit Studierenden im ersten Semester des Lehramtsstudiums, die die Kompetenzen ihrer ehemaligen Lehrkräfte retrospektiv einschätzten ($n = 169$). Es wurden wenig trennscharfe, schwer verständliche sowie Items mit zu vielen Missings oder zu hohen Nebenladungen entfernt. Gleichzeitig erfolgte eine erste kriteriale Validierung mit Hilfe der Landauer Skalen zum Sozialklima LASSO (Saldern & Littig, 1987). Es zeigten sich kleine bis große Zusammenhänge mit Merkmalen der Lehrer-Schüler-Beziehung ($.14$ [*Prozeduren*] $< r < .88$ [*Arbeitsbündnis*], $Md = .49$). Auf Ebene der LASSO-Subskalen hängt die allgemeine Zufriedenheit mit dem Unterricht am deutlichsten mit Kompetenzen im Klassenmanagement zusammen ($.43$ [*Prozeduren*] $< r < .78$ [*Arbeitsbündnis*], $Md = .60$). Die Endform von KODEK-S enthält 42 Items mit je 4 bzw. 5 Items pro Skala und konnte auf eine Bearbeitungszeit von unter zehn Minuten reduziert werden.

Die Items und Informationen über deren Herkunft können bei der Autorin angefordert werden; Itembeispiele sind in Tabelle 1 dargestellt.³

2.2.2. Stichprobe

Die Instrumente wurden im Rahmen einer Interventionsstudie entwickelt und implementiert. Ziel der Intervention war die Entwicklung von Kompetenzen im Klassenmanagement mit Hilfe von theoretischen Inputs, Reflexionsaufgaben, Videoanalysen eigenen und fremden Unterrichts sowie Microteaching (vgl. Piwowar, Thiel & Ophardt, 2013). Die Wirksamkeit des Trainings wurde im Schuljahr 2010/2011 mit 45 eigeninitiativ teilnehmenden Lehrkräften (Teilstudie 1) in einem quasi-experimentellen Prä-Post-Design mit behandelte Kontrollgruppe überprüft (ebd.). Die Interventionsbedingung wurde frei gewählt. Im Schuljahr 2011/2012 wurde Teilstudie 2 mit 50 Referendaren eines Schulpraktischen Seminars in Berlin per Wartekontrollgruppendesign realisiert—die Kontrollgruppe blieb hier unbehandelt.

² Bis zur Sekundarschulreform im Jahr 2010 existierten in der Sekundarstufe 1 vier unterschiedliche Schulformen. Mit der Sekundarschulreform wurden Gesamtschulen, Realschulen und Hauptschulen in Integrierte Sekundarschulen überführt. Die Hauptstudien (Teilstudien 1 und 2, siehe Kapitel 2.2.2) wurden nach der Reform durchgeführt.

³ Die Fragebögen können auf folgender Internetseite abgerufen werden: www.ewi-psy.fu-berlin.de/kodek.

Die Teilnahme an der Evaluation erfolgte freiwillig ($n = 43$). In beiden Teilstudien ist daher von Selbstselektionseffekten auszugehen. Zur Evaluation der Intervention wurden in beiden Teilstudien dieselben Instrumente implementiert, so dass nachfolgend die Ergebnisse für die Gesamtstichprobe dargestellt werden.⁴

Die untersuchten TeilnehmerInnen bilden eine Gelegenheitsstichprobe von 88 Lehrkräften, die an 27 Berliner und Brandenburger Schulen in der Sekundarstufe 1 (Klasse 7 bis 10) unterrichten. Die Berufserfahrung reicht von Berufsanfängern nach dem ersten Jahr des Referendariats bis zu 39 Jahren ($M = 8.0$; $SD = 10.56$) bei einem Alter von 22 bis 62 Jahren ($M = 36.73$; $SD = 10.14$), 67% der Lehrkräfte sind weiblich, 41 Lehrkräfte unterrichten an Gymnasien. Die Unterrichtsevaluation fand in allen Fächern mit Ausnahme von Musik, Sport und Kunst statt.

An den Schülerbefragungen nahmen 1811 Schüler teil. Die Klassenstärke variierte von 8 bis 31 Schülern ($M = 21.2$, $SD = 5.58$) bei einem Mädchenanteil von 12 bis 76% ($M = 9.29\%$, $SD = 14.24$).

2.2.3. Messinstrumente

Münchener Aufmerksamkeitsinventar (MAI, Helmke & Renkl, 1992). Das MAI dient der systematischen Erfassung von Aufmerksamkeit und Beteiligung der Schüler während des Unterrichts. Auf Basis eines Zeitschichprobenverfahrens wird in 5-Sekunden-Intervallen sukzessive das Aufmerksamkeitsverhalten jedes Schülers erfasst als off-task (z. B. träumen oder mit dem Nachbarn albern) vs. on-task (z. B. sich melden oder mitschreiben). Ergänzend und dazu im Wechsel (nach vier Beobachtungsdurchgängen je Schüler) wird in einer jeweils zweiminütigen Unterrichtsbeobachtung die Beteiligung der gesamten Klasse hoch-inferent auf einer fünfstufigen Skala hinsichtlich der Dauer, der Intensität und des Ausmaßes der Beteiligung beurteilt. Für die vorliegende Untersuchung wurde für jede Stunde die durchschnittliche Beteiligung sowie der prozentuale Anteil des off-task Verhaltens ermittelt.

Lehrerselbstwirksamkeit. Von Schmitz und Schwarzer (2000) liegen 10 Items mit vierstufigem Antwortformat zur Erfassung der Lehrerselbstwirksamkeit vor. Die Reliabilität lag bei Cronbachs Alpha $\alpha = .76$.

Berufliche Belastung. Über drei Subskalen (insgesamt 15 Items) wird die berufliche Belastung per Selbstbericht mit einem fünfstufigen Antwortformat erfasst (Kontrolliertheitserleben, Arbeitsüberforderung und Arbeitsunzufriedenheit, Enzmann & Kleiber, 1989). Die Reliabilität lag bei $\alpha = .86$.

⁴ Pro Lehrkraft wurde eine–zu beiden Messzeitpunkten identische–Klasse für die Evaluation ausgewählt.

2.2.4. Prozedere

Die Datenerhebung erfolgte in situ im Rahmen einer regulären Unterrichtsstunde mit drei Beobachtern. Ein Beobachter erfasste mit Hilfe des MAI simultan zum Unterricht die Schülersaufmerksamkeit und -beteiligung. Die Beurteilung des Klassenmanagements mittels KODEK-B wurde ex-post-facto auf Grundlage eines Beobachtungsprotokolls durch zwei weitere Beobachter durchgeführt. Im Anschluss an die beobachtete Unterrichtsstunde wurden nach einer standardisierten Instruktion Schüler (KODEK-S) und Lehrkräfte (Angaben zu soziodemografischen Variablen, der Lehrerselbstwirksamkeit und beruflichen Belastung) befragt. Die Beobachter wurden in mehrtägigen Schulungen umfassend qualifiziert.

2.2.5. Statistische Analysen

Aufgrund der hierarchischen Struktur der Daten wird nachfolgend zwischen Aggregat- (pro Lehrkraft über die Beurteiler gemittelte Kennwerte) und Individualebene (Kennwerte einzelner Beurteiler pro Lehrkraft) unterschieden. Da die Fragebögen Lernumgebungen (Aggregatebene) und keine Unterschiede zwischen Schülern modellieren sollen (Individualebene), finden alle kriterialen Validitätsanalysen auf Aggregatebene statt. Bis auf die Angaben zur Stabilität gehen hierbei ausschließlich die Erhebungen des ersten Messzeitpunkts ein.

Es werden verschiedene Aspekte der *Reliabilität* ermittelt: (1) die interne Konsistenz (Cronbachs Alpha, α), (2) das Ausmaß an Wahrnehmungskonsens und Differenzierbarkeit ($ICC[1]$ und $ICC[2]$) sowie (3) die zeitliche Stabilität (Retest-Reliabilität, r_{tt}). Da lediglich bei der Kontrollgruppe aus Teilstudie 2 eine echte Messwiederholung vorliegt, basieren die Angaben zur Retest-Reliabilität auf einer Substichprobe ($n = 21$) mit 3-monatigem Messintervall. Alle Reliabilitätsmaße werden ab einem Wert von 0.7 als akzeptabel, ab 0.8 als gut und ab 0.9 als exzellent bewertet (Lienert & Raatz, 1998)–ausgenommen der $ICC(1)$ für den Schülerfragebogen.

Die $ICC(2)$ gilt als Maß der Reliabilität der Aggregatwerte, die $ICC(1)$ spezifiziert die Zuverlässigkeit einzelner Urteile bzw. gibt an, wie hoch der prozentuale Anteil der Gesamtvarianz ist, der auf Unterschiede zwischen Lehrkräften zurückzuführen ist (Lüdtke, Trautwein, Kunter & Baumert, 2006). Ab 10% Varianzaufklärung durch die Klassenzugehörigkeit kann von einem Merkmal auf Klassenebene ausgegangen werden (ebd.). Ob eine geteilte Wahrnehmung im Sinne intersubjektiver Übereinstimmung der Schüler vorliegt, wird mit dem Übereinstimmungsindex AD_M (*average absolute deviation*) überprüft. Übereinstimmung liegt vor, wenn der AD_M unter dem Wert von 0.67 bleibt (Gärtner, 2010). Die Retest-Reliabilität

wird über latente, bzw. bei Skalen mit nur einem Item manifesten Korrelationen der Aggregatwerte bestimmt.

Kennwerte zur konvergenten und diskriminanten *Validität* werden über manifeste und, wenn mehrere Indikatoren vorliegen, latente Korrelationen der Aggregatwerte ermittelt. Der Korrelationskoeffizient dient als Effektstärkemaß mit kleinen Effekten über 0.1, mittleren Effekten über 0.3 und großen Effekten über 0.5 (Cohen, 1988).

Da im Beobachterfragebogen lediglich zwei Skalen aus drei Items bestehen, kann die postulierte mehrdimensionale Struktur nicht faktorenanalytisch verifiziert werden. Die Struktur des Schülerfragebogens wird durch eine Mehrebenen-Faktorenanalyse überprüft (Brown, 2006; Lüdtke, Trautwein, Schnyder & Niggli, 2007). Dabei werden ausschließlich die Handlungsstrategien der Lehrkraft einbezogen, also nicht die Skala *Störungsausmaß*. Entsprechend der in Kapitel 2.1.1 dargestellten Überlappungen und gegenseitigen Abhängigkeit der Dimensionen bzw. Skalen wird für die Faktorenanalyse kein konfirmatives, sondern ein exploratives Vorgehen gewählt (*Multilevel Exploratory Structural Equation Modeling, MESEM*, Marsh et al., 2009). Mit Hilfe eines vierschrittigen Vorgehens wird die beste Approximation an die Daten identifiziert (Muthén, 1994): (1) Ermittlung der Variation zwischen den Klassen per $ICC(1)$; (2) separate Faktorenanalysen für die pooled-within Matrix (S_{PW}) und die (3) between Matrix (S_B); (4) integrierte Analyse per type = twolevel mit der in den Schritten (2) und (3) extrahierten Faktorenanzahl. Als Rotationsverfahren wird die oblique Geomin Rotation verwendet; für die integrierte Analyse in (4) wird der *WLSMV*-Schätzer verwendet (Brown, 2006). Die Auswahl des passenden Modells wird mit Hilfe inkrementeller (*CFI* und *TFI*; akzeptabel $> .90$, gut $> .97$) und absoluter Fit-Indizes (*RMSEA* und *SRMR*; akzeptabel $< .08$, gut $< .05$) sowie den informationstheoretischen Maßen *AIC*, *BIC* und *corBIC* (für N adjustierter *BIC*) vorgenommen (ebd.).⁵ Schließlich dient eine detaillierte Evaluation der Parameterschätzungen der Modellauswahl.

Latente und SEM-Analysen werden mit dem Programm *Mplus* (Version 6.0) durchgeführt, alle übrigen Berechnungen erfolgen mit der Statistiksoftware SPSS19. Das Signifikanzniveau wird für alle Analysen auf das 5%-Niveau festgelegt.

2.3. Ergebnisse

Die aggregierten Mittelwerte und Standardabweichungen sowie alle Kennwerte zur Reliabilität und der kriterialen Validität sind in Tabelle 1 dargestellt.

⁵ Mehta und Neale (2005) weisen darauf hin, dass MESEM eine relativ junge methodische Entwicklung ist und die Angemessenheit der Grenzwerte hier noch nicht überprüft ist – die Orientierung an den zitierten Grenzwerten hat damit nur eingeschränkte Aussagekraft.

2.3.1. Beobachterfragebogen: KODEK-B

Die durchschnittlichen Beurteilungen liegen bei einem theoretischen Mittelwert von 3.5 für alle Skalenwerte im positiven Bereich; am niedrigsten werden *Regeln* bewertet ($M_{\text{agg}} = 4.05$, $SD = 1.06$), am höchsten das *Arbeitsbündnis* ($M_{\text{agg}} = 5.29$, $SD = .70$).

Die internen Konsistenzen (als Maß der *Reliabilität*) sind akzeptabel und gut. Das Urteil eines einzelnen Raters ($ICC[1]$) mit Ausnahme der $ICC(1)$ für die Skalen *Klarheit des Handlungsprogramms* und *Störungsintervention* sowie die Interrater-Reliabilität für die aggregierten Klassenmittelwerte ($ICC[2]$) sind akzeptabel bis exzellent. Die niedrigste $ICC(2)$ zeigt sich für die *Klarheit des Handlungsprogramms*, $ICC(2) = .79$. Die Korrelationen der Skalenwerte von erster und zweiter Erhebung (r_{tt}) befinden sich alle unter der Akzeptanzgrenze für Reliabilitätsmaße ($.07 < r_{tt} < .57$), lediglich *Regeln* wurden im Abstand von drei Monaten mit einem großen Effekt ähnlich eingeschätzt, $r_{tt} = .57$.

Konvergente Validität. Es liegen mittlere bis hohe relative Übereinstimmungen mit den Einschätzungen der Schüler vor ($Md = .55$); am höchsten korreliert die Einschätzung des *Störungsmaßes* ($r = .65$), am geringsten die Beurteilung von *Klarheit des Handlungsprogramms* ($r = .30$). Der Anteil von off-task Verhalten weist konsistent negative Zusammenhänge mit den durch Beobachter eingeschätzten Aspekten des Klassenmanagements auf ($Md = -.54$), mit mittleren (*Klarheit des Handlungsprogramms*, $r = -.38$) bis zu großen Effekten (*Störungsintervention*, $r = -.73$). Der Zusammenhang mit der Schülerbeteiligung ist insgesamt geringer ($Md = .28$) mit kleinen (*Klarheit des Handlungsprogramms*, $r = .22$) bis zu mittleren positiven Zusammenhängen (*Störungsmaß*, $r = .42$).

Diskriminante Validität. Die Lehrerselbstwirksamkeit weist mittlere positive Zusammenhänge mit dem *Arbeitsbündnis* und der *Störungsintervention* ($r = .36$ und $r = .40$) auf. Ansonsten zeigen sich keine bis kleine positive und negative Zusammenhänge ($-.13 < r < .25$, $Md = .20$). Die Belastung einer Lehrkraft weist keine bis kleine positive Zusammenhänge mit der Beurteilung der Kompetenzen im Klassenmanagement auf ($-.08 < r < .24$, $Md = .16$).

Tabelle 1: Anzahl Items (N_i), Beispielitems, Mittelwerte (M_{agg}) und Standardabweichung (SD) der Aggregatwerte, Maße der Reliabilität sowie kriteriale Zusammenhänge (Validität) von KODEK-S und KODEK-B

Dimension	Skala	N_i	Beispielitem	M_{agg} (SD)	Reliabilität				Validität				
					α	ICC(1)	ICC(2)	$r_{t,t}$	$r_{B,S}$	r_{OT}	r_{BET}	r_{LS}	r_{BEL}
<i>KODEK-B</i>													
	SA(-) ¹	3	<i>Der Unterricht wird massiv gestört.</i>	4.80 (1.09)	.84	.86	.92	.40	.65	-.64	.42	.17	.18
MS	RE	1	<i>In der Klasse gelten klare Verhaltensregeln.</i>	4.05 (1.06)	–	.82	.90	.57	.47	-.54	.27	.20	.20
	SI	2	<i>Der Lehrkraft gelingt es, auftretende Störungen rasch und effizient zu unterbinden.</i>	4.75 (1.10)	–	.68	.81	.45	.48	-.73	.32	.40	.19
	MO	1	<i>Die Lehrkraft bekommt alles mit, was in der Klasse vor sich geht.</i>	4.90 (.87)	–	.75	.85	.07	.60	-.49	.28	.04	.14
MI	PR	2	<i>Regelmäßig wiederkehrende Abläufe verlaufen flüssig und ohne Zeitverlust.</i>	4.81 (.95)	–	.80	.89	.49	.53	-.52	.32	.25	.24
	ZM	3	<i>Die Unterrichtszeit wird effektiv genutzt.</i>	4.91 (.78)	.78	.78	.88	.39	.55	-.63	.31	-.13	-.08
	GM	2	<i>Die Lehrkraft bindet alle Schülerinnen und Schüler aktiv in das Unterrichtsgeschehen ein.</i>	4.26 (.97)	–	.71	.83	.33	.60	-.54	.26	.24	-.05
	KH	2	<i>Das Vorgehen der Lehrkraft ist klar und strukturiert.</i>	4.56 (.74)	–	.65	.79	.16	.30	-.38	.22	.06	-.07
MB	AB	2	<i>Es herrscht ein lernfreundliches Unterrichtsklima.</i>	5.29 (.70)	–	.81	.89	.51	.58	-.55	.25	.36	.16
<i>KODEK-S</i>													
	SA(-) ¹	4	<i>Bei diesem Lehrer können wir im Unterricht ungestört arbeiten.</i>	2.86 (.45)	.80	.48	.95	.87	–	-.46	.48	.23	-.04
MS	RE	4	<i>Bei diesem Lehrer ist klar, was man machen darf und was nicht.</i>	2.99 (.31)	.78	.33	.92	.87	–	-.32	.43	.30	-.07
	SI	4	<i>Dieser Lehrer beendet Störungen von Schülern schnell und ohne Diskussion.</i>	3.14 (.39)	.79	.31	.91	.91	–	-.37	.43	.11	-.03
	MO	4	<i>Dieser Lehrer bemerkt nicht, wenn sich Schüler mit etwas anderem beschäftigen.</i>	3.08 (.35)	.72	.24	.88	.89	–	-.39	.41	.09	.18

Tabelle wird fortgesetzt

Dimension	Skala	N _i	Beispielitem	M _{agg} (SD)	Reliabilität				Validität				
					α	ICC(1)	ICC(2)	r _{t,t}	r _{B,S}	r _{OT}	r _{BET}	r _{LS}	r _{BEL}
MI	PR	4	Bei diesem Lehrer sind wichtige Abläufe so gut eingeübt, dass kaum Zeit verloren geht (z.B. bei Gruppenarbeit, am Beginn der Stunde).	2.72 (.37)	.68	.35	.92	.88	–	-.42	.48	.29	.08
	ZM	4	Bei diesem Lehrer wird im Unterricht viel Zeit vertrödelt.	2.94 (.31)	.69	.34	.92	.91	–	-.37	.43	.30	-.08
	GM	4	Bei diesem Lehrer nimmt meist die gesamte Klasse aufmerksam am Unterrichtsgeschehen teil.	2.90 (.38)	.69	.41	.94	.88	–	-.43	.45	.24	.03
	KH	4	Bei diesem Lehrer frage ich mich oft, was zu tun ist.	3.24 (.33)	.73	.29	.90	.81	–	-.30	.38	.22	.04
MB	AB	5	Diesem Lehrer ist es sehr wichtig, dass wir etwas lernen.	3.23 (.42)	.84	.41	.94	.87	–	-.32	.40	.33	-.05
	KF	4	Dieser Lehrer hilft uns, Konflikte zu klären.	2.99 (.47)	.89	.33	.91	.87	–	-.24	.36	.28	-.08

Anmerkungen. MS – Management des Schülerverhaltens, MI – Management der Instruktion, MB – Management der Lehrer-Schüler-Beziehung, SA – Störungsausmaß, MO – Monitoring, RE – Regeln, SI – Störungsintervention, PR – Prozeduren, ZM – Zeitmanagement, GM – Gruppenmobilisierung, KH – Klarheit des Handlungsprogramms, AB – Arbeitsbündnis, KF – Konflikte, B – Beobachter, S – Schüler, OT – Anteil off-task Verhalten der Schüler, BET – Ausmaß der Schülerbeteiligung, LS – Lehrerselbstwirksamkeit, BEL – berufliche Belastung der Lehrkraft.

¹invers

2.3.2. Schülerfragebogen: KODEK-S

Die durchschnittlichen Beurteilungen liegen bei einem theoretischen Mittelwert von 2.5 für alle Skalenwerte im positiven Bereich; am niedrigsten wird der Einsatz von *Prozeduren* beurteilt ($M_{\text{agg}} = 2.72$, $SD = .37$), am höchsten die *Klarheit des Handlungsprogramms* ($M_{\text{agg}} = 3.24$, $SD = .33$).

Die internen Konsistenzen (als Maß der *Reliabilität*) der individuellen Schülerurteile sind akzeptabel bis gut, mit Ausnahme der Skalen *Prozeduren*, *Gruppenmobilisierung* und *Zeitmanagement*, welche knapp unter der akzeptierten Grenze für α liegen. Die Übereinstimmung liegt überwiegend unter dem kritischen Wert von $AD_M = .67$. Es gibt keine (*Monitoring*, *Prozeduren*, *Gruppenmobilisierung*) bis maximal fünf Klassen (*Umgang mit Konflikten*), in denen dieser Wert überschritten wird. Es konnten zwischen 24 (*Monitoring*) und 48% (*Störungsmaß*) der Varianz der Skalenwerte durch die Klassenzugehörigkeit aufgeklärt werden, die Kennwerte der Interrater-Reliabilität $ICC(2)$ und Retest-Reliabilität r_{tt} sind mindestens gut.

Faktorielle Validität. Die Varianzaufklärung durch die Klassenzugehörigkeit liegt auf Itemebene bei 9 bis 39% ($Md = 26\%$), so dass eine Berücksichtigung der Mehrebenenstruktur indiziert ist (Schritt 1, siehe Kap 5.5). In der explorativen Faktorenanalyse für die gepoolte within-Matrix S_{PW} (Schritt 2) erreichten erst ab sieben Faktoren alle Fit-Maße akzeptable Werte ($\chi^2 = 1130.94$; $df = 458$; $RMSEA = .029$; $CFI = .963$; $TLI = .944$; $SRMR = .020$). Die Lösungen mit acht und neun Faktoren wiesen einen besseren Fit auf, wurden jedoch verworfen: Der achte bzw. achte und neunte Faktor zeigten keine eigenen substantziellen Ladungen, zudem gab es ab der acht-faktoriellen Lösung keine wesentliche Modellverbesserung mehr. Für die between Matrix S_B wurden bis zu fünf Faktoren extrahiert (Schritt 3), erst ab drei Faktoren ergibt sich eine interpretierbare Ladungsstruktur sowie ein $SRMR < .05$. $RMSEA$, TLI und CFI erreichen die untere Grenze bei keiner Lösung. In der vier- und in der fünf-faktoriellen Lösung weisen der vierte bzw. vierte und fünfte Faktor ausschließlich Nebenladungen der ersten Faktoren auf, sodass die drei-Faktor-Lösung gewählt wurde. Die integrierte Analyse (Schritt 4) mit sieben Faktoren innerhalb von Klassen (within) und drei Faktoren zwischen Klassen (between) zeigt eine akzeptable bis gute Anpassung des Modells ($\chi^2 = 8181.167$; $df = 1332$; $RMSEA = .017$; $CFI = .923$; $TLI = .896$; $SRMR_{\text{within}} = .019$; $SRMR_{\text{between}} = .041$). Die Faktorladungen und Faktorinterkorrelationen der integrierten Analyse sind in Tabelle 2 abgebildet. Aus individueller Schülersicht können fünf von neun Skalen des Klassenmanagements als eigener Faktor identifiziert werden; die vier

instruktionalen Strategien sind hingegen empirisch nicht trennbar: *Zeitmanagement* und *Klarheit des Handlungsprogramms* sowie *Prozeduren* und zwei Items von *Gruppenmobilisierung* bilden jeweils einen gemeinsamen Faktor, zeigen z. T. aber auch Nebenladungen auf dem je anderen Faktor. Ein semantisch stark mit *Monitoring* assoziiertes Item der Skala *Gruppenmobilisierung* (GM1) lädt auf Faktor 3 (*Monitoring*), ein weiteres (GM4) lädt auf dem Faktor 2 (*Störungsintervention*). Die Faktorinterkorrelationen sind klein bis hoch ($Md = .39$; $.12 < r < .55$). Auf der Aggregatebene kann nicht zwischen den einzelnen Skalen unterschieden werden, vielmehr lassen sich die drei Dimensionen *Management des Schülerverhaltens* (Faktor 1), *Management der Instruktion* (Faktor 2) und *Management der Lehrer-Schüler-Beziehung* (Faktor 3) extrahieren. Die Items GM1 und GM4 laden auch hier im Bereich des Verhaltensmanagements. Item SI2 (Skala *Störungsintervention*) wird in der geteilten Wahrnehmung Faktor 2, *Management der Instruktion* zugeordnet. Schließlich weist das *Arbeitsbündnis* zwei Zuordnungen auf: Die höheren Ladungen liegen auf Faktor 2, jedoch können auch für Faktor 3 substantielle Ladungen nachgewiesen werden. Die Faktorinterkorrelationen zeigen, dass verhaltensbezogenes und instruktionales Management stark assoziiert, jedoch unterscheidbar sind ($r = .67$), am geringsten hängen *Management der Instruktion* und das *Management der Lehrer-Schüler-Beziehung* zusammen ($r = .33$).

Konvergente Validität. Der Zusammenhang des off-task Verhaltens sowie der Beteiligung der Schüler mit der Beurteilung des Klassenmanagements zeigt für alle Skalen mittlere Effekte in der erwarteten positiven (Beteiligung, $Md = .43$) bzw. negativen (off-task Verhalten, $Md = -.37$) Richtung, mit Ausnahme eines kleinen Zusammenhangs von off-task Verhalten mit dem *Umgang mit Konflikten*. Während die Zusammenhänge mit der Beteiligung etwas höher sind als mit off-task Verhalten, liegt der niedrigste Zusammenhang jeweils mit dem *Umgang mit Konflikten* vor ($r = .36$ bzw. $r = -.24$) und der höchste mit dem *Störungsausmaß* ($r = .48$ bzw. $r = -.46$) sowie für die Beteiligung zusätzlich mit *Prozeduren* ($r = .48$).

Diskriminante Validität. Mit Ausnahme von *Monitoring* (kein Zusammenhang) gibt es kleine und mittlere positive Zusammenhänge mit der Lehrerselbstwirksamkeit ($.11 < r < .33$, $Md = .26$). Die berufliche Belastung ist im Mittel nicht mit der Einschätzung des Klassenmanagements assoziiert ($-.08 < r < .18$, $Md = -.01$), lediglich *Monitoring* weist einen kleinen positiven Effekt auf ($r = .18$).

Tabelle 2: Faktorladungen und Faktorinterkorrelationen der Mehrebenen-Faktorenanalyse mit 7 within-Faktoren und 3 between-Faktoren.

Faktor Itemcode	within							between		
	1	2	3	4	5	6	7	1	2	3
RE1	.481							.586		
RE2	.689							.803		
RE3	.468							.873		
RE4	.282			.308				.635		.467
SI1		.594						.652	.475	
SI2		.326							.743	
SI3		.404						.703		
SI4		.702						.664	.400	
MO1(-)			.333		.319			.674		
MO2(-)			.763					1.017		
MO3			.389					.724		
MO4			.592					1.069		
PR1(-)					.216	.505				.665
PR2						.200				.654
PR3(-)					.330	.235				.958
PR4(-)						.313				.674
ZM1(-)					.317	.428				.703
ZM2					.232					.657
ZM3(-)					.395					.864
ZM4					.211					.692
GM1(-)			.308					.963		
GM2						.490	.205	.421	.596	
GM3	.208					.418	.310	.562		
GM4		.219							.632	
KH1					.236		.203			.623
KH2(-)					.536					.828
KH3(-)					.571					.804
KH4					.235					.921
AB1							.614		.631	.543
AB2					.220		.426		.896	
AB3							.406		.550	
AB4					.282		.215		.527	.407
AB5							.670		.580	.518
KF1				.663						.782
KF2				.751				.402		.687
KF3				.858						.716
KF4				.661						.663
<i>Faktorinterkorrelationen</i>										
2	.435							.665		
3	.354	.510						.500	.333	
4	.393	.424	.400							
5	.375	.380	.455	.345						
6	.178	.311	.323	.118	.234					
7	.519	.542	.481	.548	.512	.183				

Anmerkungen. MO – Monitoring, RE – Regeln, SI – Störungsintervention, PR – Prozeduren, ZM – Zeitmanagement, GM – Gruppenmobilisierung, KH – Klarheit des Handlungsprogramms, AB – Arbeitsbündnis, KF – Konflikte. (-) indiziert negative Items. Ladungen <.2 für die within-Matrix und <.4 für die between-Matrix sind nicht angezeigt.

2.4. Diskussion

In der vorliegenden Arbeit wurde die Entwicklung und empirische Überprüfung eines Beobachter- und eines Schülerfragebogens zum Klassenmanagement dargestellt (KODEK-B und KODEK-S). Auf Grundlage von empirischen und theoretischen Forschungsarbeiten wurden Kompetenzen des Klassenmanagements identifiziert, die den Dimensionen *Management von Schülerverhalten*, *Management der Instruktion* und *Management der Lehrer-Schüler-Beziehung* zugeordnet wurden. Sowohl die mehrdimensionale Konzeptualisierung, die an den internationalen Forschungsstand anschließt, als auch die gründliche empirische Prüfung der Gütekriterien stellen eine wichtige Weiterentwicklung der existierenden Instrumente dar.

Hinsichtlich der Reliabilität führen die Individualurteile von Schülern zu hohen Übereinstimmungen innerhalb von Klassen. Mit 29 bis 48% Varianzaufklärung durch die Klassenzugehörigkeit ist die Differenzierungsfähigkeit der Schülerurteile im Vergleich zu anderen Konstrukten der Unterrichtsqualität hoch (Kunter & Baumert, 2006). Das Schülerurteil (KODEK-S) erreicht bei allen drei Reliabilitätsmaßen überwiegend exzellente, mindestens gute Kennzahlen; die internen Konsistenzen liegen z. T. knapp unter der akzeptierten Grenze. Auch KODEK-B weist mindestens akzeptable Interrater-Reliabilitäten sowie interne Konsistenzen auf. Während die *ICC(2)* auf eine gute querschnittliche Differenzierung hinweist, indizieren die maximal mittleren Korrelationen von erster und zweiter Messung, dass der Vergleich von zwei Unterrichtsstunden im Abstand von drei Monaten bezüglich Klassenmanagement nicht stabil ist bzw. mit KODEK-B nicht stabil erfasst werden kann. Dies kann einerseits darauf zurückzuführen sein, dass das Klassenmanagement beim Vergleich einzelner Unterrichtsstunden tatsächlich situativ unterschiedlich ist, z. B. in Abhängigkeit von den anwesenden Schülern oder den Lerninhalten (Doyle, 2006). Darüber hinaus beziehen sich die Kennwerte auf eine relativ kleine Gruppe von Referendaren, bei denen der Unterricht möglicherweise noch durch hohe Variabilität gekennzeichnet ist. Der Einsatz des Instruments für längsschnittliche Analysen sollte daher mindestens zwei Beobachtungen pro Messzeitpunkt umfassen. Auch könnte geprüft werden, ob auf videografierten Unterrichtsstunden basierende Einschätzungen, die das mehrmalige Beobachten einer Unterrichtsstunde ermöglichen, zu stabileren Ergebnissen führen.

Hinsichtlich der Validität lassen sowohl die deduktive Konstruktionsmethode, welche eine repräsentative Auswahl der konstruktrelevanten Facetten berücksichtigt, als auch die Prüfung der entwickelten Items durch Experten auf Inhaltsvalidität der entwickelten Instrumente schließen. Erste Hinweise auf die kriteriale Validität des Schülerfragebogens zeigten sich in einer Pilotstudie, in der der KODEK-S und der Klimafragebogen LASSO simultan

eingesetzt wurden. Merkmale der Lehrer-Schüler-Beziehung hingen erwartungskonform am höchsten mit der Beurteilung des *Arbeitsbündnisses* zusammen; die Zufriedenheit mit dem Unterricht korrelierte deutlich mit allen durch KODEK-S erfassten Aspekten.

Validierungsanalysen der vorliegenden Studie bezogen sich auf die weitere kriteriale Validierung sowie eine empirische Überprüfung der postulierten mehrdimensionalen Struktur des Schülerfragebogens. Es zeigten sich erwartungsgemäß mittlere bis hohe Zusammenhänge von Schüler- und Beobachterbeurteilung, die gegenüber Studien ohne perspektivenspezifische Items relativ hoch ausfallen (vgl. z. B. Clausen, 2002). Insbesondere jene Facetten zeigten hohe Übereinstimmung, bei denen die Beobachtbarkeit hoch und die Verständnisanforderungen gering sind, nämlich *Monitoring*, *Gruppenmobilisierung* und *Störungsmaß*. Ob diese Übereinstimmungen systematisch sind, also z. B. gleiche Aspekte ähnlicher bewertet werden als ungleiche, sollte im Rahmen weiterführender Studien mit Methoden des Multi-Trait-Multi-Method Ansatzes überprüft werden.

Die konvergente Validierung implizierte Variablen auf Schülerseite (off-task Verhalten und Schülerbeteiligung). Es zeigten sich die postulierten negativen Zusammenhänge mit dem off-task Verhalten sowie positive Zusammenhänge mit der Schülerbeteiligung. Es wurden unterschiedlich starke Zusammenhänge identifiziert, die die Unterscheidbarkeit der Skalen empirisch untermauern: So sind die Skalen *Umgang mit Konflikten* und *Klarheit des Handlungsprogramms*, die distalsten Strategien bezüglich des Schülerverhaltens, aus beiden Perspektiven am niedrigsten mit off-task Verhalten und der Schülerbeteiligung assoziiert.

Die diskriminanten Zusammenhänge mit Variablen auf Lehrerseite (Lehrerselbstwirksamkeit und berufliche Belastung) sind durchschnittlich geringer als die konvergenten, zudem zeichnet sich hier ein heterogeneres Bild. Wie erwartet zeigten sich maximal mittlere Korrelationen mit den diskriminanten Kriterien. Allerdings sind diese Zusammenhänge perspektivenspezifisch, also nur zum Teil auf die gleichen Skalen bezogen. Auffällig sind darüber hinaus Höhe und Richtung der Zusammenhänge mit der subjektiven Belastung: Während Klassenmanagement aus Schülersicht nicht mit der Belastung der Lehrkräfte assoziiert ist, zeigten sich für das Beobachterurteil in der Tendenz sogar positive Effekte—Lehrkräfte mit mehr zugeschriebenen Kompetenzen in diesen Strategien fühlten sich eher beruflich belastet. Beide Ergebnisse widersprechen empirischen Befunden, die einen negativen Zusammenhang von Klassenmanagement mit Stresserleben nachgewiesen haben (Clunies-Ross, Little & Kienhuis, 2008; Friedman, 2006)—wenn auch Klusman, Kunter, Trautwein & Baumert (2006) darauf hinweisen, dass bislang vor allem selbsteingeschätzte Kompetenzen untersucht wurden.

Da gerade Berufsanfänger Disziplinprobleme als besonders belastend empfinden (Pigge & Marso, 1997), wurden in post hoc Analysen differenzielle, latente Zusammenhänge von Kompetenzen im Klassenmanagement und subjektiver Belastung untersucht. Die getrennten Analysen von Lehrern im Referendariat ($M_{\text{Belastung}} = 2.19$, $SD = .48$) vs. Lehrern, die seit durchschnittlich 18 Jahren im Beruf stehen ($M_{\text{Belastung}} = 1.99$, $SD = .56$) weisen auf differenzielle Effekte hin: Während der Zusammenhang in der Gruppe der Referendare erwartungswidrig überwiegend und im Mittel positive Effekte indiziert ($Md_{\text{Beobachter}} = .18$, $-.14 < r < .35$; $Md_{\text{Schüler}} = .16$, $.02 < r < .42$), zeigen sich in der Gruppe der Lehrer die zu erwartenden negativen Zusammenhänge ($Md_{\text{Beobachter}} = -.10$, $-.42 < r < .17$; $Md_{\text{Schüler}} = -.17$, $-.30 < r < .09$). Im Rahmen des Referendariats fühlen sich demnach Lehrkräfte mit geringeren, durch Fremdbereicht erfassten, Kompetenzen im Klassenmanagement tendenziell weniger belastet als ihre erfolgreicherer KollegInnen. Eine mögliche Erklärung dafür könnte sein, dass Lehramtsanwärter, die sich der Herausforderung des Umgangs mit schwierigem Schülerverhalten erfolgreich stellen, dies nur unter hoher (subjektiver) Belastung zu leisten vermögen. Interessant sind in diesem Zusammenhang Befunde von Klusmann, Kunter, Voss & Baumert & (2012), die querschnittlich keinen über Persönlichkeitsmerkmale und berufliche Selbstregulation hinausgehenden Effekt von Wissen in Klassenführung auf das Beanspruchungserleben nachweisen konnten–im Längsschnitt erwies sich Klassenführungswissen jedoch durchaus als bedeutsamer Prädiktor für das Beanspruchungserleben in einem Jahr. So werden möglicherweise auch Kompetenzen im Klassenmanagement erst im Verlauf der beruflichen Entwicklung als Ressource wirksam. Ob sich die gefundenen differenziellen Effekte als robust erweisen und welche Moderatoren die interindividuellen Unterschiede im Belastungserleben neben der beruflichen Entwicklungsphase erklären könnten, sollte in weiterführenden Studien hypothesengeleitet überprüft werden. Zudem könnten methodische Einschränkungen die Ergebnisse beeinflusst haben: Es ist wahrscheinlich, dass die Auswahl der Stichprobe (freiwillige Teilnahme im Rahmen einer Interventionsstudie) gerade in Bezug auf das Belastungserleben nur eingeschränkt repräsentativ für die Population von Lehrkräften ist.

Die empirische Prüfung der durch ein rationales Vorgehen extrahierten Kompetenzfacetten zeigte selbst in dieser frühen Phase der Fragebogenentwicklung, dass aus individueller Schülerwahrnehmung die meisten Kompetenzfacetten, mit Ausnahme der Strategien *Prozeduren* und *Gruppenmobilisierung* bzw. *Zeitmanagement* und *Klarheit des Handlungsprogramms*, die je einen gemeinsamen Faktor bilden, disjunkt voneinander beurteilt werden. Dies ist insofern ein bemerkenswertes Ergebnis, da Marsh et al. (2009) darauf hingewiesen haben,

dass Faktorenanalysen mit vielen Faktoren häufig nicht die Mindestanforderungen für Grenzwerte erfüllen und jeder Faktor durch fünf bis zehn Items repräsentiert sein sollte.

Für das aggregierte Schülerurteil ließen sich demgegenüber lediglich die drei allgemeineren Dimensionen *Management von Schülerverhalten*, *Management der Instruktion* und *Management der Lehrer-Schüler-Beziehung* als getrennte Faktoren identifizieren; das *Arbeitsbündnis* ist in der geteilten Wahrnehmung stark mit dem *Management der Instruktion* assoziiert, was dessen zentrale Bedeutung für die instruktionale Steuerung hervorhebt. Ob aggregierte Schülerurteile intraindividuelle Stärken und Schwächen auf Skalenebene identifizieren können, muss damit vorerst in Frage gestellt werden. Die gefundene Dreiteilung spricht dennoch für die Güte des Instruments: In anderen Studien mit hierarchischen Datensätzen zeigte sich häufig nur ein einziger Faktor auf der Aggregatebene (Muthén & Asparouhov, 2011).

2.4.1. Limitationen

Bei der Konstruktion der Fragebögen wurden jene Facetten von Klassenmanagement einbezogen, die für die Sekundarstufe besonders zentral sind. In der angloamerikanischen Literatur wird in vielen Handbüchern auch die physische Gestaltung der Lernumgebung sowie die Relevanz von Elternarbeit, Schulkoooperation und kollegialer Zusammenarbeit explizit thematisiert (Evertson & Weinstein, 2006a). Diese Aspekte wurden hier zugunsten eines ökonomischen Instruments ausgeklammert, könnten aber in weiteren Schritten als zusätzliche Facetten mit berücksichtigt werden.

Die vorliegende Studie basiert auf einer relativ kleinen, selbstselektierten Stichprobe. Zwar sind verschiedene Unterrichtsfächer, Berufserfahrung und Schultypen in der Stichprobe repräsentiert, was die Verallgemeinerbarkeit der Ergebnisse begünstigt—die Prüfung der Sensitivität der Instrumente (z. B. systematische Unterschiede in Abhängigkeit von der Schulform oder der Berufserfahrung) konnte jedoch nicht erfolgen, da potentielle Konfundierungen (hier z. B. hoher Anteil von Referendaren an Gymnasien) nicht kontrolliert werden konnten.

Schließlich gilt die empirische Prüfung der faktoriellen Struktur nur für den Schülerfragebogen KODEK-S. Aussagen zur Differenzierbarkeit der Skalen des KODEK-B konnten ausschließlich aus den Korrelationsanalysen mit Kriterien abgeleitet werden. Zudem sollte eine Zusammenlegung der Strategien zum instruktionalen Management oder eine Itemrevison in Betracht gezogen werden.

2.4.2. Zusammenfassung und Ausblick

Sowohl KODEK-B als auch KODEK-S weisen gute bis sehr gute Testeigenschaften auf. Während Schüler stabile und zuverlässige Einschätzungen vornehmen können, sollte der KODEK-B pro Messzeitpunkt mehrere Unterrichtsstunden einbeziehen. Beide Instrumente weisen unterschiedlich hohe ökologische Zusammenhänge mit dem Aufmerksamkeitsverhalten und der Schülerbeteiligung auf–trotz der schlechten Trennbarkeit der Skalen für aggregierte Schülerurteile. Damit ist ein erster Nachweis erbracht, dass die mehrdimensionale, zunächst theoretisch abgeleitete Operationalisierung auch empirisch bestätigt werden konnte.

In weiterführenden Studien sollte der Einsatz der Instrumente in verschiedenen Lerngruppen systematisch überprüft sowie weitere Merkmale zur Validierung einbezogen werden, z. B. die emotionale Entwicklung von Schülern oder ihre akademische Leistung. Auf Basis videografiertes Unterrichtsstunden sollten darüber hinaus Generalisierbarkeitskoeffizienten ermittelt werden, um auch die Intra-Rater-Reliabilität zu quantifizieren.

2.5. Literaturverzeichnis

- Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann et al. (2009). *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Brophy, J. & Good, T. L. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Handbook of research on teaching* (pp. 328–375). New York, NY: Macmillan.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- Brouwers, A. & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education*, 16, 239–253.
- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive?* Münster: Waxmann.
- Clunies-Ross, P., Little, E. & Kienhuis, M. (2008). Self-reported and actual use of proactive and reactive classroom management strategies and their relationship with teacher stress and student behaviour. *Educational Psychology*, 28, 693–710.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum Associates.
- Doyle, W. (2006). Ecological approaches to classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management* (pp. 97–125). Mahwah, NJ: Lawrence Erlbaum Associates.
- Emmer, E. T., Evertson, C. M. & Worsham, M. E. (2003). *Classroom management for secondary teachers* (6th ed.). Boston: Allyn and Bacon.
- Enzmann, D. & Kleiber, D. (1989). *Helfer-Leiden: Streß und Burnout in psychosozialen Berufen*. Heidelberg: Asanger.
- Evertson, C. M. (1985). Training teachers in classroom management: An experimental study in secondary school classrooms. *Journal of Educational Research*, 79, 51–58.
- Evertson, C. M. & Harris, A. H. (1992). What we know about managing classrooms. *Educational Leadership*, 49, 74–78.
- Evertson, C. M. & Weinstein, C. S. (Eds.). (2006a). *Handbook of classroom management*. Mahwah, NJ: Lawrence Erlbaum Associates.

- Evertson, C. M. & Weinstein, C. S. (2006b). Classroom management as a field of inquiry. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management* (pp. 3–15). Mahwah, NJ: Lawrence Erlbaum Associates.
- Freiberg, H. J. (1999). Beyond behaviorism. In H. J. Freiberg & J. E. Brophy (Eds.), *Beyond behaviorism: Changing the classroom management paradigm* (pp. 3–20). Boston: Allyn and Bacon.
- Friedman, I. A. (2006). Classroom management and teacher stress and burnout. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management* (pp. 925–945). Mahwah, NJ: Lawrence Erlbaum Associates.
- Gärtner, H. (2010). Wie Schülerinnen und Schüler ihre Lernumwelt wahrnehmen: Ein Vergleich verschiedener Maße zur Übereinstimmung von Schülerwahrnehmungen. *Zeitschrift für Pädagogische Psychologie*, 24, 111–122.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen: Schüler als Quellen der Unterrichtsbeschreibung*. Münster: Waxmann.
- Haag, L. & Streber, D. (2012). *Klassenführung: Erfolgreich unterrichten mit Classroom Management*. Weinheim: Beltz.
- Helmke, A. & Renkl, A. (1992). Das Münchener Aufmerksamkeitsinventar (MAI): Ein Instrument zur systematischen Verhaltensbeobachtung der Schüleraufmerksamkeit im Unterricht. *Diagnostica*, 38, 130–141.
- Helmke, A., Helmke, T., Schrader, F.-W. & Wagner, W. (2007). *Der Ratingbogen der DESI-Videostudie*. Universität Koblenz-Landau, Campus Landau.
- Kounin, J. S. (1970). *Discipline and group management in classrooms*. New York, NY: Holt, Rinehart & Winston.
- Klusmann, U., Kunter, M., Trautwein, U. & Baumert, J. (2006). Lehrerbelastung und Unterrichtsqualität aus der Perspektive von Lehrenden und Lernenden. *Zeitschrift für Pädagogische Psychologie*, 20, 161–173.
- Klusmann, U., Kunter, M., Voss, T. & Baumert, J. (2012). Berufliche Beanspruchung angehender Lehrkräfte: Die Effekte von Persönlichkeit, pädagogischer Vorerfahrung und professioneller Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 26, 275–290.
- Kunter, M. & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, 9, 231–251.
- Kunter, M., Brunner, M., Baumert, J., Klusmann, U., Krauss, S., Blum, W. et al. (2005). Der Mathematikunterricht der PISA-Schülerinnen und -Schüler. Schulformunterschiede in der Unterrichtsqualität. *Zeitschrift für Erziehungswissenschaft*, 8, 502–520.

- Kunter, M. & Klusmann, U. (2010). Kompetenzmessung bei Lehrkräften. *Unterrichtswissenschaft*, 38, 68–86.
- Landrum, T. J. & Kauffman, J. M. (2006). Behavioral approaches to classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management* (pp. 47–71). Mahwah, NJ: Lawrence Erlbaum Associates.
- Leinhardt, G., Weidman, C. & Hammond, K. M. (1987). Introduction and integration of classroom routines by expert teachers. *Curriculum Inquiry*, 17, 135–176.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- Lotz, M. & Gabriel, K. (2012, März). *Niedrig und hoch inferente Verfahren zur Erfassung von Unterrichtsqualität in der Grundschule*. Vortrag auf dem 23. Kongress der Deutschen Gesellschaft für Erziehungswissenschaft, Osnabrück.
- Lüdtke, O., Trautwein, U., Kunter, M. & Baumert, J. (2006). Analyse von Lernumwelten. Ansätze zur Bestimmung der Reliabilität und Übereinstimmung von Schülerwahrnehmungen. *Zeitschrift für Pädagogische Psychologie*, 20, 85–96.
- Lüdtke, O., Trautwein, U., Schnyder, I. & Niggli, A. (2007). Simultane Analysen auf Schüler- und Klassenebene. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 39, 1–11.
- Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S. et al. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, 16, 439–476.
- Martin, N. K. & Yin, Z. (1999). Beliefs regarding classroom management style. *Journal of Research in Rural Education*, 15, 101–105.
- Mayr, J. (2006). Klassenführung auf der Sekundarstufe II. *Schweizerische Zeitschrift für Bildungswissenschaften*, 28, 227–242.
- Mayr, J., Eder, F. & Fartacek, W. (2002). *Linzer Diagnosebogen zur Klassenführung (LDK)*. Retrieved March 23, 2013, from <http://ius.uni-klu.ac.at/projekte/ldk/>.
- Mehta, P. D. & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, 10, 259–284.
- Moos, R. H. & Trickett, E. J. (1974). *Classroom Environment Scale*. Palo Alto, CA: Consulting Psychologists Press.
- Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, 22, 376–398.

- Muthén, B. O. & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp.15-40). New York, NY: Routledge.
- Nie, Y. & Lau, S. (2009). Complementary roles of care and behavioral control in classroom management: The self-determination theory perspective. *Contemporary Educational Psychology*, 34, 185–194.
- Oevermann, U. (1996). Theoretische Skizze einer revidierten Theorie professionalisierten Handelns. In A. Combe & W. Helsper (Eds.), *Pädagogische Professionalität* (pp. 70–182). Frankfurt am Main: Suhrkamp.
- Ophardt, D. & Thiel, F. (2013). *Klassenmanagement. Ein Handbuch für Studium und Praxis*. Stuttgart: Kohlhammer.
- Pianta, R. C. & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38, 109–119.
- Pigge, F. L. & Marso, R. N. (1997). A seven year longitudinal multi-factor assessment of teaching concerns development through preparation and early years of teaching. *Teaching and Teacher Education*, 13, 225–235.
- Piwowar, V., Thiel, F. & Ophardt, D. (2013). Training inservice teachers' competencies in classroom management: A quasi-experimental study with teachers of secondary schools. *Teacher and Teacher Education*, 30, 1-12.
- Rakoczy, K. & Pauli, C. (2006). Hoch inferentes Rating: Beurteilung der Qualität unterrichtlicher Prozesse. In E. Klieme, C. Pauli & K. Reusser (Eds.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie "Unterrichtsqualität, Lehrerverhalten und mathematisches Verständnis"* (pp. 206-233). Frankfurt am Main: Gesellschaft zur Förderung Internationaler Pädagogischer Forschung.
- Saldern, M. & Littig, K. E. von. (1987). *Landauer Skalen zum Sozialklima (LASSO 4-13): Beiheft mit Anleitung und Normtabellen*. Weinheim: Beltz.
- Schmitz, G. S. & Schwarzer, R. (2000). Selbstwirksamkeitserwartung von Lehrern: Längsschnittbefunde mit einem neuen Instrument. *Zeitschrift für Pädagogische Psychologie*, 14, 12–25.
- Schönbächler, M.-T. (2008). *Klassenmanagement: Situative Gegebenheiten und personale Faktoren in Lehrpersonen- und Schülerperspektive*. Bern: Haupt.
- Seidel, T. (2009). Klassenführung. In E. Wild & K. Möller (Eds.), *Pädagogische Psychologie* (pp. 135–148). Heidelberg: Springer.

- Seidel, T. & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of educational research*, 77, 454–499.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1-22.
- Thiel, F., Richter, S. G. & Ophardt, D. (2012). Steuerung von Übergängen im Unterricht. Eine Experten-Novizen-Studie zum Klassenmanagement. *Zeitschrift für Erziehungswissenschaft*, 4, 727-752.
- van Tartwijk, J., den Brok, P., Veldman, I. & Wubbels, T. (2009). Teachers' practical knowledge about classroom management in multicultural classrooms. *Teaching and Teacher Education*, 25, 453–460.
- Voss, T., Kunter, M. & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, 103, 952–969.
- Walker, H. M., Ramsey, E. & Gresham, F. M. (2004). *Antisocial behaviour in school: Strategies and best practices*. Pacific Grove, CA: Brooks/Cole.
- Wang, M. C., Haertel, G. D. & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of educational research*, 63, 249–294.
- Weinstein, C. S. (1999). Reflections on best practices and promising programs: Beyond assertive classroom discipline. In H. J. Freiberg & J. E. Brophy (Eds.), *Beyond behaviorism* (pp. 147–163). Boston: Allyn and Bacon.
- Weinstein, C. S. (2007). *Middle and secondary classroom management: Lessons from research and practice*. Boston: Mc Graw Hill.
- Woolfolk Hoy, A. & Weinstein, C. S. (2006). Student and teacher perspectives on classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management* (pp. 181–219). Mahwah, NJ: Lawrence Erlbaum Associates.

2.6. Anhang

2.6.1. Item-Skalen-Zuordnung, Itemherkunft sowie Itemkürzel von KODEK-S

Skala / Item	Herkunft	Kürzel
Störungsausmaß		
Bei diesem Lehrer können wir im Unterricht ungestört arbeiten.	angelehnt an Kunter et al., 2002	-
Bei diesem Lehrer ist unsere Klasse während des Unterrichts aufmerksam und konzentriert.	angelehnt an Helmke & Jäger, 2004	-
Bei diesem Lehrer wird der Unterricht oft sehr gestört.	Baumert et al., 1997	-
Dieser Lehrer muss im Unterricht ständig Schüler ermahnen.	angelehnt an Baumert et al., 1997	-
Regeln		
Bei diesem Lehrer ist uns Schülern klar, warum bestimmte Regeln wichtig sind.	Eigenentwicklung	RE1
Bei diesem Lehrer sind die Regeln, die man einhalten muss, allen bekannt.	Baumert et al., 1997	RE2
Bei diesem Lehrer ist klar, was man machen darf und was nicht.	Baumert et al., 1997	RE3
Dieser Lehrer hat uns klar gemacht, wie wir miteinander umgehen sollten.	angelehnt an Neuen-schwander et al. 2003	RE4
Störungsintervention		
Dieser Lehrer beendet Störungen von Schülern schnell und ohne Diskussion.	Eigenentwicklung	SI1
Dieser Lehrer reagiert angemessen auf Störungen (also nicht zu streng und nicht zu mild).	angelehnt an Schön-bächler, 2005	SI2
Dieser Lehrer reagiert so auf Störungen, dass die Schüler aufhören zu stören.	Eigenentwicklung	SI3
Diesem Lehrer gelingt es, Störungen schnell zu beenden.	angelehnt an Schön-bächler, 2005	SI4
Monitoring		
Dieser Lehrer ist oft so auf den Stoff konzentriert, dass er Störungen gar nicht bemerkt.	Eigenentwicklung	MO1
Dieser Lehrer bemerkt nicht, wenn sich Schüler mit etwas anderem beschäftigen.	angelehnt an Baumert et al., 1997	MO2
Dieser Lehrer merkt, was beim Rest der Klasse vorgeht, auch wenn er mit einzelnen Schülern beschäftigt ist.	Schönbächler, 2005	MO3
Dieser Lehrer merkt sofort, wenn Schüler beginnen, sich mit etwas anderem zu beschäftigen.	angelehnt an Baumert et al., 1997	MO4
Prozeduren		
Bei diesem Lehrer dauert es zu Beginn der Stunde sehr lange, bis die Schüler ruhig werden und zu arbeiten beginnen.	Baumert et al., 1997	PR1
Bei diesem Lehrer sind wichtige Abläufe so gut eingeübt, dass kaum Zeit verloren geht (z.B. bei Gruppenarbeit, am Beginn der Stunde oder beim Verteilen von Arbeitsblättern).	angelehnt an Wannack et al., 2009	PR2
Bei diesem Lehrer nehmen Übergänge von einem Unterrichtsabschnitt zum nächsten (z.B. vom Klassengespräch zur Stillarbeit) oft viel Zeit in Anspruch.	Eigenentwicklung	PR3
Wenn Hilfsmaterialien beschafft werden, entsteht viel Unruhe.	angelehnt an Schön-bächler, 2005	PR4

Tabelle wird fortgesetzt

130 TEILSTUDIE 1: ERFASSUNG VON KLASSENMANAGEMENT

Skala / Item	Herkunft	Kürzel
Zeitmanagement		
Bei diesem Lehrer wird im Unterricht viel Zeit vertrödelt.	angelehnt an Baumert et al., 1997	ZM1
Dieser Lehrer behält die Zeit im Blick.	Angelehnt an Helmke et al., 2007	ZM2
Dieser Lehrer wird mit dem Stoff, den er in einer Unterrichtsstunde behandeln will, oft nicht fertig.	Eigenentwicklung	ZM3
Dieser Lehrer kann die Zeit, die wir für die Bearbeitung von Aufgaben brauchen, meistens richtig einschätzen.	Eigenentwicklung	ZM4
Gruppenmobilisierung		
Wenn ich bei diesem Lehrer eine Aufgabe nicht bearbeite, fällt das gar nicht auf.	Eigenentwicklung	GM1
Bei diesem Lehrer arbeiten immer nur einige wenige Schüler mit.	Eigenentwicklung	GM2
Bei diesem Lehrer nimmt meist die <i>gesamte</i> Klasse aufmerksam am Unterrichtsgeschehen teil.	angelehnt an Helmke et al., 2007	GM3
Dieser Lehrer sorgt dafür, dass alle etwas zu tun haben.	angelehnt an Mayr et al., 2002	GM4
Klarheit des Handlungsprogramms		
Bei diesem Lehrer sind die Aufgaben immer ganz klar.	Eigenentwicklung	KH1
Bei diesem Lehrer weiß ich oft gar nicht, wo wir im Stoff gerade sind.	Eigenentwicklung	KH2
Bei diesem Lehrer frage ich mich oft, was zu tun ist.	angelehnt an Baumert et al., 1997	KH3
Dieser Lehrer geht im Unterricht in einer logischen Reihenfolge vor.	Baumert et al., 2009	KH4
Arbeitsbündnis		
Dieser Lehrer hat eine gute Beziehung zu den Schülern.	Ditton & Merz, 2000	AB1
Dieser Lehrer verhält sich uns gegenüber fair.	angelehnt an Mayr, 1998	AB2
Dieser Lehrer schafft es, dass sich seine Schüler freiwillig anstrengen.	Eigenentwicklung	AB3
Diesem Lehrer ist es sehr wichtig, dass wir etwas lernen.	Eigenentwicklung	AB4
Mit diesem Lehrer macht das Lernen Spaß.	Eigenentwicklung	AB5
Konflikte		
Dieser Lehrer hört sich bei Streit zwischen Schülern beide Seiten an.	Eigenentwicklung	KF1
Diese Lehrerin hilft uns, Konflikte zu klären.	angelehnt an Mayr, 1998	KF2
Diese Lehrerin hilft uns, Streit untereinander zu schlichten	Eigenentwicklung	KF3
Diesem Lehrer gelingt es, Auseinandersetzungen unter den Schülern fair zu klären.	Eigenentwicklung	KF4

2.6.2. Item-Skalen-Zuordnung und Itemherkunft von KODEK-B

Skala / Item	Herkunft
Störungsausmaß	
Der Unterricht wird massiv gestört	angelehnt an Baumert et al., 1997
Der Unterricht verläuft nahezu störungsfrei.	angelehnt an Helmke et al., 2007
Die Lehrkraft ist ständig damit beschäftigt, auf Störungen zu reagieren.	Eigenentwicklung
Regeln	
In der Klasse gelten klare Verhaltensregeln.	angelehnt an Helmke et al., 2007
Störungsintervention	
Der Lehrkraft gelingt es, auftretende Störungen rasch und effizient zu unterbinden.	Eigenentwicklung
Zurechtweisungen bei Störungen sind angemessen.	angelehnt an Schönbächler, 2005
Monitoring	
Die Lehrkraft bekommt alles mit, was in der Klasse vor sich geht.	angelehnt an Mayr, 1998
Prozeduren	
Regelmäßig wiederkehrende Abläufe verlaufen flüssig und ohne Zeitverlust.	angelehnt an Helmke et al. 2007
Übergänge zwischen einzelnen Unterrichtsphasen laufen routiniert ab.	angelehnt an Schönbächler, 2005
Zeitmanagement	
Das, was die Lehrkraft sich für die Stunde vorgenommen hatte, wird in der Stunde auch erreicht.	Eigenentwicklung
Die Lehrkraft schreitet in angemessenem Tempo (nicht zu zügig und nicht zu langsam) durch den Stoff.	angelehnt an Baumert et al., 1997
Die Unterrichtszeit wird effektiv genutzt.	angelehnt an Helmke et al., 2007
Gruppenmobilisierung	
Die Lehrkraft bindet <i>alle</i> Schülerinnen und Schüler aktiv in das Unterrichtsgeschehen ein.	Eigenentwicklung
Die Lehrkraft macht deutlich, dass sie eine aktive und engagierte Bearbeitung der Aufgaben von jedem einzelnen bzw. jeder einzelnen erwartet.	angelehnt an Helmke et al., 2007
Klarheit des Handlungsprogramms	
Das Vorgehen der Lehrkraft ist klar und strukturiert.	Eigenentwicklung
Die Lehrkraft formuliert Aufgaben so, dass die Schüler genau wissen, was zu tun ist.	angelehnt an Baumert et al., 2009
Arbeitsbündnis	
Die Schülerinnen und Schüler ziehen mit der Lehrkraft an einem Strang.	Eigenentwicklung
Es herrscht ein lernfreundliches Unterrichtsklima.	Eigenentwicklung

2.6.3. Literatur

- Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U. et al. (2009). *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente (Materialien aus der Bildungsforschung Nr. 83)*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Baumert, J., Gruehn, S., Heyn, S., Köller, O. & Schnabel, K.-U. (1997). *Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU). Dokumentation: Bd 1. Skalen Längsschnitt I, Welle 1–4*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Ditton, H. & Merz, D. (2000). *Qualität von Schule und Unterricht. Kurzbericht über erste Ergebnisse einer Untersuchung an bayerischen Schulen*. Eichstätt: Katholische Universität.
- Helmke, A., Helmke, T., Schrader, F.-W. & Wagner, W. (2007). *Der Ratingbogen der DESI-Videostudie*. Landau: Universität Koblenz-Landau.
- Helmke, A. & Jäger, R. S. (2002). *Das Projekt MARKUS. Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext*. Landau: Empirische Pädagogik.
- Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M. et al. (2002). *PI-SA 2000: Dokumentation der Erhebungsinstrumente (Materialien aus der Bildungsforschung 72)*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Mayr, J. (1998). Mehr Flow als Frust. Qualitäten des Erlebens bei LehrerstudentInnen und LehrerInnen. In H. Altrichter, K. Krainer & J. Thonhauser (Hrsg.), *Chancen der Schule – Schule als Chance* (S. 235-250). Innsbruck: Studienverlag.
- Mayr, J., Eder, F. & Fartacek, W. (2005). *Linzer Diagnosebogen zur Klassenführung (LDK)*. Zugriff am 10.03.2012 von <http://ius.uni-klu.ac.at/projekte/ldk/>.
- Neuenschwander, M. P., Goltz, S., Balmer, T., Gasser, A., Hirt, U., Ryser, H. et al. (2003). *Eltern, Lehrpersonen und Schülerleistungen. Dokumentation des Schülerfragebogens*. Bern: Stelle für Forschung und Entwicklung.
- Schönbächler, M.T. (2005). *Klassenmanagement auf der Primarstufe. Dokumentation zur Datenerhebung bei den Lehrpersonen und bei den Schülerinnen und Schülern (Forschungsbericht Nr. 32)*. Bern: Universität Bern, Institut für Pädagogik und Schulpädagogik, Abteilung Pädagogische Psychologie.

Wannack, E., Herger, K., Gruber, M. & Barblan, A. (2009). *Classroom Management in der Schuleingangsstufe. Dokumentation der Fragebogenstudie*. Bern: PH Bern, Zentrum für Forschung und Entwicklung.

2.7. Autorinnenhinweis

Die Daten für die hier präsentierten Analysen stammen aus dem durch das Bundesministerium für Bildung und Forschung finanzierten Projekt „Kompetenzen des Klassenmanagements (KODEK). Entwicklung und Evaluation eines Fortbildungsprogramms für Lehrkräfte zum Klassenmanagement.“ (FKZ: 01JH0920, Leitung: Prof. Dr. Felicitas Thiel, Dr. Diemut Ophardt, Freie Universität Berlin).

Ich danke allen Schülern, Lehrkräften und beteiligten Testleitern für ihre engagierte Mitarbeit. Besonderer Dank gilt darüber hinaus Prof. Dr. Nagy und Markus Szczesny für ihre methodischen Hinweise zur Mehrebenenfaktorenanalyse. Schließlich möchte ich mich bei den anonymen Gutachtern für ihre wertvollen Kommentare und Anregungen zur Verbesserung des Manuskripts bedanken.

Die Autorin kann über folgende E-Mailadresse kontaktiert werden: v.piwowar@fu-berlin.de

3

Teilstudie 2: Training Inservice Teachers' Competencies in Classroom Management. A Quasi-Experimental Study with Teachers of Secondary Schools.

Notice: This is the authors' version of a work that was published in *Teaching and Teacher Education* (© 2013 by Elsevier):

Piwowar, V., Thiel, F. & Ophardt, D. (2013). Training Inservice Teachers' Competencies in Classroom Management. A Quasi-Experimental Study with Teachers of Secondary Schools. *Teaching and Teacher Education*, 30, 1-12.

[doi:10.1016/j.tate.2012.09.007](https://doi.org/10.1016/j.tate.2012.09.007)

Abstract

The current study evaluated the effectiveness of a training program for inservice secondary school teachers in classroom management. In a non-randomized pre-post-design, 19 teachers participated in a newly developed training (the intervention group) and 18 teachers participated in a control training (the control group). All participants reported better knowledge of classroom management after training. However, hypothesized positive effects on teachers' competencies and increased student engagement occurred only in the intervention group. These findings are supported by participants' reported high subjective validity of the training. Further research is needed to study sustainability of the observed effects.

Keywords: classroom management – teacher training – video – inservice teachers – secondary school – professional development

3. Training Inservice Teachers' Competencies in Classroom Management. A Quasi-Experimental Study with Teachers of Secondary Schools.

3.1. Introduction

Classroom management is one of the main domains of teacher expertise and a critical component in effective teaching (Brophy & Good, 1986; Gettinger & Kohler, 2006; Hattie, 2009; Kunter, Baumert, Blum, Klusmann, Krauss, & Neubrand, 2011; Seidel & Shavelson, 2007). Many studies have shown that successful classroom management enhances students' academic learning time by influencing their attention, engagement, and motivation positively (Oliver, Wehby, & Reschly, 2011; Wang, Haertel, & Walberg, 1993). For teachers, classroom management is one of the most enduring and widespread challenges (Everston & Weinstein, 2006a; Merrett & Wheldall, 1993; Pigge & Marso, 1997); difficulty in establishing and maintaining effective classroom management has shown to correlate with teacher burnout and job dissatisfaction (Brouwers & Tomic, 2000; Friedman, 2006; Melnick & Meister, 2008).

Classroom management competencies are an integral part of the "landscape of professional knowledge" (Leinhardt, McCarthy Young, & Merriman, 1995) and encompass a wide variety of skills and theoretical frameworks (Emmer, Evertson, & Worsham, 2003; Evertson & Weinstein, 2006a; Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008) that may no longer be best conceived as referring primarily to handling disciplinary problems (Evertson & Weinstein, 2006b). Characteristics of expert teaching such as "fluid performance" and "intuitive grasp of the situation" (Berliner, 2004, p. 22) are connected not only with the application of preventive strategies such as establishing rules, monitoring student behavior, or cultivating a functioning working alliance, but also with reactive strategies such as effectively dealing with disruption or resolving conflicts (Hardin, 2008; Little & Akin-Little, 2008). According to this broad understanding, classroom management goes beyond mere behavior management but also supports instruction and gives credit to the relevance of the teacher-student-relationship (Brophy, 2006; Manning & Bucher, 2007; Martin & Sass, 2010; Pianta, 2006).

The specific classroom management skills required vary depending on myriad factors. At secondary schools, required competencies differ markedly from those needed for managing primary students (Baer, 1999; Emmer et al., 2003; Emmer & Gerwels, 2006). For example, rates of aggressive and antisocial behaviors are higher in secondary schools, and the student-teacher relationship faces challenges that are less common or less severe in primary

schools (e. g. conflicts between students, adolescents' emotional and personal needs). Cognitive and organizational demands on secondary school students are higher as well (Emmer et al., 2003; Emmer & Gerwels, 2006; Pomeroy, 1999; Weinstein, 2007). All these conditions require specific skills for motivating, instructing and managing secondary school students compared to primary school students.

Numerous cross-sectional and longitudinal studies have shown that competencies in classroom management can be improved through specifically designed courses at universities or inservice teacher training programs (Evertson & Harris, 1999; Freiberg & Lapointe, 2006; Weinstein, 1999). However, most of these training programs are designed for primary schools (Freiberg & Lapointe, 2006) or focus on the needs of preservice teachers (Jones, 2006; Stough, 2006). Furthermore, Jones (2006) concluded that too many of those programs fail to support teachers in transferring their newly acquired skills into practice. There is still a need for effective inservice teacher programs in classroom management, particularly relating to secondary schools.

The purpose of this study was twofold: first, to evaluate the effectiveness of KODEK (Kompetenzen des Klassenmanagements: classroom management competencies), a training program developed by the authors for inservice secondary school teachers aimed at improving competencies in classroom management; second, to investigate the impact of different instructional methods on teachers' evaluation of specific training modules.

We designed KODEK with two objectives in mind: to convey empirically sound classroom competencies, and to use methods that have been proven reliable for training experienced teachers. In this paper, we first review what constitutes effective classroom management. Then, we suggest methods to provide appropriate learning opportunities for training experienced teachers. Finally, we outline the KODEK training program, our research hypotheses, and the results of our initial investigation.

3.1.1. Strategies and Prerequisites for Effective Classroom Management

Following the ecological approach (Doyle, 1984, 2006; Kounin, 1970), "classroom management is about how order is established and maintained in classroom environments" (Doyle, 2006, p. 99). Whether teachers succeed in classroom management "depends upon the strength and durability of the primary [. . .] vector of action" (Doyle, 2006, p. 106), which is strongly connected to a teachers' intended "programme [sic] of action" (p. 99).

The notions of program of action and vector of action are crucial to the ecological view of classroom management. In order to guide students' attention toward the target of the les-

son, teachers develop a program of action encompassing a planned sequence of activities. Experienced teachers frequently use “general guidelines for lessons that are designed to be responsive to the unpredictability of classroom events” (Borko & Livingston, 1989, p. 476). In a specific classroom situation, teachers must implement their program of action in response to the current circumstances by establishing a vector of action, which orients students’ attention and behavior and points it toward learning goals. “Once entered into, [vectors of action] pull events and participants along their courses” (Doyle, 2006, p. 102). In classrooms where the vector of action tends to be weak and unstable, teachers stabilize it by continuously signaling expected behavior. They must shield their program of action against disruptions, which Doyle (2006) calls “secondary vectors” (p. 114). According to this approach, student misbehavior is not a “property of an action” but “any behaviour [sic] by one or more students that is perceived by the teacher to initiate a vector of action that competes with or threatens the primary vector of action at a particular moment in a classroom activity” (Doyle, 2006, p. 112).

Order is the main precondition for an effective program of action. However, order in the classroom does not necessarily mean that students merely follow rigid rules or carry out standardized procedures (Doyle, 2006). The notion of appropriate order depends on contextual elements such as the curriculum, lesson subject, activity structures, and student ability (Berliner, 1983; Leinhardt, Weidman, & Hammond, 1987). The ecological approach underlines how the dynamic aspect of order can be described as a social condition: The teacher introduces expected conduct, but in the end the teacher and the students jointly constitute an orderly learning environment that “supports and facilitates both academic and social emotional learning” (Evertson & Weinstein, 2006a, p. 4).

To make effective use of the ecological approach, teachers need delicate and complex competencies, which are best thought of as comprising a holistic approach to classroom management, rather than a set of discrete techniques or skills (Brophy, 2006; Doyle, 2006). However, there are certain practices that have been shown to contribute consistently to successful classroom management, which we will classify broadly as preventive and reactive strategies (Hardin, 2008; Little & Akin-Little, 2008) in the following.

Preventive strategies were first identified by Kounin (1970), who found that successful teachers are capable of management strategies like withitness, overlapping, and group mobilization. Group mobilization supports student activation and leads to more engagement and on task behavior (Copeland, 1987; Kounin, 1970; Leflot, van Lier, Onghena, & Colpin, 2010). Being able to simultaneously engage students individually and the class as a whole (i.e. overlapping) is crucial for maintaining the vector of action (Doyle, 2006; McGarity & Butts, 1984;

Stallings, 1980). Continuously monitoring each student's learning activities while teaching allows teachers to acknowledge students who are engaging and to identify undesired developments early on (Doyle, 2006; Evertson & Emmer, 1982). This helps keep the learning process running while dealing with disruptions in a timely and unobtrusive manner, which in turn prevents ripple effects (Witt, van der Heyden, & Gilbertson, 2004). Moreover, teachers who provide a clear and structured program of action may encourage and enable students to engage in academic tasks over the course of the lesson (Chilcoat, 1989; Doyle, 1984; Smith, 1985).

Another important prerequisite for organizing and structuring learning processes is skillful time management (Coddling & Smyth, 2008; Doyle, 2006; Weinstein, 2007). Clear procedures and shared routines for recurring situations (e.g. procedures for taking attendance, distribution of working materials, and transitions) help teachers allocate academic learning time effectively (Arlin, 1979; Leinhardt et al., 1987; Nash, 2009; Woolfolk & Brooks, 1985). As Evertson and Harris (1999) pointed out, it is crucial to establish these routine interactions at the beginning of the school year. This enables teachers to use simple cues and prompts to initiate routines (Kounin & Gump, 1974), allowing more time for academic learning. A second step to scaffold interaction in classrooms is to set up clear behavior expectations, a system of sound rules that is established consensually with students (Evertson & Harris, 1999; Little & Akin-Little, 2008; Marzano, 2005).

Teachers use reactive strategies to deal with inattention and disruption such as misbehavior, talking, or call-outs (Hardin, 2008; Landrum & Kauffmann, 2006; Witt et al., 2004). Stabilizing and sheltering the primary vector against these disturbances while teaching is a challenging task. Teachers need to decide whether to ignore inappropriate student behavior and proceed with the current learning activity or to intervene and reprimand a disruptive student, a course that might lead to leaving the program of action (Doyle, 2006). Ideally, interventions dealing with disruptions should be appropriate, timely, and efficient (Doyle, 1984; Landrum & Kauffmann, 2006), and are more likely to be successful when referring to previously established rules (Emmer et al., 2003). Teachers also shelter the primary vector of action and maintain expected student behavior through praise and encouragement (Brophy, 1981; Chadwick & Day, 1971).

Beyond these individual practices, recent research has expanded the notion of classroom management to acknowledge that the quality of the teacher-student relationship is an important aspect (Pianta, 2006; van Tartwijk, den Brook, Veldman, & Wubbels, 2009). According to Evertson & Weinstein (2006a), teachers should operate as "warm demanders" (p.11):

empathic, responsive, caring and supportive with high expectations of their students. This kind of working alliance (Oevermann, 1996; Rogers, 2009) between students and teachers creates an environment that promotes learning: It can contribute to increased motivation and readiness to learn (Furrer & Skinner, 2003; Nie & Lau, 2009; Slavin, Hurley, & Chamberlain, 2003). How teachers handle conflict is a critical component in the teacher-student relationship—one that puts particular demands on teachers' competencies (Emmer et al., 2003; Ophardt & Thiel, 2008). Teachers can actively address issues for example by isolating, postponing or delegating the conflict (Johnson & Johnson, 2006; Walker, Ramsey, & Gresham, 2004).

3.1.2. Designing Learning Opportunities to Improve Inservice Teachers' Competencies

In this article we refer to Weinert's (2001) definition of competencies as "cognitive abilities and skills that individuals possess or are able to acquire which allow them to solve specific problems. Competencies encompass motivational, volitional and social dispositions and abilities that allow for successful and responsible application of the problem solutions in variable situations." (p. 27f; translated by the authors).

According to this definition, competencies are based on teacher knowledge, which can be classified by various domains of knowledge and forms for representing that knowledge, "some of which derive from research whereas others originate in the wisdom of practice" (Shulman, 1986, p. 9). Yet, the relevance of theoretical and practical knowledge for competent performance is discussed controversially (for an overview see Munby, Russel, & Martin, 2001). Theoretical, declarative knowledge enables teachers to perform "reasoning processes such as labeling, codifying, describing, analyzing and justifying" (Leinhardt et al., 1995, p. 402).

Content knowledge as a major part of theoretical teacher knowledge (Shulman, 1986) has been proved to be an important prerequisite to competent performance and student learning (Baumert, Kunter, Blum, Brunner, Voss, Jordan et al., 2010). In contrast, practical knowledge, being highly relevant for competencies of classroom management, is contextually developed knowledge which according to Leinhardt et al. (1995) is "situational, intuitive, and tacit" (p. 401). It is connected to a particular environment, can be applied flexibly with high speed, and its guiding principles remain usually unexplicated (Berliner, 2004). As a consequence, these two forms of knowledge exist "separately, discrete, non-overlapping, and inert" (Leinhardt et al., 1995, p. 402) and the practitioner might lack cognitive tools to revise and improve practical knowledge deliberately. For this reason knowledge integration and "fusing

experience and theory” (Bromme & Tillema, 1995, p. 261) seem to be powerful prerequisites to foster professional learning and “reflection on action” (Schoen, 1983, p. 68). Considering the tension between the two types of knowledge, integration is not trivial and involves enabling “learners to make universal, formal, and explicit knowledge that often remains situational, intuitive, and tacit; and to transform universal, formal, explicit knowledge for use in situ” (Leinhardt et al., 1995, p. 403).

Referring to this idea of knowledge integration, Simons and Ruijters (2004) suggested that professional learning should address and systematically connect practical and theoretical knowledge by using three principles: elaborating, expanding and externalizing knowledge.

Elaborating practical knowledge means to make procedural, usually implicit knowledge explicit and to “abstract principles from particulars” (Leinhardt et al., 1995, p. 403). Since elaboration should be directly supported, experienced teachers “need to think of concrete situations and experiences in order to become aware of their implicit learning outcomes” (Simons & Ruijters, 2004, p. 215). Research on teacher learning has shown that simulations (Grossman, 2005) and video based analysis (Brophy, 1988, 2004) are effective at providing situated learning and reflection of practical experience. Elaborating practical knowledge is also an important prerequisite to the other principles of professional learning.

Expanding declarative knowledge can be achieved by imparting elaborate schemata and concepts that refer explicitly to research theories and findings. When carefully connected to practical experience, expanding knowledge may contribute to reflective practice. Theoretical reasoning may be used to “slow down time, facilitate reflective revision, and examine the consequences of various actions” (Leinhardt et al., 1995, p. 404). Teachers get the opportunity to develop and transfer alternative strategies to their own set of classroom knowledge.

Externalizing knowledge refers to learning through collaborative exchange and reflection on practical knowledge. In teacher education, especially the application of video documented lessons in collaborative settings enables the discussion of contextual conditions and alternative options (Fukkink, Trienekens, & Kramer, 2010; Krammer & Reusser, 2005; Santagata & Guarino, 2011). In these settings new perspectives from peers encourage reflection and extension of the action repertoire.

Competencies of classroom management can be effectively developed by observational learning and role-playing because these methods comprise the processes of expanding and elaborating (Kubany, Sloggett, & Ogata, 1974). Microteaching, in particular, helps reduce the complexity of a real classroom situation, allowing teachers to explore single interventions and discuss them from different perspectives (Grossman, 2005)—key components in facilitating

the elaboration of practical knowledge and the integration of alternative strategies. In addition, video based analysis has been shown to initiate knowledge externalization and to enhance development of competencies in classroom management (Brophy, 2004; Hylton, 2000; Sariscsany & Pettigrew, 1997).

Several research-based programs suitable for secondary schools exist: The Classroom Organization and Management Program (COMP, Evertson & Harris, 1999), Consistency Management and Cooperative Discipline (CMCD, Freiberg, Stein, & Huang, 1995), a school-based program by Gottfredson, Gottfredson, & Hybl (1993), and Incredible Years (Shernoff & Kratochwill, 2007), a self-training program.

However, in Germany, there is only one classroom management training program (Münchener Lehrertraining—Munich's teacher training, Havers, 2001), and it targets university students and preservice teachers rather than inservice teachers. As we have demonstrated above, inservice teachers have distinct needs due to their accumulated practical knowledge. Also, there is an unmet need for classroom management training for inservice secondary school teachers in a German context.

3.1.3. KODEK: a New Inservice Teacher Training Program

KODEK, a teacher training program developed by the authors, comprises three modules, each with distinct learning opportunities and objectives (see Table 1). In the two sessions of Module 1, concepts and findings of classroom management research are conveyed to establish a theoretical framework for reflecting on teachers' classroom management practice and for introducing a wide spectrum of strategies of classroom management (*expanding*). The first module has three objectives: first, to supply the participant with research-based, domain-specific declarative knowledge on classroom management; second, to establish a framework to enable participants to discuss classroom situations using shared and precise terms; third, to support the extension of the established classroom management strategies through active construction of alternative strategies.

The first session starts with a lecture along with a video presentation and discussion. Trainers describe the depicted aspects of classroom management strategies to establish and maintain expected behavior, to direct learning activities, and to solve conflicts. Video analyses are used in two ways: to help transfer general principles of classroom management to practical examples, and to help teachers reflect on their own classroom management techniques. In the second session, participants analyze staged video cases in small groups in order to generate and discuss alternative strategies.

In Module 2, participants reflect their practical knowledge and try out new classroom management strategies in simulations using microteaching and role-playing methods (*elaboration*). The main objectives of Module 2 are to allow teachers to reflect on implicit routines in dynamic situations, to implement new strategies in a sheltered environment, and to examine multiple perspectives (i.e. teacher, student, and observer) in critical situations. Another objective is to introduce standards for giving feedback to fellow trainees, an important preparation for the third module.

The first session of Module 2 uses an adaptation of the traditional microteaching approach. We developed a script around a common classroom situation (disruptive behavior during the introduction of seatwork) with four roles: a) the teacher, b) students following the program of action, c) students showing deviant behavior, and d) detached observers who evaluate the teachers' performance using an observation form. Each simulation takes 5-10 minutes, after which it is discussed from each of the four perspectives. The second session consists of individual role-playing situations where teachers can practice conflict resolution strategies as well as discuss potential alternatives.

Table 3.1: Overview of the training program KODEK

	Learning Opportunities	Objectives	Number and Duration of Sessions
Module 1	- Lecture - Small group video analyses	- Establish common framework - Extend repertoire of classroom management strategies	2 x 4h
Module 2	- Microteaching - Role-playing	- Reflect implicit routines - Try out new strategies in staged situations - Introduce feedback	2 x 4h
Module 3	- Individual reflection on own video-taped performance ("video circle")	- Implement new strategies - Evaluate and discuss own practice	1 x 5h 2h preparation
Total			5 units (23 h)

In Module 3, "video circles" are implemented to allow teachers to reflect collaboratively on how successfully they were able to put the newly acquired principles of classroom management into practice (*externalization*). A video circle consists of a group of four participants mediated by a coach. After attending the first two modules, participants are videotaped in their classrooms while trying to implement the newly learned strategies. Each participant views the videotaped lesson individually and selects two segments of 2 to 4 minutes to be discussed. During the video circle session, participants analyze these selected segments and generate alternative strategies for how to manage the classroom. Referring back to the lessons

in the first module, the third module helps teachers implement new classroom management strategies by giving focused feedback and by generating alternative practices in a co-constructive way.

Participants receive a supplemental reader (Ophardt, Piwovar, Thiel, 2009) for further individual study.

3.1.4. Purpose and Research Hypotheses

Our first objective was to investigate KODEK's effectiveness in enhancing inservice secondary school teachers' knowledge and competencies relating to classroom management. We used a quasi-experimental design to examine this question. The intervention group (IG) participated in Modules 1, 2, and 3. The control group (CG) only participated in Module 1, which is presumed to represent the traditional instruction approach. Our second objective was to gain insight into the utility of three modules individually, as well as how they each impacted teachers' perception of the program.

Investigated outcome measures were organized with respect to Kirkpatrick's (1998) four levels of evaluation: reaction, learning, behavior, and results.

Hypothesis 1: With regard to learning outcomes, we first hypothesize that both the intervention group and the control group will demonstrate increased declarative knowledge of classroom management, as they both participated in learning opportunities that foster declarative knowledge.¹

Hypothesis 2: On the behavioral level, we must differentiate between expected effects on teachers' classroom management competencies in the long term and in the immediate term. We anticipate teachers' strategies to guide and monitor student behavior will be affected more immediately. However, the working alliance, teachers' manner of handling conflicts, and the effective use of rules and procedures will take more time to change notably, since they strongly depend on being established at the beginning of the school year. Consequently, we hypothesize that immediately after the KODEK training, we will see effects on teachers' competencies with regard to dealing with disruption effectively, time management, group mobilization, and clarity of program of action. No effects on competencies are expected in the control group.

Hypothesis 3: As a result, we hypothesize that the KODEK training will have an effect on students, specifically off-task behavior should decrease and student engagement should increase. No treatment effects on students are expected in the control group.

¹ The terms "knowledge" and "declarative knowledge" are used interchangeably in the following.

Hypothesis 4: Finally, we hypothesize increases over the course of the three modules in teachers' perceived ability for self-reflection regarding their own practice as well as perceived training transferability. These two aspects refer to teachers' reaction to the intervention and affect the subjective validity and evaluation of the implemented learning opportunities.

3.2. Method

3.2.1. Sample

A total of 37 secondary school teachers from nine urban schools in the area of Berlin and Brandenburg (Germany) volunteered to participate in the study. To enhance compliance, the teachers themselves decided whether to enter the intervention group or the control group. Of the 19 teachers in the intervention group (IG), 63% were female. Their ages ranged from 29 to 62 years ($M = 47.56$ years, $SD = 8.71$), and they had an average of 20.26 years of teaching experience ($SD = 10.52$). Of the 18 teachers in the control group (CG), 56% were female. These teachers were between 22 and 60 years of age ($M = 39.56$ years, $SD = 11.40$), and they had an average of 11.48 years of teaching experience ($SD = 10.12$).

Teachers rated themselves positively on teacher self-efficacy ($M_{CG} = 3.17$, $SD_{CG} = 0.40$; $M_{IG} = 3.03$, $SD_{IG} = 0.34$), assessed by an instrument developed by Schmitz & Schwarzer (2000) on a 4-point scale. Working stress was high ($M_{CG} = 3.88$, $SD_{CG} = 0.61$; $M_{IG} = 4.07$, $SD_{IG} = 0.50$) as measured by an instrument developed by Enzmann & Kleiber (1989) on a 5-point scale.

666 students were enrolled in the classes of the teachers who participated in the study. In the IG teachers' classes, 51% of the students were female. The mean age was 13.50 years ($SD = 1.18$, Range = 11-17 years); class sizes ranged from 10 to 29 students ($M = 23.1$, $SD = 4.34$). In the CG teachers' classes, 38% of the students were female. The mean age was 14.48 years ($SD = 1.39$, Range = 11-18 years); class sizes ranged from 17 to 28 students ($M = 21.44$, $SD = 2.87$).

The two groups were comparable with regard to teacher sex, teacher self-efficacy, and teacher work stress. Systematic differences were found with regard to teacher age, teaching experience, student age and student sex: IG teachers were older, $t(34) = -2.37$, $p = .024$ and reported more years of teaching experience, $t(35) = -2.59$, $p = .014$. Students were older in CG teachers' classes, $t(34) = 2.29$, $p = .028$; there were more female students in IG classes compared to CG classes, $t(34) = -2.84$, $p = .009$.

3.2.2. Design

Effectiveness and Treatment Conditions: We employed a quasi-experimental pre-post design to study the effects of the KODEK training program. Whereas the intervention group participated in the entire KODEK program, the control group attended only Module 1. Participants' knowledge on classroom management, participants' competencies in classroom management and their students' behavior in class (dependent variables) were assessed before and after the intervention using four distinct measurements. Knowledge was assessed using a self-rating. Students and observers rated teachers' competencies in classroom management. Student behavior (student attention and off-task behavior) was assessed by an additional observer rating. We conducted pre-post comparisons of the dependent variables to provide an indirect measurement of change.

Subjective validity and learning opportunities: An acceptability rating for participants served as an indicator for acceptance of the intervention, as well as to compare the three modules with each other regarding two aspects: first, its usefulness for teachers' own practice (transferability) and second, the amount of reflection on own teaching in each module. The rating was completed after each session to gather information on immediate reactions to the training sessions.

Only non-personalized information was collected from teachers and students. Consent was obtained from school principals, teachers, students, and the parents of students younger than 14 years old.

3.2.3. Measures

All instruments assessing the dependent variables of classroom management were developed by our research group (Piwowar, Ophardt & Thiel, 2011). These instruments constitute a multidimensional assessment of classroom management, putting special focus on preventive and proactive strategies. The following distinct dimensions were assessed and described for most effective teacher behavior:

- a) "Rules": A system of rules is established and works well: students stick to that system and have internalized it
- b) "Procedures": The teacher has established procedures and has transformed patterns of activity into smooth routines (e.g. routines for attendance check, distribution of working materials and transitions)

- c) "Dealing with disruption" (reactive strategy): The teacher's reactions to disruptive students are timely, efficient, and unobtrusive. Reprimands and interventions are appropriate for the misbehavior and mindful of the working alliance
- d) "Time management": Allocated time, instructional time, engaged time, and academic learning time are used efficiently; activities and transitions are well prepared and organized; the lesson proceeds smoothly
- e) "Monitoring": The teacher keeps track of all student activities while teaching and is able to detect undesired developments early on
- f) "Working alliance": A collaborative and productive working alliance between students and the teacher is denoted by cohesion, mutual engagement and an environment that promotes learning (e.g. encouraging comments, praise, error-friendliness)
- g) "Conflicts among students" (reactive strategy): The teacher is actively involved in the resolution of conflicts, intervenes fairly and constructively
- h) "Group mobilization": The teacher maintains group focus by involving and activating every single student and the class as a whole; students are held accountable
- i) "Clarity of program of action": The teacher's intended program of action (single tasks as well as the progression of the entire lesson) becomes clear and transparent to the students.

Knowledge on classroom management

To obtain an assessment of knowledge, we asked teachers to fill out a form rating themselves. A written introduction instructed teachers to refer to their theoretically available knowledge only, disregarding whether they succeeded in real classroom situations. Teachers rated statements such as "I am familiar with various strategies for involving as many students as possible actively in the classroom" (group mobilization) and "I am familiar with strategies for dealing with severe problem behavior" (Dealing with disruption) on an 8-point scale (1 [*strongly disagree*] to 8 [*strongly agree*]). Each scale consisted of three items. Internal consistency computed using coefficient Cronbach's Alpha was good to excellent ($M = .88$, $Md = .87$, Range = .82; .95). Monitoring was not modeled as part of the knowledge rating because it cannot be clearly formulated as declarative knowledge.

Competencies in classroom management

Two informant ratings were administered to measure teachers' competencies in classroom management: a student rating and an observer rating.

Students were asked to rate their teachers' competencies in classroom management on a 4-point scale (1 [*strongly disagree*] to 4 [*strongly agree*]). Items were designed to be low-inferent. Sample items included: "This teacher does not notice when students are busy with something off-topic.(-)" (monitoring), and "This teacher makes sure that everyone is involved" (group mobilization). Scales consisted of four to six items. Internal consistency assessed by Cronbach's Alpha was acceptable to good ($M = .77$, $Md = .77$, Range = .71; .86). Intraclass coefficients ($ICC[2]$) for aggregated values on class level were good to excellent ($M = .89$, $Md = .90$, Range = .82; .93). Criterion related validity was established by medium to high correlations to observer ratings on classroom management and medium to high correlations with on-task behavior and student engagement.

A high-inferent observer rating was adopted to assess participants' competencies in classroom management from an impartial perspective. Two observers per lesson independently rated the items on a 6-point scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). Sample items for the rating included: "The teacher is aware of everything that is going on in the classroom." (monitoring), and "The teacher involves all students actively in the classroom." (group mobilization). Scales are represented by one to three items. Means per scale were aggregated for the two observers and then used as outcome measures. Coefficient Alpha for the three-item scales "group mobilization" and "working alliance" were good (.81) and excellent (.92), respectively. Intraclass coefficients ($ICC[2]$) showed acceptable to excellent reliability for aggregated observer judgments ($M = .88$, $Md = .89$, Range = .75; .95). This rating did not include the scale "conflicts among students" because such incidents are comparatively rare and often take place outside the classroom. Criterion related validity was established by medium to high correlations to student ratings on classroom management and medium to high correlations with on-task behavior and student engagement. Observers participated in three days of training, which included a detailed manual and exercises with videotapes. Ratings were conducted blind to the condition.

Student engagement and student on task behavior

We used the Munich Observation Inventory (MAI) (Helmke, 1988) to measure student behavior. The MAI is a standardized rating instrument that allows observers to systematically assess students' attention and engagement in class. Student attention is operationalized as on-task and off-task behavior as a low inference measure: For an interval of 5 seconds, each student is observed and rated to be on-task (0) or off-task (1). On-task students are characterized by active participation (raising one's hand, answering a question) as well as passive participa-

tion (listening, taking notes). Off-task codes cover disruptive and undesirable behaviors (talking with one's neighbor, walking around without permission) but also passive behaviors (looking out of the window, drawing on the table). The outcome measurement is the mean percentage of student off-task time per lesson. In addition, student engagement is rated every two minutes on a 5-point, high inferent scale (1 [*low engagement*] to 5 [*high engagement*]) regarding three aspects: number of engaged students, intensity of student engagement, and duration of student engagement within the two minute interval. The average of these ratings indicates the mean student engagement per lesson. Reliability coefficients and validity coefficients were reported to be good (Helmke, 1992). Student engagement and attention was rated by one observer per lesson who was blind to the condition. Observers had been trained in a two day workshop.

Acceptability

An acceptability rating was conducted referring to Lipowski's (2010) criteria of relevant aspects to evaluate interventions and aspects of subjective validity. After each session, participants evaluated personal relevance, acceptance of used methods, organization of the intervention, trainers' competencies, and relevance on an 8-point scale (1 [*strongly disagree*] to 8 [*strongly agree*]). To obtain measures on amount of reflection on own practice and the transferability of the training, two 5-item scales were designed which indicated good internal consistency (Cronbach's Alpha coefficient at .81 and .87, respectively). Sample items included "I became more aware of my action in the classroom." (amount of reflection) and "I've learned things which I can immediately adapt to my teaching practice." (transferability).

3.2.4. Data analysis

Data were analyzed using SPSS 19. Analysis of Covariance (ANCOVA) and repeated measures Analyses of Variance (ANOVA) were employed to analyze the data.

Analysis of Covariance (ANCOVA) controlling for pretest values as a covariate was conducted to compare mean group improvement after the intervention regarding classroom management competencies and student behavior outcomes (second and third hypotheses). Assumptions for conducting an ANCOVA were met. There were no statistically significant differences at baseline between the two groups, except from students' ratings on group mobilization.

Repeated measures Analyses of Variance (ANOVA) with factors time (pretest vs. posttest) and group (control group vs. intervention group) were employed to test the expected increase of knowledge in classroom management (first hypothesis). The same procedure was

used to test the hypothesized increase in reflection and transferability over the course of the modules (Module 1 vs. Module 2 vs. Module 3, fourth hypothesis) in the IG. All assumptions for the repeated measures ANOVA were met. There were no statistically significant differences at baseline among the two groups.

For all statistical analyses, the Alpha-level was set at $p < .05$. Partial eta square (η_p^2) served as an indicator of effect size (Richardson, 2011), which are judged to be small at .01, medium at .06 and large at .14 (Cohen, 1988).

Descriptive statistics were used to illustrate the acceptability of the training in the separate intervention groups. The assessments after each session of the IG were summarized to a mean value in order to compare the outcomes of the IG to those of the CG, which only attended Module 1.

3.3. Results

The results and discussion sections are organized to reflect three topics of analysis: pre-post-tests indicating effectiveness of the training with regard to knowledge, competencies and influence on student behavior, acceptability of the training, and differences between the modules regarding perceived transferability and reflection.

3.3.1. Pre-post comparisons

Effects on knowledge

Pretest means, posttest means, standard deviations, and results of repeated measures analysis of variance with regard to knowledge on classroom management are illustrated in Table 2. Referring to descriptive statistics, group means increased compared to baseline on all dimensions of classroom management in both groups, except for a decrease in CG teachers' time management, $M_{CG,t1} = 6.52$, $SD_{CG,t1} = 1.40$; $M_{CG,t2} = 6.09$, $SD_{CG,t2} = 1.43$, group mobilization, $M_{CG,t1} = 6.28$, $SD_{CG,t1} = 1.07$; $M_{CG,t2} = 6.06$, $SD_{CG,t2} = 1.31$, and clarity of program of action, $M_{CG,t1} = 6.31$, $SD_{CG,t1} = 1.16$; $M_{CG,t2} = 6.26$, $SD_{CG,t2} = 1.37$.

ANOVAs revealed significant change over time for all participants on the dimensions of rules, $F(1,35) = 7.15$, $p = .011$, $\eta_p^2 = .17$, procedures, $F(1,35) = 8.01$, $p = .008$, $\eta_p^2 = .19$, dealing with disruption, $F(1,35) = 12.62$, $p = .001$, $\eta_p^2 = .27$, and working alliance, $F(1,35) = 16.33$, $p < .001$, $\eta_p^2 = .32$, indicating large effects. Two further medium effects occurred regarding conflicts among students, $F(1,35) = 2.21$, $p = .109$, $\eta_p^2 = .07$, and clarity of program of action, $F(1,35) = 2.09$, $p = .157$, $\eta_p^2 = .06$, neither was statistically significant. There were medium interaction effects of group and time on the dimensions rules,

$F(1,35) = 3.25, p = .080, \eta_p^2 = .09$, procedures, $F(1,35) = 2.55, p = .119, \eta_p^2 = .07$, dealing with disruption, $F(1,35) = 2.86, p = .100, \eta_p^2 = .08$, and working alliance, $F(1,35) = 2.83, p = .132, \eta_p^2 = .06$, indicating a stronger increase in knowledge in the IG than in the CG; medium interaction effects on the dimensions time management, $F(1,35) = 4.31, p = .045, \eta_p^2 = .11$, group mobilization, $F(1,35) = 3.90, p = .056, \eta_p^2 = .10$, and clarity of program of action, $F(1,35) = 3.55, p = .068, \eta_p^2 = .09$ indicate increased mean knowledge in the IG but decreased mean knowledge in the CG. Only the interaction regarding time management revealed a significant effect.

Table 3.2: Pretest Means, Posttest Means, Standard Deviations, and Repeated Measures Analysis of Variance Results of Knowledge on Dimensions of Classroom Management.

	Main Effects									Interaction		
	Pretest	Posttest	Group (G)			Time (T)			G x T			
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>F</i>	<i>p</i>	η_p^2	<i>F</i>	<i>p</i>	η_p^2	<i>F</i>	<i>p</i>	η_p^2	
Rules			.14	.713	.00	7.15	.011	.17	3.25	.080	.09	
CG	6.30 (1.27)	6.43 (1.10)										
IG	6.14 (1.01)	6.81 (0.65)										
Procedures			.05	.880	.00	8.01	.008	.19	2.55	.119	.07	
CG	6.21 (1.45)	6.39 (1.17)										
IG	6.04 (0.87)	6.67 (0.75)										
Dealing with disruption			.53	.472	.02	12.62	.001	.27	2.86	.100	.08	
CG	5.72 (1.58)	6.07 (1.29)										
IG	5.66 (1.12)	6.65 (0.74)										
Time management			.45	.508	.01	.00	.956	.00	4.31	.045	.11	
CG	6.52 (1.40)	6.09 (1.43)										
IG	6.33 (1.18)	6.74 (0.71)										
Working alliance			.71	.405	.02	16.33	<.001	.32	2.38	.132	.06	
CG	5.46 (1.58)	5.93 (1.46)										
IG	5.47 (1.28)	6.51 (0.64)										
Conflicts among students			.94	.340	.03	2.21	.109	.07	1.63	.211	.04	
CG	6.06 (1.14)	6.11 (1.51)										
IG	6.18 (0.84)	6.61 (0.65)										
Group mobilization			.49	.491	.01	.73	.398	.02	3.90	.056	.10	
CG	6.28 (1.07)	6.06 (1.31)										
IG	6.09 (1.06)	6.65 (0.78)										
Clarity of program of action			.50	.486	.01	2.09	.157	.06	3.55	.068	.09	
CG	6.31 (1.16)	6.26 (1.37)										
IG	6.32 (1.05)	6.74 (0.77)										

Note. Means are based on 8-point scales. 1 = low declarative knowledge; 8 = high declarative knowledge. CG = control group, IG = intervention group. $n_{CG} = 18, n_{IG} = 19. df = 1, 35$.

Effects on competencies²

Students' ratings on competencies. Descriptive statistics showed that IG teachers had higher scores than CG teachers on all tested dimensions of classroom management at posttesting (see Table 3). Whereas all means increased or remained stable in the IG compared to baseline, means in the CG remained stable or decreased.

ANCOVAs indicate large positive effects at posttesting in favor of IG teachers regarding monitoring, $F(1,32) = 5.31, p = .028, \eta_p^2 = .14$, and group mobilization, $F(1,32) = 25.00, p < .001, \eta_p^2 = .44$. Medium positive effects occurred regarding teachers' time management, $F(1,32) = 2.95, p = .096, \eta_p^2 = .08$, and clarity of program of action, $F(1,32) = 4.39, p = .044, \eta_p^2 = .12$. Only the effects on monitoring, group mobilization, and clarity of program of action revealed statistically significant results.

Observers' ratings on competencies. Descriptive statistics showed that IG teachers had higher scores on all tested dimensions of classroom management at posttesting, except for time management, $M_{IG} = 4.96, SD_{IG} = .19; M_{CG} = 5.16, SD_{CG} = .20$, respectively (see Table 3). Compared to baseline, all means increased in both groups, except for a decrease in monitoring in the CG, $M_{CG,t1} = 4.36, SD_{CG,t1} = 1.00; M_{CG,t2} = 3.96, SD_{CG,t2} = 0.28$. ANCOVAs revealed no statistically significant differences between groups after training, but there were medium positive effects regarding monitoring and group mobilization, $F(1,30) = 2.56, p = .120, \eta_p^2 = .08$; $F(1,30) = 2.75, p = .108, \eta_p^2 = .08$, respectively.

Effects on students

As shown in Table 3, off task behavior decreased when compared to baseline in both groups. Students in IG classes showed off-task behavior 26% of the time at posttests on average ($SD_{IG} = 2.7$) compared to 30% in CG classes ($SD_{CG} = 2.8$) which is neither statistically significant nor indicative of a sizeable effect, $F(1, 30) = 0.66, p = .422, \eta_p^2 = .02$.

Student engagement was higher in IG classes than in CG classes after training, indicating a statistically significant and large positive effect, $F(1, 30) = 15.55, p < .001, \eta_p^2 = .34$. Whereas mean student engagement increased in IG, it decreased in CG compared to baseline.

² Analyzed were only those dimensions of classroom management that were specified in the second hypothesis (3.1.4) with regard to improving competencies. In each group (IG and CG) one teacher was absent at pretesting and another teacher at posttesting.

Table 3.3: Pretest Means, Adjusted Posttest Means, Standard Deviations, and Analysis of Covariance Results of Competencies on Dimensions of Classroom Management and on Student Behavior.

	Pretest		Posttest		<i>F</i>	<i>p</i>	η_p^2
	CG	IG	CG	IG			
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M_{adj}</i> (<i>SD</i>)	<i>M_{adj}</i> (<i>SD</i>)			
Dimensions of Classroom Management							
Dealing with disruption							
Students ^a	3.04 (0.31)	3.16 (0.29)	3.05 (0.06)	3.15 (0.06)	1.79	.191	.05
Observers ^b	4.75 (0.90)	4.67 (0.93)	4.76 (0.23)	5.03 (0.23)	.68	.417	.02
Time management							
Students ^a	3.05 (0.39)	3.06 (0.31)	2.95 (0.06)	3.10 (0.06)	2.95	.096	.08
Observers ^b	4.93 (0.87)	4.76 (0.79)	5.16 (0.20)	4.96 (0.19)	.54	.469	.02
Monitoring							
Students ^a	3.02 (0.28)	3.08 (0.23)	2.97 (0.05)	3.13 (0.05)	5.31	.028	.14
Observers ^b	4.36 (1.00)	4.06 (1.04)	3.96 (0.28)	4.60 (0.27)	2.56	.120	.08
Group Mobilization							
Students ^a	3.09 (0.35)	2.60 (0.21)	2.73 (0.10)	3.07 (0.09)	25.00	<.001	.44
Observers ^b	4.25 (0.93)	4.22 (0.86)	4.42 (0.21)	4.92 (0.21)	2.75	.108	.08
Clarity of program of action							
Students ^a	3.25 (0.36)	3.22 (0.38)	3.11 (0.06)	3.30 (0.06)	4.39	.044	.12
Observers ^b	4.59 (0.84)	4.46 (0.80)	4.98 (0.18)	5.09 (0.18)	.20	.660	.01
Student behavior							
Amount of off-task ^c	34.32 (19.72)	31.23 (13.50)	29.50 (2.80)	26.30 (2.70)	.66	.422	.02
Student engagement ^d	3.20 (0.58)	3.06 (0.79)	2.75 (0.17)	3.67 (0.16)	15.55	<.001	.34

Note. CG = control group, IG = intervention group

^a Means are based on 4-point scales. 1 = low competencies; 4 = high competencies. $n_{CG} = 17$. $n_{IG} = 18$. $df = 1, 32$.

^b Means are based on 6-point scales. 1 = low competencies; 6 = high competencies. $n_{CG} = 16$. $n_{IG} = 17$. $df = 1, 30$.

^c in %. $n_{CG} = 16$. $n_{IG} = 17$. $df = 1, 30$. ^d Means are based on 5-point scales. 1 = low engagement. 5 = high engagement. $n_{CG} = 16$. $n_{IG} = 17$. $df = 1, 30$

3.3.2. Acceptability

Both CG teachers and IG teachers reported an overall favorable view of their training sessions. Ratings on acceptability of the training were based on 8-point scales (1[disagree] and 8[agree]) and were higher in the IG than in the CG; highest acceptability referred to the organization of the training ($M_{CG} = 6.94$, $SD_{CG} = 0.94$, $M_{IG} = 7.46$, $SD_{IG} = 0.36$) and trainer competence ($M_{CG} = 6.76$, $SD_{CG} = 1.09$, $M_{IG} = 7.51$, $SD_{IG} = 0.27$) in both groups. Training sessions were judged to be educational ($M_{CG} = 5.83$, $SD_{CG} = 1.69$; $M_{IG} = 6.69$, $SD_{IG} = 0.69$, respectively) and personally relevant ($M_{CG} = 5.67$, $SD_{CG} = 1.88$; $M_{IG} = 6.72$, $SD_{IG} = 0.83$, respectively). Total scores regarding amount of reflection and transferability were medium in both

groups ($M_{CG} = 4.81$, $SD_{CG} = 1.09$; $M_{IG} = 6.13$, $SD_{IG} = 0.84$, and $M_{CG} = 4.50$, $SD_{CG} = 1.10$; $M_{IG} = 5.67$, $SD_{IG} = 1.09$, respectively).

3.3.3. Amount of Reflection and Transferability over the Course of the Modules

IG teachers participated in the whole training program—all three modules of KODEK. As shown in Figure 1, perceived amount of reflection and transferability increased over the course of the three modules ($M_1 = 5.54$, $SD_1 = 0.94$; $M_2 = 6.35$, $SD_2 = 0.81$; $M_3 = 6.79$, $SD_3 = 0.98$, and $M_1 = 5.13$, $SD_1 = 1.29$; $M_2 = 5.62$, $SD_2 = 1.08$; $M_3 = 6.46$, $SD_3 = 1.21$, respectively). Repeated measures ANOVA indicate that these increases are statistically significant, $F(2,34) = 18.02$, $p < .001$, $\eta_p^2 = .52$, and $F(2,34) = 18.37$, $p < .001$, $\eta_p^2 = .52$, respectively. With regard to amount of reflection, post hoc tests using the Bonferroni correction showed that Module 2 and Module 3 elicited a significant increase when compared to Module 1 ($p = .003$, $p = .001$, respectively). Post hoc tests for transferability showed a significant increase from Module 1 to Module 3 as well as from Module 2 to Module 3 ($p < .001$, $p = .003$, respectively).

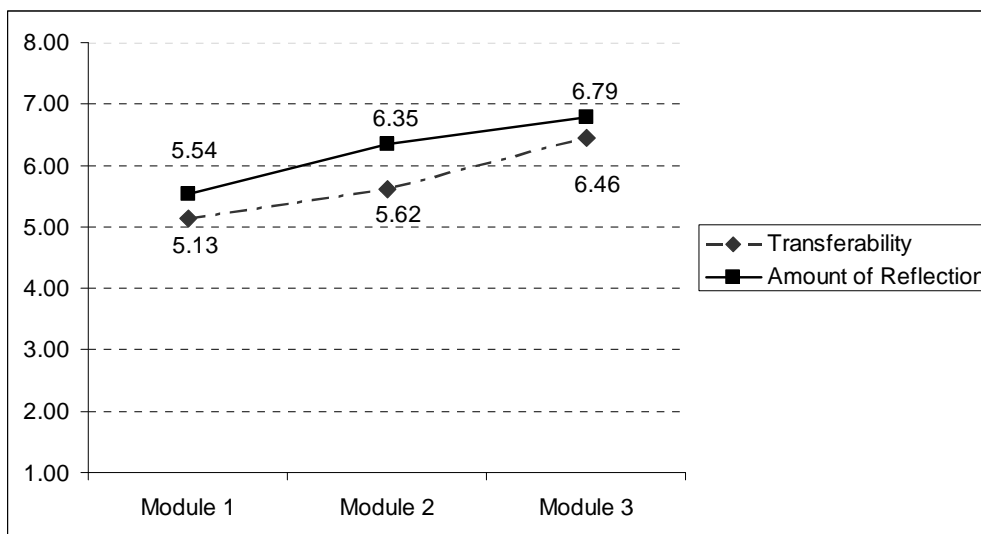


Figure 3.1. Mean scores for 'Amount of reflection' and 'Transferability' over the course of the three Modules. Means are based on 8-point scales. 1 = low; 8 = high. $n = 18$. $df = 2, 34$.

3.4. Discussion

In this study, we investigated whether participation in KODEK (treatment of the intervention group, IG), would improve teachers' classroom management competencies compared to participation in a training program corresponding to the traditional instruction approach (treatment of the control group, CG). The first subject of analysis was whether teachers' declarative knowledge on classroom management increased after training for all participants (first hypothesis). Indeed, reported declarative knowledge increased considerably on six out of eight dimensions compared to baseline scores (high to medium effects, in descending order):

working alliance, dealing with disruption, procedures, rules, conflicts among students and clarity of program of action. There are even stronger increases in IG teachers' mean knowledge on these six dimensions, except for the knowledge of handling of conflicts among students. Contrary to our hypotheses, knowledge of time management and group mobilization—the remaining two dimensions—increased for the IG but decreased for the CG.

These findings show that both KODEK and traditional training approaches can positively affect declarative knowledge on classroom management. However, KODEK, which applied all principles of professional learning (i.e. elaborating, expanding, and externalizing) and aligned training to practical experience, even had a stronger positive influence on most dimensions of teachers' declarative knowledge. It is important to note that the results regarding declarative knowledge are based on self-report measures. As Howard and Dailey (1979) have shown, indirect measurement of change using a subjective measure entails the risk of a "response shift." This shift occurs when a participant uses a different internal understanding of the construct being measured to complete the pretest and the posttest. As a consequence, non-significant findings or even unexpected decreases in dependent variables can occur, even though latent positive changes exist (e.g. due to pretest overestimation or biased self-presentation, Aiken & West, 1990). Given these methodological concerns, the decrease in time management knowledge in the CG could be due to a ceiling effect in combination with response shift: Time management is the area where teachers reported highest knowledge at baseline.

Per our second hypothesis, we expected IG teachers to perform better after training than CG teachers in the measured dimensions of classroom management: preventive strategies relating to time management, monitoring, group mobilization, and clarity of program of action, as well as reactive practice relating to dealing with disruption. However, the findings relating to improvement in competencies were mixed. When rated by students, all assessed preventive practices of classroom management improved in the IG over time and were greater than in the CG at posttesting. In contrast, the quality of reactive strategies did not change, although teachers reported an increase in declarative knowledge. These findings could indicate that teachers' practical knowledge is focused on reactive strategies. KODEK might have enabled teachers to better explicate their knowledge, but it is possible that greater potential for improving competencies lays with preventive and proactive management skills.

Observer ratings did not validate all of the hypothesized effects either. Only on two out of five tested dimensions did observers' ratings show medium positive effects in favor of IG teachers at posttesting: first, strategies of group mobilization and second, monitoring student

behavior. Although results with regard to teachers' competencies revealed some positive trends and effects favoring KODEK, there was not sufficient evidence to conclude that IG teachers enhanced their competencies in all hypothesized areas of classroom management.

These inconsistent findings could derive from the fact that observers rated teachers' competencies in classroom management based on only one sample before and after training (i.e. a single 45 minute lesson). However, a teacher's classroom management strongly depends on lesson subject as well as activity structures; these go hand in hand with different demands on students that influence their interactions with each other, their learning processes, and their academic and social engagement (Berliner, 1983; Doyle, 2006; Leinhardt et al., 1987). So the fact that there was only one observation session might have influenced the internal validity of the results, because the assessed lessons before and after training might not have been entirely comparable with respect to the classroom management demands on teachers. In contrast, students can refer to many samples of classroom management skills, which is a major advantage of student ratings on classroom variables (Clausen, 2002; Webb, Campbell, Schwartz, & Sechrest, 1966).

The students' reported improvement of preventive classroom management strategies is confirmed by a change in their behavior (third hypothesis). A large positive effect occurred in the student engagement in IG teachers' classrooms. Moreover, off-task behavior decreased to a greater degree than in CG classes. This did not yield a significant effect, but it indicates a trend which might become more pronounced when testing for follow-up developments. More samples of classroom teaching might have been useful to gain more representative data on off-task behavior as it was described for observations of teacher behavior.

As measured by learning (teacher knowledge) and behavior outcomes (teacher competencies, student behavior), the learning opportunities presented in KODEK proved to be appropriate and mainly successful. These findings are supported by results of the acceptability ratings conducted after each of the three modules in the IG. These ratings showed significantly increased reflection and transferability, as was proposed (fourth hypotheses). Concordant with the aims of the implemented learning arrangements, lecture and video analyses in Module 1, microteaching and role playing in Module 2, and the analysis of own videotaped teaching in Module 3 each affected reflection and transferability differently. The stronger increase with regard to transferability took place between Module 2 and Module 3. This validates the use of analyzing one's own videotaped teaching in small groups in order to maximize subjective usefulness for teachers' own practice.

A special note must be made with regard to the fact that some of the effects are due to a

decrease in outcome measures for CG teachers. One possible explanation for these effects is the demands inservice teachers face when trying to improve previously practiced classroom management and reorganize practical knowledge in that domain. The CG training could have caused negative short-term effects; it is possible that while focusing on improving certain aspects of classroom management, CG teachers inadvertently allowed other aspects of their practice to deteriorate. The methods that were implemented with the IG in Module 2 and 3 but not in the CG proved essential to successfully transfer acquired knowledge into teachers' own classroom management. These learning opportunities initiated the processes of elaborating and externalizing knowledge and consequently helped develop expertise (Simons & Ruijters, 2004). This point is of special interest because CG teachers were on average younger and less experienced than IG teachers. Differing short-term and long-term effects on less experienced and more experienced teachers could be investigated in order to find out more about underlying learning principles of inservice teachers. A second possible explanation might be drawn from the study's quantitative approach, where classroom management was operationalized as a set of discrete competencies. While we believe in the value of the quantitative approach presented here, it likely did not capture all facets of classroom management in their full complexity.

3.4.1. Limitations

The current study was limited in several ways. First, the relatively small sample of secondary school teachers may have restricted the power of statistical measures. Although the IG consistently had scores above the CG, some of these practically relevant effects were not statistically significant. In addition, differential effects due to school type, motivation, or years of teaching experience could be investigated with larger samples.

Second, there were specific limitations on the design of the study. The pre-post-tests methodology failed to test for the sustainability of the measured effects or to detect sleeper effects, which are to be expected in the working alliance, the handling of conflicts and the implementation of rules and procedures. In addition, internal validity could have been threatened due to non-randomization. For example, self-selection bias (e.g. different motivation, level of work load, maturation between groups) may have affected the sample, even though the two groups were comparable with regard to pretest scores on measures of declarative knowledge, competencies and student behavior. Furthermore, treatment conditions were not comparable regarding the length of training which might have triggered a dose-effect-relationship (Howard, Kopta, Krause, & Orlinsky, 1986).

Third, more sensitive measures could have been used. Instead of subjective ratings on knowledge, an objective test could have been carried out. In addition, there could have been multiple observations before and after the training as well as open ended classroom observations and case-studies based on qualitative methods that go beyond the assessment of discrete practices. This both could have broadened the study's view of classroom management and might have highlighted its contextualized and situated nature more comprehensively. We decided against these advantages of measurement in order to raise attendance and avoid dropout due to such time-costly requirements. However, all these issues together affect the generalizability of the results and take a limited view on processes of change.

3.4.2. Conclusions

Capturing and developing teachers' competencies that are based on practical, situated and implicit knowledge is one of the most challenging affordances to teacher education and teacher development (Berliner, 2004; Leinhardt et al., 1995; van Driel & Berry, 2012). KODEK is based on principles of professional learning (Simons & Ruijters, 2004) and was developed in order to increase teachers' classroom management skills under conditions of typical practice. In mediated video circles in Module 3, teachers reported the highest amount of reflection and transferability to their teaching. Despite the lack of congruent improvement in competencies, teachers who participated in KODEK exceeded teachers who participated in the training based on the traditional instruction approach on the majority of the conducted measures. The strongest indicators of KODEK's effectiveness are students' perceived improvement in their teachers' preventive classroom management strategies and increased student engagement. With regard to declarative knowledge, our findings emphasize that when training experienced teachers, it is crucial to integrate new ideas into their own practice.

In summary, our study demonstrated that a teacher training program using lecturing, simulations (role-playing and microteaching), and mediated video circles to analyze own videotaped teaching could succeed in improving experienced teachers' classroom management competencies, especially student engagement. In addition, teachers reported high subjective validity for the training: They found the training to be well organized, educational and relevant to them personally. Further investigation is needed to ascertain which learning processes lead to improvement and how they operate, how sustainable the detected effects are, as well as if and which long-term effects might occur when teachers start off with a new class at the beginning of the year. In addition, intervention effects on working stress and teacher self-efficacy could be investigated.

3.5. References

- Aiken, L., & West, S. (1990). Invalidity of true experiments: Self-report pretest biases. *Evaluation Review*, 14(4), 374-390. doi: 10.1177/0193841X9001400403
- Arlin, M. (1979). Teacher transitions can disrupt time flow classrooms. *American Educational Research Journal*, 16(1), 42-56. doi: 10.2307/1162402
- Baer, J. (1999). Adolescent development and the junior high school environment. *Social Work in Education*, 21(4), 238-248. doi: 10.1093/cs/21.4.238
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133-180. doi:10.3102/0002831209345157
- Berliner, D. C. (1983). Developing conceptions of classroom environments: Some light on the T in classroom studies of ATI. *Educational Psychologist*, 18(1), 1-13. doi: 10.1080/00461528309529256
- Berliner, D. C. (2004). Expert teachers: Their characteristics, development and accomplishments. In R. Batllori i Obiols, A. E. Gomez Martinez, M. Oller i Freixa, & J. Pages i Blanch (Eds.), *De la teoria...a l'aula: Formacio del professorat ensenyament de las ciències socials* (pp. 13-28). Barcelona, Spain: Departament de Didàctica de la Llengua de la Literatura I de les Ciències Socials, Universitat Autònoma de Barcelona.
- Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal*, 26(4), 473-498. doi:10.3102/00028312026004473
- Bromme, R., & Tillema, H. (1995). Fusing experience and theory: The structure of professional knowledge. *Learning and Instruction*, 5(4), 261-267. doi:10.1016/0959-4752(95)00018-6
- Brophy, J. (1981). Teacher praise: A functional analysis. *Review of Educational Research*, 51(1), 5-31. doi:10.2307/1170249.
- Brophy, J. (1988). Educating teachers about managing classrooms and students. *Teaching and Teacher Education*, 4(1), 1-18. doi:10.1016/0742-051X(88)90020-0
- Brophy, J. (Ed.). (2004). *Using video in teacher education*. Oxford: Elsevier.
- Brophy, J. (2006). History of research on classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management. Research, practice and contemporary issues* (pp. 17-43). Mahwah, NJ: Lawrence Erlbaum Associates.

- Brophy, J., & Good, T. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–375). New York, NY: Macmillan.
- Brouwers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education, 16* (2), 239–253. doi:10.1016/S0742-051X(99)00057-8
- Chadwick, B. A., & Day, R. C. (1971). Systematic reinforcement. Academic performance of underachieving students. *Journal of Applied Behavior Analysis, 4*(4), 311–319. doi:10.1901/jaba.1971.4-311
- Chilcoat, G. (1989). Instructional behaviors for clearer presentations in the classroom. *Instructional Science, 18*(4), 289–314. doi:10.1007/BF00118015
- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität. Pädagogische Psychologie und Entwicklungspsychologie: Vol. 29*. Münster: Waxmann.
- Codding, R. S., & Smyth, C. A. (2008). Using performance feedback to decrease classroom transition time and examine collateral effects on academic engagement. *Journal of Educational and Psychological Consultation, 18*(4), 325–345. doi:10.1080/10474410802463312
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Copeland, W. D. (1987). Classroom management and student teachers' cognitive abilities: A relationship. *American Educational Research Journal, 24*(2), 219–236. doi:10.2307/1162892
- Doyle, W. (1984). How order is achieved in classrooms: An interim report. *Journal of Curriculum Studies, 16*(3), 259–277.
- Doyle, W. (2006). Ecological approaches to classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management. Research, practice and contemporary issues* (pp. 97–125). Mahwah, NJ: Lawrence Erlbaum Associates.
- Emmer, E. T., Evertson, C. M., & Worsham, M. E. (2003). *Classroom management for secondary teachers* (6th ed.). Boston: Allyn and Bacon.
- Emmer, E. T., & Gerwels, M. C. (2006). Classroom management in middle and high school classrooms. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management. Research, practice and contemporary issues* (pp. 407–437). Mahwah, NJ: Lawrence Erlbaum Associates.

- Enzmann, D., & Kleiber, D. (1989). *Helfer-Leiden: Stress und Burnout in psychosozialen Berufen*. Heidelberg: Asanger.
- Evertson, C. M., & Emmer, E. T. (1982). Effective management at the beginning of the year in junior high classes. *Journal of Educational Psychology*, 74(4), 485-498. doi:10.1037/0022-0663.74.4.485
- Evertson, C. M., & Harris, A. H. (1999). Support for managing learning-centered classrooms: The classroom organization and management program. In H. J. Freiberg & J. E. Brophy (Eds.), *Beyond behaviorism. Changing the classroom management paradigm* (pp. 59–74). Boston: Allyn and Bacon.
- Evertson, C. M., & Weinstein, C. S. (2006a). Classroom management as a field of inquiry. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management. Research, practice and contemporary issues* (pp. 3–15). Mahwah, NJ: Lawrence Erlbaum Associates.
- Evertson, C. M., & Weinstein, C. S. (Eds.).(2006b). *Handbook of classroom management. Research, practice and contemporary issues*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Freiberg, J. H., & Lapointe, J. M. (2006). Research-based programs for preventing and solving discipline problems. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management. Research, practice and contemporary issues* (pp. 735–786). Mahwah, NJ: Lawrence Erlbaum Associates.
- Freiberg, J. H., Stein, T. A., & Huang, S. (1995). Effects of a classroom management intervention on student achievement in inner-city elementary schools. *Educational Research and Evaluation*, 1(1), 36–66. doi:10.1080/1380361950010103
- Friedman, I. A. (2006). Classroom management and teacher stress and burnout. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management. Research, practice and contemporary issues* (pp. 925-945). Mahwah, NJ: Lawrence Erlbaum Associates.
- Fukkink, R. G., Trienekens, N., & Kramer, L. J. C. (2010). Video feedback in education and training: Putting learning in the picture. *Educational Psychology Review*, 23(1), 45-63. doi:10.1007/s10648-010-9144-5
- Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, 95(1), 148–162. doi:10.1037/0022-0663.95.1.148

- Gettinger, M., & Kohler, K. M. (2006). Process-outcome approaches to classroom management and effective teaching. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management. Research, practice and contemporary issues* (pp. 73-95). Mahwah, NJ: Lawrence Erlbaum Associates.
- Gottfredson, D. C., Gottfredson, G. D., & Hybl, L. G. (1993). Managing adolescent behaviour: A multiyear, multi-school study. *American Educational Research Journal*, 30(1), 179–215.
- Grossman, P. L. (2005). Research on pedagogical approaches in teacher education. In M. Cochran-Smith & K. Zeichner (Eds.), *Studying teacher education* (pp. 425-476). Mahwah, NJ: Lawrence Erlbaum.
- Hardin, C. J. (2008). *Effective classroom management: Models and strategies for today's classrooms* (2nd ed.). Upper Saddle River, NJ: Pearson/Merrill/Prentice Hall.
- Hattie, J. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge, Taylor & Francis Group
- Havers, N. (2001). Mit Disziplinschwierigkeiten umgehen lernen: Das Münchener Lehrertaining. *Grundschule*, 30(4), 33-35.
- Helmke, A. (1988). *Das Münchener Aufmerksamkeitsinventar (MAI): Manual für die Beobachtung des Aufmerksamkeitsverhaltens von Grundschulern während des Unterrichts*. München: Max-Planck-Institut für psychologische Forschung.
- Helmke, A. (1992). Das Münchener Aufmerksamkeitsinventar (MAI): Ein Instrument zur systematischen Verhaltensbeobachtung der Schüleraufmerksamkeit im Unterricht. *Diagnostica*, 38(2), 130–141.
- Howard, G. S., & Dailey, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology*, 64(2), 144-150. doi: 10.1037/0021-9010.64.2.144
- Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose–effect relationship in psychotherapy. *American Psychologist*, 41(2), 159–164. doi:10.1037/0003-066X.41.2.159
- Hylton, I. E. (2000). *Classroom management skills: Can video-modeling make a difference?* New York University: School of Education.
- Johnson, D. W., & Johnson, R. T. (2006). Conflict resolution, peer mediation, and peacemaking. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management. Research, practice and contemporary issues* (pp. 803-832). Mahwah, NJ: Lawrence Erlbaum Associates.

- Jones, V. (2006). How do teachers learn to be effective classroom managers? In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management. Research, practice and contemporary issues* (pp. 887-907). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kirkpatrick, D. L. (1998). *Evaluating training programs: The four levels* (2nd ed.). San Francisco, CA: Berrett-Koehler.
- Kounin, J. S. (1970). *Discipline and group management in classrooms*. New York, NY: Holt, Rinehart, & Winston.
- Kounin, J. S., & Gump, P. V. (1974). Signal systems of lesson settings and the task-related behavior of preschool children. *Journal of Educational Psychology*, 66(4), 554–562. doi 10.1037/h0036748
- Krammer, K., & Reusser, K. (2005). Unterrichtsvideos als Medium der Aus- und Weiterbildung von Lehrpersonen. *Beiträge zur Lehrerbildung*, 23(1), 35-50.
- Kubany, E. S., Sloggett, B. B., & Ogata, R. F. (1974). Training teachers in classroom management skills. *American Journal of Community Psychology*, 2(2), 139–149.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Eds.). (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Landrum, T. J., & Kauffman, J. M. (2006). Behavioral approaches to classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management. Research, practice and contemporary issues* (pp. 47-71). Mahwah, NJ: Lawrence Erlbaum Associates.
- Leflot, G., van Lier, P. A. C., Onghena, P., & Colpin, H. (2010). The role of teacher behavior management in the development of disruptive behaviors: An intervention study with the Good Behavior Game. *Journal of Abnormal Child Psychology*, 38(6), 869–882. doi:10.1007/s10802-010-9411-4
- Leinhardt, G., McCarthy Young, K., & Merriman, J. (1995). Integrating professional knowledge: The theory of practice and the practice of theory. *Learning and Instruction*, 5, 401-408. doi:10.1016/0959-4752(95)00025-9
- Leinhardt, G., Weidman, C., & Hammond, K. M. (1987). Introduction and integration of classroom routines by expert teachers. *Curriculum Inquiry*, 17(2), 135–176.
- Lipowski, F. (2010). Lernen im Beruf – Empirische Befunde zur Wirksamkeit von Lehrerfortbildung. In F. Müller, A. Eichenberger, M. Lüders, & J. Mayr (Eds.), *Lehre-*

- rinnen und Lehrer lernen – Konzepte und Befunde zur Lehrerfortbildung* (pp. 51-72). Münster: Waxmann.
- Little, S. G., & Akin-Little, A. (2008). Psychology's contributions to classroom management. *Psychology in the Schools, 45*(3), 227–234. doi:10.1002/pits.20293
- Manning, M. L., & Bucher, K. T. (2007). *Classroom management: Models, applications, and cases* (2nd ed.). Upper Saddle River, NJ: Pearson /Merrill Prentice Hall.
- Martin, N. K., & Sass, D. A. (2010). Construct validation of the Behavior and Instructional Management Scale. *Teaching and Teacher Education, 26*(5), 1–12. doi:10.1016/j.tate.2009.001
- Marzano, R. J. (2005). *A handbook for classroom management that works*. Alexandria, Va: Association for Supervision and Curriculum Development.
- McGarity, J. R., & Butts, D. P. (1984). The relationship among teacher classroom management behavior, student engagement and student achievement of middle and high school science students of varying aptitude. *Journal of Research in Science Teaching, 21*(1), 55–61. doi:10.1002/tea.3660210107
- Melnick, S. A., & Meister, D. G. (2008). A comparison of beginning and experienced teachers' concerns. *Educational Research Quarterly, 31*(3), 39-56.
- Merrett, F., & Wheldall, K. (1993). How do teachers learn to manage classroom behavior? A study of teachers' opinions about their initial training with special reference to classroom behavior management. *Educational Studies, 19*(1), 91-106. doi:10.1080/0305569930190106
- Munby, H., Russell, T., & Martin, A. K. (2001). Teachers' knowledge and how it develops. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 877-904). Washington, DC: American Educational Research Association.
- Nash, R. (2009). *The active classroom: Practical strategies for involving students in the learning process*. Thousand Oaks, CA: Corwin Press.
- Nie, Y., & Lau, S. (2009). Complementary roles of care and behavioral control in classroom management: The self-determination theory perspective. *Contemporary Educational Psychology, 34*(3), 185–194. doi:10.1016/j.cedpsych.2009.03.001
- Oevermann, U. (1996). Theoretische Skizze einer revidierten Theorie professionalisierten Handelns. In A. Combe & W. Helsper (Eds.), *Pädagogische Professionalität. Untersuchungen zum Typus pädagogischen Handelns* (pp. 70 – 182). Suhrkamp: Frankfurt/Main.

- Oliver, R. M., Wehby, J. H., & Reschly, D. J. (2011). Teacher classroom management practices: Effects on disruptive or aggressive student behavior. *Campbell Systematic Reviews*. Advance online publication. doi: 10.4073/csr.2011.4
- Ophardt, D., & Thiel, F. (2008). Klassenmanagement als Basisdimension der Unterrichtsqualität. In M. K. W. Schweer (Ed.), *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungsperspektiven und methodische Zugänge* (2nd ed., pp. 259–284). Wiesbaden: Verlag für Sozialwissenschaften.
- Ophardt, D., Thiel, F. & Piwowar, V. (2009). *KODEK - Kompetenzen des Klassenmanagements. Reader zur Fortbildung mit Beispielen und Reflexionsaufgaben*. Unveröffentlichtes Manuskript, Fachbereich Erziehungswissenschaft und Psychologie, Freie Universität Berlin, Berlin.
- Pianta, R. C. (2006). Classroom management and relationships between children and teachers: Implications for research and practice. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management. Research, practice and contemporary issues* (pp. 685-710). Mahwah, NJ: Lawrence Erlbaum Associates.
- Pigge, F. L., & Marso, R. N. (1997). A seven-year longitudinal multi-factor assessment of teaching concerns development through preparation and early years of teaching. *Teaching and Teacher Education, 13*(2), 225-235. doi:10.1016/S0742-051X(96)00014-5
- Piwowar, V., Ophardt, D. & Thiel, F. (2011, September). *A multidimensional rating system assessing competencies in classroom management*. Paper presented at the European Conference on Educational Research, Berlin, Germany.
- Pomeroy, E. (1999). The teacher-student relationship in secondary school: Insights from excluded students. *British Journal of Sociology of Education, 20*(4), 465-482. doi:10.1080/01425699995218
- Richardson, J. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review, 6*(2), 135–147. doi:10.1016/j.edurev.2010.12.001
- Rogers, D. (2009). The working alliance in teaching and learning: Theoretical clarity and research implications. *International Journal for the Scholarship of Teaching and Learning, 3*(2). Retrieved from http://academics.georgiasouthern.edu/ijstl/v3n2/essays_about_sotl/PDFs/Essay_Rogers.pdf
- Santagata, R., & Guarino, J. (2011). Using video to teach future teachers to learn from teaching. *ZDM Mathematics Education, 43*(1), 133–145. doi: 10.1007/s11858-010-0292-3

- Sariscsany, M. J., & Pettigrew, F. (1997). Effectiveness of interactive video instruction on teacher's classroom management declarative knowledge. *Journal of Teaching in Physical Education, 16*(2), 229-240.
- Schmitz, G. S., & Schwarzer, R. (2000). Selbstwirksamkeitserwartung von Lehrern: Längsschnittbefunde mit einem neuen Instrument. *Zeitschrift für Pädagogische Psychologie, 14*(1), 12-25. doi:10.1024//1010-0652.14.1.12
- Schoen, D. A.(1983). *The reflective practitioner. How professionals think in action*. London: Temple Smith.
- Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4), 454–499. doi:10.3102/0034654307310317
- Shernoff, E. S., & Kratochwill, T. R. (2007). Transporting an evidence-based classroom management program for preschoolers with disruptive behavior problems to a school: An analysis of implementation, outcomes, and contextual variables. *School Psychology Quarterly, 22*(3), 449–472. doi:10.1037/1045-3830.22.3.449
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14. doi:10.3102/0013189X015002004
- Simons, P. R. J., & Ruijters, M. C. P. (2004). Learning professionals: Towards an integrated model. In H. P. A. Boshuizen, R. Bromme, & H. Gruber (Eds.), *Professional learning: Gaps and transitions on the way from novice to expert* (pp. 207-229). Dordrecht: Kluwer.
- Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education and Treatment of Children, 31*(3), 351–380. doi: 10.1353/etc.0.0007
- Slavin, R. E., Hurley, E. A., & Chamberlain, A. (2003). Cooperative learning and achievement: Theory and research. In W. M. Reynolds & G. E. Miller (Eds.), *Handbook of psychology: Educational psychology*. New York, NY: Wiley.
- Smith, L. R. (1985). A low-inference indicator of lesson organization. *The Journal of classroom interaction, 21*(1), 25-30.
- Stallings, J. (1980). Allocated academic learning time revisited, or beyond time on task. *Educational Researcher, 9*(11), 11–16. doi:10.3102/0013189X009011011
- Stough, L. M. (2006). The place of classroom management and standards in teacher education. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management*.

- Research, practice and contemporary issues* (pp. 909-923). Mahwah, NJ: Lawrence Erlbaum Associates.
- van Driel, J. H., & Berry, A. (2012). Teacher professional development focusing on pedagogical content knowledge. *Educational Researcher*, 41(1), 26–28. doi:10.3102/0013189X11431010
- van Tartwijk, J., den Broek, P., Veldman, I., & Wubbels, T. (2009). Teachers' practical knowledge about classroom management in multicultural classrooms. *Teaching and Teacher Education*, 25(3), 453–460. doi:10.1016/j.tate.2008.09.005
- Walker, H. M., Ramsey, E., & Gresham, F.M. (2004). *Antisocial behaviour in school: Strategies and best practices* (2nd ed.). Pacific Grove, CA: Brooks/ Cole.
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63(3), 249–294. doi:10.3102/00346543063003249
- Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago, IL: Rand McNally.
- Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Ed.), *Leistungsmessung in Schulen* (pp.17-32). Weinheim: Beltz.
- Weinstein, C. S. (1999). Reflections on best practices and promising programs: Beyond assertive classroom discipline. In H. J. Freiberg & J. E. Brophy (Eds.), *Beyond behaviorism. Changing the classroom management paradigm* (pp. 147–163). Boston: Allyn and Bacon.
- Weinstein, C. S. (2007). *Middle and secondary classroom management: Lessons from research and practice*. Boston: Mc Graw Hill.
- Witt, J. C., van der Heyden, A. M., & Gilbertson, D. M. (2004). Instruction and classroom management: Prevention and intervention research. In R. B. Rutherford (Ed.), *Handbook of research in emotional and behavioral disorders* (pp. 426–445). New York, NY: Guilford Press.
- Woolfolk, A. E., & Brooks, D. M. (1985). Beyond words: The influence of teachers' nonverbal behaviors on students' perceptions and performances. *Elementary School Journal*, 85, 513-528.

3.6. Acknowledgements

The research project is funded by the BMBF (Bundesministerium für Bildung und Forschung; Federal Ministry of Education and Research of Germany). We would like to thank the students, teachers, and the administrative staff involved for this research. We would also like to thank Benjamin Beckett and Franziska Pfitzner-Eden for editing our manuscript. Finally, we would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the manuscript.

4

Teilstudie 3: Evaluating Response Shift in Training Evaluation. Comparing the Retrospective Pre-Test with a Structural Equation Modeling Approach in a Classroom Management Training Pro- gram.

Notice: This is the authors' version of a work that has been submitted for publication as:

Piwowar, V., & Thiel, F. (2013). Evaluating Response Shift in Training Evaluation. Comparing the Retrospective Pre-Test with a Structural Equation Modeling Approach in a Classroom Management Training Program. *Manuscript submitted for publication.*

Abstract

Response shift (RS) is of particular concern in pre-post designs as it threatens their internal validity. The two most common approaches to deal with RS are the retrospective pre-test and a structural equation modeling technique. However, their results in identifying RS were rarely studied simultaneously. We evaluated RS by implementing both procedures and compared the occurrence of types of RS in two treatment conditions (short-term vs. long-term classroom management training program). Participants were 146 pre-service teachers who completed pre- (before training), post- and retrospective self-ratings (after training) on classroom management knowledge. Results indicated that the two approaches do not yield the same results. Structural equation modeling identified more and also group specific RS, as opposed to the findings of the retrospective pre-test, which found less and only little evidence for group specific RS. Further research is needed to study the usability and validity of the respective approaches in case of RS.

Keywords: response shift – retrospective pre-test – training evaluation – pre-post design – classroom management

4. Evaluating Response Shift in Training Evaluation. Comparing the Retrospective Pre-Test with a Structural Equation Modeling Approach in a Classroom Management Training Program.

4.1. Introduction

A principal issue in evaluating student teachers' learning is using valid methods to measure improvement (Moore & Tananis, 2009; Pratt, McGuigan, & Katzev, 2000). However, a direct and proximal assessment is not always feasible (Desimone, 2009). Instead, self-reports are often employed, which provide the basis for an indirect measurement of change via pre-post designs (Lucas & Baird, 2006).

Pre-post designs relying on self-report measures are well established in the area of training evaluation. However, there are two main drawbacks that jeopardize their internal validity: pre-testing effects (e.g., priming effects—pre-testing as a treatment in its own) and response-shift bias (Aiken & West, 1990; Lam, 2003). Response shift (*RS*) “refers to a change in the meaning of one's self-evaluation of a target construct” (Sprangers & Schwartz, 1999, p. 1508). Sprangers and Schwartz (1999) identified three types of RS: “(1) recalibration, that is, a change in the respondent's internal standards of measurement; (2) reprioritization, that is, a change in the respondent's values (i.e., reevaluation of the importance of [indicators] constituting the target construct); or (3) reconceptualization, that is, a redefinition of the target construct” (Sprangers & Schwartz, 1999, p. 1508, as cited in Oort, Visser, & Sprangers, 2009, p. 1127).

RS is of particular concern in group-based educational training programs, where participants gain new insights into a target construct, discuss it and also interact with each other (Gibbons, 1999; Nolte, Elsworth, Sinclair, & Osborne, 2009). Accordingly, a reconceptualization RS is most likely to occur when the target construct is abstract or unfamiliar pre-intervention or when the training program explicitly intends to elucidate concepts (e.g., the understanding of effective rules in a classroom). Consequently, a shift will probably not occur for untrained controls (Aiken & West, 1990) or may vary in different treatments (e.g., subjects who discuss the usability of rules may experience different RS than subjects who try to implement stipulated rules in a role-play). A reprioritization RS points to changes in the connotation of single indicators (e.g., the connotation of “clear” behavior expectations to students) - the longitudinal in-homogeneity of indicators that constitute a latent construct. In particular, formative constructs, such as health-related quality of life (*HR-QOL*), are prone to a

changed meaning of indicators (Donaldson, 2005). A recalibration RS indicates a change in the latent metric of the construct. The frame of reference and meaning of options for answers have changed over time, for example due to changed awareness or social comparison, since participants may interpret items and their related responses differently post-intervention (Gibbons, 1999; Hill & Betz, 2005). Hence, improvement after attending a training program may reflect other than a “true change”, but biased measurement (Oort et al., 2009). Conversely, RS may mask underlying training-effects (false-negative results), for example when individuals overestimate their initial status (Howard, 1980; Howard & Dailey, 1979).

Various efforts have been undertaken in order to address the issue of RS bias (for an overview, cf. Schwartz & Sprangers, 1999; Barclay-Goddard, Epstein, & Mayo, 2009). The present study implemented two procedures to examine RS: First, the then-test procedure (pre-test data derived in retrospect—a design approach) and second, the evaluation of level of measurement invariance (a statistical approach).

Within the field of educational research, the design approach is the most commonly used method for measuring RS (Hill & Betz, 2005; Pratt et al., 2000; Sibthorp, Paisley, Gookin, & Ward, 2007). In addition to—or sometimes instead of—the conventional pre-test, subjects rate their initial status using a retrospective pre-test (then-test) at the end of the training program. As then- and post-test are completed at the same occasion, it is assumed, that subjects refer to the same internal metric and understanding of the target construct. Differences between the initial and retrospective ratings are interpreted as RS bias (Hill & Betz, 2005).

The then-test was originally developed to measure recalibration RS, assuming that reprioritization and reconceptualization do not occur. If this assumption is violated, then-test results are difficult to interpret due to a changed understanding of the target concept between assessments (Visser, Oort, & Sprangers, 2005; Schwartz, Sprangers, Carey, & Reed, 2004). Hence, in recreation and HR-QOL research, a statistical approach evolved which investigates RS by means of structural equation modeling (*SEM*, Schwartz & Sprangers, 1999). In particular, the measurement of invariance (*MI*, Meredith, 1993) using SEM serves to evaluate recalibration, reprioritization, and reconceptualization in longitudinal data in terms of different factor patterns, different factor loadings and different intercepts for repeated measures, respectively (Ahmed, Mayo, Corbiere, Wood-Dauphinee, Hanley, & Cohen, 2005; Gandhi, Ried, Huang, Kimberlin, & Kauf, 2013; Oort, 2005).

The aim of the current study was to address three different issues. First, even though both the design and the statistical approaches can be used to successfully detect RS, they have rarely been combined, and only refer to HR-QOL research, mostly using the SF-36 health

survey (Schwartz, Bode, Repucci, Becker, Sprangers, & Fayers, 2006). Second, MI was predominantly applied for pre-post comparison, whereas then-post invariance remained untested (Nolte et al., 2009, Schwartz & Sprangers, 2010). Third, treatment-specific impacts on the various types of RS need to be studied, as they are of particular interest to training evaluation that relies on control-group designs (Ahmed, Mayo, Corbiere, et al., 2005; Aiken & West, 1990).

4.1.1. Response shift and the then-test procedure

The most common way to face RS is to implement a then-test, as a “quasi-indirect” measurement of change (e.g., Cantrell, 2003; Drennan & Hyde, 2008; Holden, Barker, Rosenberg, & Onghena, 2008; Moore & Tananis, 2009; Pratt et al., 2000; Sibthorp et al., 2007). Hoogstraten (1982) introduced the then-test procedure to the educational training context. He found that both a task performance test and the then-post comparison showed evidence for subjects’ improvement, whereas the traditional pre-post comparison did not. Also, pre-then comparisons revealed that subjects initially over-estimated their performance, indicating a RS. Numerous educational and medical studies replicated these early findings of over-estimation (e.g., Cantrell, 2003; Drennan & Hyde, 2008; Hill & Betz, 2005; Holden et al., 2008; Hoogstraten, 1985; Moore & Tananis, 2009; Pratt et al., 2000), founding larger effects for the then-post as opposed to the pre-post comparison (e.g., Cantrell, 2003; Pratt et al., 2000, Sibthorp et al., 2007). Furthermore, then-tests tend to be better predictors for objective measures and other external validation criteria than are traditional pre-tests (e.g. interview data or self-rated change; Ahmed, Mayo, Corbiere, Wood-Dauphinee, Hanley, & Cohen, 2005; Cantrell, 2003; Hoogstraten, 1985; Pratt et al., 2000; Sibthorp et al., 2007), and they are less correlated with social desirability measures (Aiken & West, 1990; Hill & Betz, 2005).

However, it is widely discussed that then-tests may increase bias rather than reduce it (Hill & Betz, 2005; Schwartz & Sprangers, 2010). Implicit theories of change might operate, when pre- and then-test scores differ from each other: Subjects may underscore initial status in their retrospective ratings in order to reduce cognitive dissonance (effort justification bias; Aiken & West, 1990, Sibthorp et al., 2007). Also, then-tests might be confounded with recall bias, since subjects differ in their ability to remember prior status (Schwartz et al., 2004). Accordingly, a difference in pre-then scores may indicate mere invalidity (e.g., poor memory) of then-scores instead of a RS bias. Albeit, previous findings relating to recall bias and then-test are inconsistent and strongly depend on the studied time frame and the outcome measures (Henry, Moffitt, Caspi, Langley & Silva, 1994; Schwartz et al., 2004, Schwartz & Sprangers,

2010). In a study with university students that participated in outdoor leadership courses, subjects changed their retrospective self-rating even though they were able to remember their pre-scoring (Sibthorp et al., 2007). Consequently, it remains unclear, whether pre-then differences indicate over-estimation at pre-testing, deliberate adjustment of then-scores (e.g. in terms of effort justification), or simply recall bias at then-testing (Norman, 2003).

4.1.2. Response shift and the measurement of invariance

All statistical methods that have been proposed for assessing RS belong to the SEM framework (Schwartz & Sprangers, 1999). Various procedures evolved to adequately use techniques of SEM in order to detect recalibration, reprioritization, and reconceptualization RS (e.g., Ahmed, Mayo, Corbiere, et al., 2005; Gandhi et al., 2013; Oort, 2005; Schmitt, 1982). In particular, the longitudinal measurement of invariance (Meredith, 1993), which merged SEM and item response theory, gained wide acceptance for RS-evaluation, as it analyzes testable assumptions and explicitly allows for the statistical comparison of separate components of the measurement model over time (Visser et al., 2005; Schwartz et al., 2004).

Oort's (2005) statistical approach became most prevalent—it allows for the testing of all types of RS and maximizes statistical power (Gandhi et al., 2013, see also Table 1). Nolte et al. (2009) extended Oort's approach and assigned the terminology of different types of RS to corresponding levels of MI, which had not been made explicit in the original RS model. Subsequently, we are further adding the assignment of strict invariance to non-uniform recalibration RS. Identification of RS and MI testing are combined as follows:

- 1) *Reconceptualization* corresponds to *configural invariance (CI)*: The overall number of common factors and pattern of the common factor loadings are invariant across occasions (CI). Differences between patterns indicate that the meaning of the construct has changed over time (reconceptualization).
- 2) *Reprioritization* corresponds to *weak (also metric) invariance (WI)*: In addition to CI, factor loadings are invariant across occasions (WI). Differences between item factor loadings indicate that a variable has become more indicative or less indicative of the latent construct (reprioritization).

Oort (2005) also introduced the differentiation between *uniform recalibration* and *non-uniform recalibration*:

- 3) *Uniform Recalibration* corresponds to *strong invariance (SI)*: In addition to WI, item intercepts are invariant across occasions (SI). Differences between item intercepts indicate a recalibration in the mean structure (uniform recalibration).

4) *Non-uniform Recalibration* corresponds to *strict invariance (STI)*: In addition to SI, items' residual variances are invariant across occasions (STI). Differences between residual variances indicate a recalibration to the covariance structure (item uniqueness) of the observed variables that cannot be attributed to a change in the common factor variances (non-uniform recalibration).

At least strong MI needs to be met for an unbiased comparison of factor means over time (Brown, 2006; Meredith, 1993). Accordingly, only non-uniform recalibration RS (with the highest level of MI being strong invariance) still allows for a valid pre-post comparison.

Studies that investigated RS using SEM, predominantly refer to the SF-36. Depending on the applied procedures, the subjects and state of health studied, SEM identified no (Ahmed, Mayo, Corbiere et al., 2005; Nolte et al., 2009) to all kinds of RS (Ahmed, Schwartz, Ring, & Sprangers, 2009; Gandhi et al., 2013, Oort, Visser, & Sprangers, 2005). Corresponding to the findings of design approaches, HR-QOL-research that modeled RS statistically, also found increased estimates of true change in HR-QOL when accounting for RS (Ahmed et al., 2009; Oort et al., 2005, Schwartz et al., 2004).

In spite of the recognized strengths of SEM techniques, there are certain drawbacks that may compromise its feasibility, reliability and validity, for example the requirement of relatively large sample sizes, or the question of how to handle reconceptualization when no MI is achieved (Ahmed, Mayo, Corbiere, et al., 2005; Schwartz & Sprangers, 1999). Donaldson (2005) critically postulated that SEM techniques are not appropriate for detecting RS, as they operate at a construct level, whereas RS operates at the individual level.

As opposed to the commonly applied then-test, validity studies examining the MI approach to determining RS are quite rare. Studies, that investigated MI and the then-test simultaneously, reported inconsistent findings. Whereas Visser et al. (2005) found convergence in RS detected by both approaches for subjects with cancer, Ahmed, Mayo, Wood-Dauphinee, Hanley, and Cohen. (2005) found scale recalibration in then-post comparison and individual interviews, but no RS conducting MI for post-stroke subjects. A single study explored RS with SEM for stroke people vs. a control group without stroke and found that either group experienced RS, even though an occurrence of RS in the stroke group was expected (Ahmed, Mayo, Corbiere et al., 2005).

For two main reasons, simply transferring the presented findings to the context of training evaluation is insufficient. First, the construct of HR-QOL is rather formative than recursive (Donaldson, 2005). It is measured by cause indicators—where the indicators constitute the construct, as opposed to reflexive indicators (also effect indicators)—where the indicators are

equivalent manifestations of the latent construct (Bollen & Lennox, 1991; Vandenberg & Lance, 2000). Consequently, a reprioritization has a different meaning for both types of constructs: Formative indicator reprioritization alters the meaning of the latent construct, as the sum of indicators defines the construct; recursive indicator reprioritization alters the effect of the latent construct on a reprioritized indicator. Second, outcome variables and treatments in educational research differ markedly from those in clinical trials. The former targets competence, knowledge, or attitudes, whereas the latter targets the objective recreation and subjective appraisal of physical and mental health status. Thus, research findings of medical interventions studies determining RS can hardly be assigned to the educational context.

4.1.3. Research questions

Pre-post designs are prone to RS bias, where the frame of reference, priorities of facets, or the target concept is not intra- and/or inter-individually stable across time. Even though the phenomenon and proper handling of RS was studied intensively, none of the developed approaches raised to be the gold standard. Still, several research gaps remain:

First, although MI is common to the field of education and psychology, research into RS using MI was developed in the area of HR-QOL and stayed there. As opposed to the then-test procedure, MI holds testable assumptions for analyzing the various types of RS bias (Schwartz et al., 2004; Visser et al., 2005). Consequently, we deployed a MI approach (pre-post comparison) within the field of educational training evaluation and compared its conclusions to the then-test procedure. As participants of the training are supposed to gain an in-depth understanding of the target construct (that is, classroom management), we hypothesize, that (H1) the training program affects MI of pre-post comparison in terms of reconceptualization, reprioritization and recalibration RS. Schwartz et al. (2006) determined moderators (i.e., approach to study RS) for RS in a meta-analysis and found no interaction of identified RS and the applied approach (either then-test or MI). Therefore, we further hypothesize, that (H2) the occurrence of RS is equivalent for MI and the then-test procedure.

Second, the implicit assumption that subjects refer to the same set of internal standards for then- and post-test (that is, then- and post-test are supposed to be invariant across occasions), has yet to be tested and confirmed (Schwartz & Sprangers, 2010). Nolte et al. (2009) evaluated a chronic disease self-management course, where then-post comparison revealed even less invariant items than pre-post comparison. The authors challenged the psychometric properties of the then-test as opposed to the pre- and the post-test. However, Nolte et al. presented each then- and post item simultaneously, which may have reinforced, for example ef-

fort justification bias. Thus, we hypothesize that when completed consecutively, (H3) then-post comparisons are measurement-invariant (that is, at least strong invariance) and reveal higher levels of MI than pre-post comparisons.

Third, the then-test is often applied within single-group designs (Pratt et al., 2000). Recognizing that a treatment may affect RS (Oort et al., 2009) occurrence and type of RS are hypothesized to be distinct in different treatment conditions depending on treatment intensity (in terms of length and teaching-methods applied). Accordingly, H1 is specified as follows: We hypothesize, that individuals exposed to a more interactive training (intervention group) would experience a greater amount and a greater degree of RS (in particular: reconceptualization RS) as compared to participants of less interactive training (control group) as tested via MI as well as the then-test procedure. We expect no group differences with regard to hypotheses H2 and H3.

4.2. Method

4.2.1. Study design and setting

We applied a quasi-experimental design, where eight study groups were randomly assigned to either the control group (*CG*) or the intervention group (*IG*). The training program was implemented within a university course that aimed at improving teachers' knowledge on classroom management for master degree student teachers in their first year. Participants would have had some level of elementary knowledge on classroom management, however as student teachers obtained their bachelor degree at different universities, and classroom management is not a fixed part in teacher training curricula, that level might have varied significantly between participants.

Participants provided pre-test data at the start of the course (*T1*). Post-measures were administered consecutively (post-test, *T2*, and then-test, *T1r*). In total, 176 student teachers participated in the course. Evaluation was optional. Only those subjects that completed all three questionnaires (pre-, post-, and then-test) were considered for the present analyses. Hence, the final sample consisted of $N = 146$ student teachers (83% of the initial sample, $N_{CG} = 81$, $N_{IG} = 65$). Groups were comparable with regard to their age ($M = 25$ years, $SD = 3.35$), their teaching experience ($Md = 12$ hours of own teaching), their motivations for vocational choice (Thiel, & Blüthman, 2009), and teacher self-efficacy (Schwarzer & Hallum, 2008). *IG* student teachers were more satisfied with their studies ($M_{IG} = 6.67$, $SD = 1.20$, and $M_{CG} = 6.40$, $SD = 1.35$, on an 8-point scale, 8 indicating high satisfaction, Thiel, & Blüthmann, 2009) at pre-testing and were less often female ($IG = 55\%$, $CG = 79\%$).

4.2.2. Treatment conditions

The training program consisted of three blocks: (1) basics of classroom research (three sessions à 90 minutes), (2) basics of classroom management (two sessions), and (3) practice of classroom management (three sessions). Only block (2) and (3) focused on improving classroom management (CM). In block (2), students got introduced to the following aspects of CM strategies: establishing and maintaining expected behavior, directing learning activities, and solving conflicts (i.e., dealing with disruptive student behavior, using rules, providing structured and clear tasks, maintaining activity flow via procedures and time management, building a working alliance, and handling conflicts among students). This was realized via lecture along with a video presentation and self-reflection exercises. In block (3), student teachers analyzed staged video cases focusing on disruptive student behaviors in small groups in order to actively generate and discuss alternative strategies (two sessions), and tried out new CM strategies in simulations using the microteaching method (one session, Piwovar, Ophardt, & Thiel, 2013; Grossman, 2005). Within these three sessions, the main focus was on meeting disruptive student behavior (dealing with disruptions and referring to rules), while keeping the learning process running as well as sheltering the working alliance between students and the teacher. Both video analyses and micro-teaching required active participation of each student and included intensive collaborative discussion with peers.

The control group completed the post- and then-ratings after attending block (1) and (2) (eight sessions) without participating in the interactive sessions in block (3). The intervention group completed the post- and then-ratings after completing the whole training (eleven sessions).

4.2.3. Material

Knowledge on classroom management was assessed using the self-rating *SEWIK* (*Selbsteingeschätztes Wissen im Klassenmanagement*, self-rated knowledge on classroom management, Piwovar, et al., 2013; Thiel, Ophardt, & Piwovar, 2013). The SEWIK constitutes a multidimensional assessment of CM knowledge, containing the following dimensions: (a) *Clarity of Program of Action (CL)*, (b) *Conflicts among Students (CO)*, (c) *Dealing with Disruption (DD)*, (d) *Group Mobilization (GM)*, (e) *Procedures (PR)*, (f) *Rules (RU)*, (g) *Time Management (TM)*, and (h) *Working Alliance (WA)*. Student teachers were instructed to refer to their theoretical knowledge only, disregarding if they would succeed in real a classroom situation. Student teachers rated statements such as “*I am familiar with various strategies for involving as many students as possible actively in the classroom.*” (Group Mobilization) and

“*I am familiar with strategies for dealing with severe problem behavior.*” (Dealing with Disruption) on an 8-point scale ranging from 1 (*strongly disagree*) to 8 (*strongly agree*). Each scale consists of three items (24 items in total) and scale scores were calculated by averaging. Cronbach’s Alpha was good to excellent ($Md_{T1} = .87$, $.78 < \alpha_{T1} < .95$; $Md_{T2} = .87$, $.79 < \alpha_{T2} < .95$; and $Md_{T1r} = .94$, $.85 < \alpha_{T1r} < .96$). Previous studies indicated SEWIK’s good predictive validity, as it showed positive relationships between knowledge and teaching experience: Student teachers had the lowest, pre-service teachers had moderate, and in-service teachers had the highest knowledge on dimensions of classroom management (Thiel et al., 2013).

4.2.4. Data analysis

SEWIK-Items for T1, T2, and T1r had mean missings of 0.26%, 0.03%, and 0.15%, respectively. As missings were rare, they were replaced by scale scores. In SEM-analysis, the maximum-likelihood estimate was used to estimate missing values (Muthén & Muthén, 1998-2010).

In order to evaluate RS based on the then-test framework, t-tests of individual difference scores ($\Delta T1-T1r$) were performed separately for both the CG and the IG ($H_0 =$ mean difference scores are zero). Difference scores provide an estimate of the direction and the magnitude of the RS effect. Positive difference scores indicate over-estimation, and negative difference scores indicate under-estimation (Schwartz & Sprangers, 2010). Effect size values of $d = 0.2$, 0.5 , and 0.8 are considered ‘small’, ‘medium’, and ‘large’, respectively (Cohen, 1988). They were calculated by dividing means by the standard deviation.

In order to evaluate RS based on the MI framework, we compared pre- and post-test scores (T1 vs. T2) as well as then- and post-test scores (T1r vs. T2) by following Oort’s (2005) three-step within-group procedure (CG and IG separately, cf. Nolte et al, 2009)¹:

(1) Establishing a measurement model (configural invariance of occasions). The measurement model serves as a baseline model with identical factor patterns across measurement occasions, and no imposed constraints. This step confirms, if the hypothesized measurement model fits the data. If the fit is unsatisfactory, the analyses end here, concluding that a reconceptualization of the construct has occurred. If configural invariance holds, we proceed to step two.

¹ Oort (2005) specified a fourth step to further identify true change and calculate effect size. As this is not the focus of the present paper, we only run through steps one to three.

(2) Overall test of RS (strict invariance of occasions). In step two, all model parameters constraints across occasions in sense of strict MI were imposed to test for overall RS. In case of unsatisfactory fit, at least one type of RS has occurred (i.e., one model parameter is not invariant) and we proceed to step three. If the fit was satisfactory and the fully constrained model was not significantly worse than the baseline model, we can stop the analysis without occurrence of RS indicating strict invariance.

(3) Response shift detection (test of level of measurement invariance). The third step started with Model 2 and consisted of testing for specific types of RS: Parameter constraints for factor loadings (reprioritization), then intercepts (uniform recalibration), and then residuals (non-uniform recalibration) are released in the hierarchy of measurement invariance (Gandhi et al., 2103; Nolte et al., 2009). The evaluation of fit indices, modification indices, and standardized residuals, guided this step (Muthén & Muthén, 1998-2010). Each modification was then compared to the baseline model using the χ^2 difference testing (Brown, 2006) in order to determine the level of invariance (i.e., the final model).

Table 4.1: Three Step Procedure for Response Shift Detection, Invariance Hypothesis, and Attribution to Level of Measurement Invariance and Type of Response Shift

Step (associated model)	Invariance hypothesis	Invariance constraints	Level of MI	Interpretation in case of non-invariance ^a
1. Establishing a measurement model (Measurement model)	$\text{Patt}(\Lambda_1) = \text{Patt}(\Lambda_2)$	Factor patterns	Configural invariance	Reconceptualization
2. Overall test of response shift (No response shift model)	$\text{Diag}(\Omega_{11}) = \text{Diag}(\Omega_{22})$	Residual variances	Strict invariance	Recalibration (non-uniform)
3. Response shift detection (Response shift model)	$\Lambda_1 = \Lambda_2$ $\tau_1 = \tau_2$	Factor loadings Intercepts	Weak invariance Strong invariance	Reprioritization Recalibration (uniform)

Note. Adapted from Oort (2005, p. 593) and Nolte et al. (2009, p. 1176).

^a The interpretation in terms of RS only applies to pre- vs. post-test.

This procedure is summarized in Table 1. Whereas the statistical procedure applied to both pre-post and then-post comparison, the assignment of RS to the level of MI only applied to the former. We evaluated if assumptions for the SEM analysis (e.g. multivariate normality, multicollinearity, outliers) were met. Covariances between the errors of scales were allowed, because it improved the longitudinal model fit (Sörbom, 1975). Model evaluation was based on goodness-of fit indices (Brown, 2006), that is, χ^2 ($p > .05$ indicates model fit), *RMSEA*

(<.08 reasonable fit, <.05 close fit), *SRMR* (<.10 reasonable fit, <.05 close fit, Vandenberg & Lance, 2000) as well as *CFI*, and *TFI* (> .95 close fit, > .90 reasonable fit). χ^2 -difference testing was used for model comparison, as models were nested. The sample size was too small to include the 24 SEWIK-items in a single model, thus analyses were conducted at the subscale level (Tanaka, 1987).

SPSS19 software was used to carry out t-tests; the *Mplus* software (Version 6.0) was used for SEM analyses. For all statistical analyses, the Alpha-level was set at $p < .05$.

4.3. Results

T1, T1r, and T2 means, standard deviations and difference scores, as well as the final level of measurement invariance for each dimension, are shown in Table 2. In both pre- and then-test self-ratings students mean initial knowledge on CM was moderate, with the highest knowledge on Clarity ($M_{CG,T1} = 5.38$, $M_{CG,T1r} = 5.35$; $M_{IG,T1} = 5.02$, $M_{IG,T1r} = 5.03$) and the least knowledge on Dealing with Disruptions ($M_{CG,T1} = 3.87$, $M_{CG,T1r} = 4.23$; $M_{IG,T1} = 4.05$, $M_{IG,T1r} = 3.48$). Mean post-test scores indicated higher knowledge on all dimensions as compared to the traditional as well as the retrospective pre-test scores in both groups: After the training, students reported the least knowledge on Conflicts ($M_{CG,T2} = 4.86$, $M_{IG,T2} = 5.18$) and the highest knowledge on Clarity in the CG ($M_{CG,T2} = 5.77$), and Dealing with Disruptions in the IG ($M_{IG,T2} = 6.12$), respectively.

Mean pre-then difference scores ($\Delta T1-T1r$) were rather small in both groups. Within group t-tests indicate initial under-estimation in both groups with regard to the Working Alliance ($d_{CG} = -.27$, $d_{IG} = -.28$), Dealing with Disruptions in the CG ($d_{CG} = -.26$), and Conflicts in the CG ($d_{CG} = -.21$). Over-estimation was only present with regard to Dealing with Disruptions ($d_{IG} = .33$) and Group Mobilization in the IG ($d_{IG} = .20$). The effects were small and non-significant for Conflicts and Group Mobilization in both groups.

The results of the three-step SEM procedure are shown in Table 3. Comparing pre- and post-test identified a RS on every dimension in both treatment groups (see Tables 2 and 3). In the CG, non-uniform recalibration RS with regard to Clarity, Dealing with Disruptions, Group Mobilization, and Rules occurred (highest level of MI: strong invariance); further, reprioritization RS occurred with regard to Conflicts and Time Management (highest level of MI: configural invariance); and redefinition RS occurred on the two remaining dimensions Procedures and Working Alliance (no level of MI achieved). In the IG, non-uniform recalibration RS occurred with regard to Conflicts and Group Mobilization, and redefinition RS occurred on all six remaining dimensions. Thus, uniform recalibration RS bias was not present, in either the

CG or the IG (highest level of MI: weak invariance). Further, level of MI was higher in the CG than in the IG in four out of eight dimensions and the same in three out of eight dimensions. Only one dimension (conflicts) had a higher degree of measurement invariance in the IG. Model misfit was predominantly indicated by the RMSEA index.

Table 4.2: Means (*M*), Standard Deviations (*SD*) and Response Shift for T1, T1r, and T2

Scale		T1		T1r		T2		$\Delta T1-T1r$	Level of MI	
		<i>M</i>	(<i>SD</i>)	<i>M</i>	(<i>SD</i>)	<i>M</i>	(<i>SD</i>)	<i>d</i>	T1, T2	T1r, T2
CL	CG	5.38	(1.02)	5.35	(1.34)	5.77	(1.06)	.03	3	3
	IG	5.02	(1.49)	5.03	(1.44)	5.83	(1.17)	-.01	0	3
CO	CG	4.20	(1.33)	4.49	(1.72)	4.86	(1.40)	-.21	1	4
	IG	4.31	(1.84)	4.23	(1.68)	5.18	(1.41)	.05	3	0
DD	CG	3.87	(1.38)	4.23	(1.60)	4.89	(1.27)	-.26 *	3	2
	IG	4.05	(1.65)	3.48	(1.69)	6.12	(1.17)	.33 *	0	2
GM	CG	5.24	(1.26)	5.19	(1.31)	5.63	(1.15)	.03	3	4
	IG	4.83	(1.63)	4.54	(1.53)	5.71	(1.22)	.20	3	4
PR	CG	4.96	(1.21)	4.89	(1.49)	5.56	(1.04)	.05	0	3
	IG	4.38	(1.72)	4.14	(1.81)	5.49	(1.36)	.16	0	4
RU	CG	4.73	(1.49)	4.72	(1.66)	5.52	(1.29)	.01	3	0
	IG	4.35	(1.61)	4.12	(1.65)	5.92	(1.12)	.14	0	0
TM	CG	5.08	(1.45)	5.07	(1.63)	5.42	(1.26)	.01	1	0
	IG	4.57	(1.86)	4.82	(1.90)	5.33	(1.62)	-.16	0	3
WA	CG	4.37	(1.21)	4.76	(1.50)	5.18	(1.17)	-.27 *	0	4
	IG	4.24	(1.69)	4.65	(1.68)	5.88	(1.20)	-.28 *	0	4

Note. Means are based on 8-point scales, 1 – low declarative knowledge; 8 – high declarative knowledge; CG – control group, IG – intervention group; T1 – pre-test, T1r – then-test, T2 – post-test, $\Delta T1-T1r$ – Pre-then difference scores; Level of MI: 0 – no invariance, 1 – configural invariance, 2 – metric invariance, 3 – strong invariance, 4 – strict invariance; CL – Clarity of Program of Action, CO – Conflicts among Students, DD – Dealing with Disruptions, GM – Group Mobilization, PR – Procedures, RU – Rules, TM – Time Management, WA – Working Alliance.

* $p > .05$.

For the then-post comparison, the frequency pattern of level of MI was identical for the CG and the IG: Strict MI was present in three out of eight scales (CG: Group Mobilization, Working Alliance, Conflicts; IG: Group Mobilization, Working Alliance, Procedures), strong MI in two scales (CG: Clarity, Procedures; IG: Clarity, Time Management), weak invariance in one scale (both groups: Dealing with Disruptions), and no level of MI in two scales (CG: Rules, Time Management; IG: Rules, Conflicts). Level of MI was the same in both the CG and the IG on five out of eight dimensions; the CG revealed higher levels of MI on one (Conflicts) and lower levels of MI on two dimensions as compared to the IG (Procedures, Time Management). Level of MI was at least strong in five out of eight dimensions in both the CG and the IG. There was no change in factor loadings (highest level of MI: weak invariance) on

Table 4.3: Three Step Procedure for Pre-Post and Then-Post Comparison

Scale	CG							IG						
	Step	χ^2 (df)	RMSEA	CFI	TLI	SRMR	χ^2_{Diff} (df)	Step	χ^2 (df)	RMSEA	CFI	TLI	SRMR	χ^2_{Diff} (df)
Pre-Post Comparison (T1 vs. T2)														
CL	1	3.11(5)	.00	1.00	1.03	.04		1	8.45(5)	.10	.99	.95	.04	
	2	25.11(12)*	.12	.93	.91	.07								
	3_SI(14,15)	9.10(10)	.00	1.00	1.01	.06	5.99(5)							
CO	1	1.26(5)	.00	1.00	1.02	.01		1	7.29(5)	.08	1.00	.99	.03	
	2	28.27(12)*	.13	.97	.96	.08		2	18.76(12)	.09	.99	.98	.06	
	3_WI(all)	14.38(7)*	.11	.98	.97	.09	13.13(2)*	3_SI(17)	14.74(11)	.07	.99	.99	.07	7.45(6)
DD	1	1.23(5)	.00	1.00	1.06	.02		1	17.76(5)*	.20	.93	.79	.06	
	2	22.38(5)*	.10	.95	.93	.09								
	3_SI(all)	1.10(5)	.04	.99	.99	.06	8.86(4)							
GM	1	5.15(5)	.02	1.00	1.00	.04		1	4.12(5)	.00	1.00	1.01	.03	
	2	2.06(12)	.09	.97	.96	.10		2	19.36(12)	.10	.97	.96	.12	
	3_SI(12)	13.38(11)	.05	.99	.99	.07	8.22(6)	3_SI(all)	8.48(9)	.00	1.00	1.00	.08	4.44(4)
PR	1	14.13(5)*	.15	.96	.87	.07		1	7.58(5)	.09	.99	.97	.04	
RU	1	6.90(5)	.07	.99	.98	.06		1	8.60(5)	.11	.98	.95	.05	
	2	37.05(12)*	.16	.92	.91	.08								
	3_SI(all)	9.99(9)	.04	1.00	1.00	.08	3.09(4)							
TM	1	7.78(5)	.08	.99	.97	.04		1	9.95(5)	.12	.99	.96	.04	
	2	26.14(12)*	.12	.96	.94	.13								
	3_WI(all)	12.71(7)	.10	.98	.96	.10								
WA	1	10.72(5)	.12	.97	.92	.05		1	8.59(5)	.11	.99	.95	.05	

continued

186 TEILSTUDIE 3: EVALUATING RESPONSE SHIFT IN TRAINING EVALUATION

Scale	CG							IG						
	Step	χ^2 (df)	RMSEA	CFI	TLI	SRMR	χ^2_{Diff} (df)	Step	χ^2 (df)	RMSEA	CFI	TLI	SRMR	χ^2_{Diff} (df)
Then-Post Comparison (T1r vs. T2)														
CL	1	2.58(5)	.00	1.00	1.03	.04		1	2.58(5)	.00	1.00	1.03	.04	
	2	18.37(12)	.08	.98	.97	.07	15.79(7)*	2	16.25(12)	.07	.99	.98	.11	
	3_SI(14)	12.57(11)	.04	1.00	.99	.07	9.99(6)	3_SI(13,14)	11.80(10)	.05	.99	.98	.09	9.22(5)
CO	1	2.92(5)	.00	1.00	1.01	.01		1	7.59(5)	.09	1.00	.98	.03	
	2	8.49(12)	.00	1.00	1.01	.03	5.60(7)							
DD	1	7.41(5)	.08	.99	.98	.03		1	5.26(5)	.03	1.00	1.00	.05	
	2	28.33(12)*	.13	.95	.94	.09		2	18.67(12)	.09	.97	.97	.11	
	3_WI(2)	9.08(8)	.04	1.00	.99	.04	1.67(3)	3_WI(1,2)	6.34(7)	.00	1.00	1.01	.06	1.08(2)
GM	1	3.83(5)	.00	1.00	1.01	.03		1	4.70(5)	.00	1.00	1.00	.02	
	2	1.70(12)	.00	1.00	1.01	.06	6.87(7)	2	13.97(12)	.05	.99	.99	.10	9.27(7)
PR	1	3.94(5)	.00	1.00	1.01	.03		1	2.23(5)	.00	1.00	1.02	.02	
	2	2.10(12)	.09	.98	.97	.15		2	15.47(12)	.07	.99	.99	.08	13.23(7)
	3_SI(8, 9)	13.74(10)	.07	.99	.98	.08	9.80(5)							
RU	1	11.53(5)*	.13	.99	.96	.06		1	9.12(5)	.11	.99	.96	.09	
TM	1	15.02(5)*	.16	.98	.93	.06		1	6.77(5)	.07	1.00	.99	.02	
								2	19.94(12)	.10	.98	.98	.04	
								3_SI(22,24)	8.99(10)	.00	1.00	1.00	.03	2.22(5)
WA	1	4.97(5)	.00	1.00	1.00	.02		1	4.85(5)	.00	1.00	1.00	.03	
	2	17.94(12)	.08	.98	.98	.09	12.96(7)	2	15.44(12)	.07	.99	.98	.08	10.59(7)

Note. Model: 1 – configural invariance, 2 – strict invariance, 3 – adapted model with highest level of invariance (either SI – strong invariance or WI – weak invariance) and number of item where restrictions were released; χ^2_{Diff} – comparison to Model 1; RMSEA – Root Mean Square Error of Approximation, CFI – Comparative Fit Index, TLI – Tucker Lewis Index, SRMR – Standardized Root Mean Square Residual; CG – Control Group, IG – Intervention Group; CL – Clarity of Program of Action, CO – Conflicts among Students, DD – Dealing with Disruptions, GM – Group Mobilization, PR – Procedures, RU – Rules, TM – Time Management, WA – Working Alliance

any dimension, in either the CG or the IG. Model misfit was predominantly indicated by the RMSEA index.

Results for MI of pre-post vs. then-post comparison indicated that the invariance level was higher for then-post comparisons than for pre-post comparisons in both groups on the majority of dimensions. Exceptions were as follows: The level of MI was the same for Clarity in the CG (only strong MI) and Rules in the IG (no level of MI achieved). The level of MI was lower for then-post comparison for Dealing with Disruptions (CG), Time Management (CG), and Conflicts (IG).

4.4. Discussion

RS bias is of particular concern for research into the effectiveness of training programs (Pratt et al., 2000, Gandhi, 2013). A basic assumption is that observed change may be the result of both true change and RS. The two most common approaches to study and account for RS are the then-test procedure and a SEM approach. The former can be used instead of a traditional pre-test to arrive at a RS adjusted estimate of change. The latter serves to identify invariant item parameters, account for possible RS and assess true change.

This study had roughly three objectives with regard to aspects of reliability and validity of the respective approaches. First, to implement the SEM approach in an educational training context. Second, to apply MI to the then-post comparison. Third, to study differential RS effects due to the kind of treatment on indices of RS.

Concordant with our first hypothesis, RS was present on all target dimensions as measured by MI for pre-post comparison. Only one dimension met strong invariance for both groups. The majority of pre-post scores did not yield invariant measures (Meredith, 1993). Type of RS appeared to be group specific. In the IG, only two types of RS occurred: reconceptualization RS (changed patterns of the common factor loadings over time) and non-uniform recalibration RS (changed meaning of the option of answers), whereas the former was the most frequent. In contrast, reprioritization, reconceptualization, and non-uniform recalibration RS occurred in the CG, whereas the latter was the most frequent. Thus, level of MI tends to be higher (four dimensions) or the same (three dimensions) in the CG as compared to the IG, indicating that there was less RS in the CG than the IG. These findings show that CM dimensions were most often reconceptualized in the IG as opposed to a more frequent change in the range of answers in the CG. Apparently, the IG treatment, where students not only gained new insights into strategies to cope with and manage problem behavior, but also interacted with peers and deployed learned strategies in role-plays, induced a profound redefinition

of CM aspects. In addition, no uniform recalibration RS (i.e., changed mean structure of the observed variables) and only two reprioritization RS (only in the CG) occurred. Whereas HR-QOL research repeatedly reported findings of those shifts, it needs to be studied in future research, whether changed mean structure and changed loadings are really an RS-issue for recursive constructs. Recently introduced individualized SEM approaches could broaden our knowledge about interactions of type of RS and the type of construct studied (Mayo, Scott, & Ahmed, 2009).

Contrary to our second hypothesis, the then-test procedures did not yield the same results as the SEM approach. Instead, only two dimensions were object to RS in both groups and one additional dimension in each group. Patterns of RS comparing both approaches could not be identified either, for example that pre-then RS was present on those dimensions that did not yield any level of MI. These findings are opposite to those of Ahmed, Mayo, Wood-Dauphinee et al. (2005) who found no shift with SEM, but some shift with then-tests, and incongruent with Visser et al.'s (2005) findings, who found convergence of the two approaches. In addition, contrary to the results based on SEM and contrary to our hypotheses, the IG did not have a larger amount of RS than the CG in pre-then scores. This raises the question of what is assessed by the depicted methods and which explanatory variables lead to or do not lead to a shift in responses. Apparently, either SEM is more sensitive in detecting RS, or the then-test procedure is too general to detect RS bias adequately. Probably, the diverse types of RS interfere with each other resulting in comparable pre-then means even though RS did occur. Consequently, the then-test procedure may not be reliable enough to detect treatment specific RS (Hill & Betz, 2005). However, those findings are only first indicators for the complex effects of treatment type on RS as studied in our two kinds of treatment groups. In order to deepen our understanding of RS as a potential side-effect in control-trials, future studies should also incorporate a waiting control group or implement a design, where subjects completing then-tests stay without pre-tests (Hoogstraten, 1982).

Interestingly, identified RS based on the pre-then difference scores indicate, that subjects did also underestimate initial knowledge: In the IG, one out of three RS effects indicate initial under-estimation. In the CG, all three RS effects indicate under-estimation. With regard to their relative mean levels, under-estimation was present on those dimensions that were rated lower than others at pre-testing. These findings suggest two conclusions: First, training programs have the ability to strengthen participants' confidence in judging their own knowledge, as is shown by the over-estimation of the working-alliance. Both the CG and the IG subjects realized that they already knew a lot about this aspect of CM, even though they

were not that confident when they entered the course. This interpretation supports Hill & Betz's (2005) conclusion of the then-test being an effective procedure to reinforce feelings of confidence and promote this type of reflection after the course. Second, these findings again suggest a treatment specific effect: It could be, that only the interactive practices led to critical self-reflection, as it may have changed the frame of reference to a more realistic or an altered social norm, for example with regard to dealing with disruptions (Norman, 2003). Both conclusions are of particular relevance for the composition of university courses and stress the importance of interactive opportunities to learn.

Our third hypothesis could be confirmed only partially. As hypothesized, then-post comparison disclosed predominantly higher levels of MI as opposed to MI of pre-post comparison. In addition, the comparison of the two treatment groups indicated the same level of MI on five out of eight dimensions, suggesting MI of then-post self-ratings is rather not affected by type of treatment. However, in both groups only four out of eight dimensions met at least strong invariance and even two dimensions did not meet any level of MI. Accordingly, also then-post comparison was not consistently measurement invariant. Comparing our findings to those of Nolte et al. (2009), who implemented the then-test in direct comparison to the post-test, systematic research seems necessary to detect those conditions that lead to at least strong invariant then-post measures. The stability in respective estimates may be strengthened by responding *consecutively* (Aiken & West, 1990; Schwartz & Sprangers, 2010). In addition, research should systematically address this issue by varying procedures and instructions in order to find the optimal strategy for an invariant measurement of then- and post-scores (e.g., the mere announcement of external validation, or encouragement of explicit comparison to a known reference group or an ideal norm, Aiken & West, 1990; Norman, 2003). Also, implementation of individual techniques (such as personal interviews or direct assessment of change) could uncover, why measures at the same occasion lead to invariant parameters.

4.4.1. Limitations

The current study was limited in several ways. First, our research design only allowed for the comparison of two treatment groups. An additional waiting control group could have broadened our understanding of treatment-specific RS, and would further have facilitated the disentanglement of the RS from treatment effects (Schwartz & Sprangers, 2010). In addition, the timeframe between pre- and post-measures were not identical in the two groups. It is possible that identified group specific findings may have interfered with this fact. Also, group differences in ratio of females/males could be relevant to our findings. Moore & Tananis

(2009) found higher pre- and then-scores for males as opposed to females with regard to self-efficacy beliefs.

Second, identified level of measurement invariance strongly depends on the chosen cut-off values of fit indices (Brown, 2006). Predominantly, constraints needed to be relieved because the RMSEA achieved a critical value. Hu & Bentler (1998) pointed to the fact that the RMSEA index is less preferable at small sample sizes (≤ 250). Even though we have chosen a rather liberal cut-off value ($< .08$) it needs to be further analyzed whether larger sample sizes tend to identify less serious RS (i.e., uniform and non-uniform recalibration RS).

4.4.2. Conclusions

The current study evaluated RS based on the then-test procedure and on tests of measurement of invariance. We could demonstrate the feasibility of the MI approach to the detection of RS to other concepts than HR-QOL. However, the two approaches did not lead to the same results. In particular, MI seemed to be more sensitive to RS; the then-test identified strongest RS just on those dimensions that were particularly focused in both treatments and appear to have increased most after treatments. For both reasons, it seems that treatment-effects and RS bias are rather confounded in the then-test procedure. Still, we do not conclude that MI outperformed the then-test. The found levels of MI render it impossible to compare pre-post-scores. In contrast, then-post comparison revealed high levels of invariance on the majority of studied dimensions indicating that change scores which were derived from the then-post comparison, may be less robust when compared to the traditional pre-post method for assessing program outcomes.

Contrary to current research praxis, the evaluation of RS needs to be an integral part of program evaluation as it is essential for interpreting treatment-effects. If RS is not tested, we cannot be sure whether a change in observed test-scores fully represents true change in the target construct, or also a change in the response behavior of the respondent (Vandenberg & Lance, 2000; Oort, 2005). However, there seems to be no single best approach to ascertain RS to date (Barclay-Goddard et al., 2009). Hill and Betz (2005) suggested that in the context of program evaluation, traditional pre-post-designs are preferable when subjects are already familiar with the target construct when they enter training, as it precludes effort justification or recall bias. Furthermore, when pre-then comparisons reveal no RS, pre-post comparisons are supposed to lead to more valid results. In contrast, the then-test procedure appears to be more adequate if pre-testing effects are probable and over-estimation of program effects is unproblematic (Hill & Betz, 2005), or if the subjective experiences of program-related change is of

interest. Still, it is crucial to test whether comparison is meaningful—that is, assumptions of measurement invariance are met—and whether violation of MI is small enough to justify at least partial MI in order to obtain changes in latent means. In either pre-post or then-post comparison, it is advisable to use more sophisticated analytic techniques in addition to simple change scores or group differences, which help to tease out error sources (Barclay-Goddard, Lix, Tate, Weinberg, & Mayo, 2009; King-Kallimanis, Oort, Visser, & Sprangers, 2009; Rapkin, 2009). Eventually, as RS was found to differ across groups, it seems promising to analyze predictors of a RS in general but also of the various types of RS in more detail (Barclay-Goddard, Epstein et al., 2009).

4.5. References

- Ahmed, S., Mayo, N. E., Corbiere, M., Wood-Dauphinee, S., Hanley, J., & Cohen, R. (2005). Change in quality of life of people with stroke over time: True change or response shift? *Quality of Life Research, 14*(3), 611–627. doi:10.1007/s11136-004-3708-0
- Ahmed, S., Mayo, N. E., Wood-Dauphinee, S., Hanley, J. A., & Cohen, S. R. (2005). The structural equation modeling technique did not show a response shift, contrary to the results of the then test and the individualized approaches. *Journal of Clinical Epidemiology, 58*(11), 1125–1133. doi:10.1016/j.jclinepi.2005.03.003
- Ahmed, S., Schwartz, C., Ring, L., & Sprangers, M. A. (2009). Applications of health-related quality of life for guiding health care: Advances in response shift research. *Journal of Clinical Epidemiology, 62*(11), 1115–1117. doi:10.1016/j.jclinepi.2009.04.006
- Aiken, L. S., & West, S. G. (1990). Invalidity of true experiments: Self-report pretest biases. *Evaluation Review, 14*(4), 374–390. doi:10.1177/0193841X9001400403
- Barclay-Goddard, R., Epstein, J. D., & Mayo, N. E. (2009). Response shift: A brief overview and proposed research priorities. *Quality of Life Research, 18*(3), 335–346. doi:10.1007/s11136-009-9450-x
- Barclay-Goddard, R., Lix, L. M., Tate, R., Weinberg, L., & Mayo, N. E. (2009). Response shift was identified over multiple occasions with a structural equation modeling framework. *Journal of Clinical Epidemiology, 62*(11), 1181–1188. doi:10.1016/j.jclinepi.2009.03.014
- Bollen, K., Lennox, & R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*(2), 305–314.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- Cantrell, P. (2003). Traditional vs. retrospective pretests for measuring science teaching efficacy beliefs in preservice teachers. *School Science and Mathematics, 103*(4), 177–185. doi:10.1111/j.1949-8594.2003.tb18116.x
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum Associates.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181–199. doi:10.3102/0013189X08331140

- Donaldson, G. W. (2005). Structural equation models for quality of life response shifts: Promises and pitfalls. *Quality of Life Research*, 14(10), 2345–2351. doi:10.1007/s11136-005-3977-2
- Drennan, J., & Hyde, A. (2008). Controlling response shift bias: the use of the retrospective pre-test design in the evaluation of a master's programme. *Assessment & Evaluation in Higher Education*, 33(6), 699–709. doi:10.1080/02602930701773026
- Gandhi, P. K., Ried, L. D., Huang, I.-C., Kimberlin, C. L., & Kauf, T. L. (2013). Assessment of response shift using two structural equation modeling techniques. *Quality of Life Research*, 22(3), 461–471. doi:10.1007/s11136-012-0171-1
- Gibbons, F. X. (1999). Social comparison as a mediator of response shift. *Social Science & Medicine*, 48(11), 1517–1530. doi:10.1016/S0277-9536(99)00046-5
- Grossman, P. L. (2005). Research on pedagogical approaches in teacher education. In M. Cochran-Smith, & K. Zeichner (Eds.), *Studying teacher education* (pp. 425-476). Mahwah, NJ: Lawrence Erlbaum.
- Hill, L. G., & Betz, D. L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation*, 26(4), 501–517. doi:10.1177/1098214005281356
- Henry, B., Moffitt, T. E., Caspi, A., Langley, J., & Silva, P. A. (1994). On the “remembrance of things past”: A longitudinal evaluation of the retrospective method. *Psychological Assessment*, 6(2), 92–101. doi:10.1037/1040-3590.6.2.92
- Holden, G., Barker, K., Rosenberg, G., & Onghena, P. (2008). The evaluation self-efficacy scale for assessing progress toward CSWE accreditation related objectives: A replication. *Research on Social Work Practice*, 18(1), 42–46. doi:10.1177/1049731507303954
- Hoogstraten, J. (1982). The retrospective pretest in an educational training context. *Journal of Experimental Education*, 50(4), 200–204.
- Hoogstraten, J. (1985). Influence of objective measures on self-reports in a retrospective pre-test-posttest design. *Journal of Experimental Education*, 53(4), 207–210.
- Howard, G. S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. *Evaluation Review*, 4(1), 93–106. doi:10.1177/0193841X8000400105
- Howard, G. S., & Dailey, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology*, 64(2), 144–150.

- Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424–453. doi:10.1037/1082-989X.3.4.424
- King-Kallimanis, B., Oort, F. J., Visser, M., & Sprangers, M. (2009). Structural equation modeling of health-related quality-of-life data illustrates the measurement and conceptual perspectives on response shift. *Journal of Clinical Epidemiology*, 62(11), 1157–1164. doi:10.1016/j.jclinepi.2009.04.004
- Lam, T. C. M. (2003). A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation*, 24(1), 65–80. doi:10.1177/109821400302400106
- Lucas, R. E., & Baird, B. M. (2006). Global self-assessment. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (1st ed., pp. 29–59). Washington, DC: American Psychological Association.
- Mayo, N. E., Scott, S. C., & Ahmed, S. (2009). Case management poststroke did not induce response shift: the value of residuals. *Journal of Clinical Epidemiology*, 62(11), 1148–1156. doi:10.1016/j.jclinepi.2009.03.020
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543. doi:10.1007/BF02294825
- Moore, D., & Tananis, C. A. (2009). Measuring change in a short-term educational program using a retrospective pretest design. *American Journal of Evaluation*, 30(2), 189–202. doi:10.1177/1098214009334506
- Muthén, L.K. & Muthén, B.O. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.
- Nolte, S., Elsworth, G. R., Sinclair, A. J., & Osborne, R. H. (2009). Tests of measurement invariance failed to support the application of the “then-test”. *Journal of Clinical Epidemiology*, 62(11), 1173–1180. doi:10.1016/j.jclinepi.2009.01.021
- Norman, G. (2003). Hi! How are you? Response shift, implicit theories and differing epistemologies. *Quality of Life Research*, 12, 239–249.
- Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of Life Research*, 14(3), 587–598. doi:10.1007/s11136-004-0830-y
- Oort, F. J., Visser, M. R. M., & Sprangers, M. A. G. (2005). An application of structural equation modeling to detect response shifts and true change in quality of life data from cancer patients undergoing invasive surgery. *Quality of Life Research*, 14(3), 599–609. doi:10.1007/s11136-004-0831-x

- Oort, F. J., Visser, M. R., & Sprangers, M. A. (2009). Formal definitions of measurement bias and explanation bias clarify measurement and conceptual perspectives on response shift. *Journal of Clinical Epidemiology*, *62*(11), 1126–1137. doi:10.1016/j.jclinepi.2009.03.013
- Piwowar, V., Thiel, F. & Ophardt, D. (2013). Training inservice teachers' competencies in classroom management: A quasi-experimental study with teachers of secondary schools. *Teaching and Teacher Education*, *30*, 1-12. doi: 10.1016/j.tate.2012.09.007
- Pratt, C. C., McGuigan, W. M., & Katzev, A. R. (2000). Measuring program outcomes: Using retrospective pretest methodology. *American Journal of Evaluation*, *21*(3), 341–349. doi:10.1016/S1098-2140(00)00089-8
- Rapkin, B. D. (2009). Considering the application of the trait/state distinction for response shift research: Continuing the conversation. *Journal of Clinical Epidemiology*, *62*(11), 1124–1125. doi:10.1016/j.jclinepi.2009.02.009
- Schmitt, N. (1982). The use of analysis of covariance structures to assess beta and gamma change. *Multivariate Behavioral Research*, *17*, 343–358.
- Schwartz, C. E., & Sprangers, M. A. G. (1999). Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Social Science & Medicine*, *48*(11), 1531–1548. doi:10.1016/S0277-9536(99)00047-7
- Schwartz, C. E., & Sprangers, M. A. G. (2010). Guidelines for improving the stringency of response shift research using the thentest. *Quality of Life Research*, *19*(4), 455–464. doi:10.1007/s11136-010-9585-9
- Schwartz, C. E., Bode, R., Repucci, N., Becker, J., Sprangers, M. A. G., & Fayers, P. M. (2006). The clinical significance of adaptation to changing health: A meta-analysis of response shift. *Quality of Life Research*, *15*(9), 1533–1550. doi:10.1007/s11136-006-0025-9
- Schwartz, C. E., Sprangers, M. A. G., Carey, A., & Reed, G. (2004). Exploring response shift in longitudinal data. *Psychology & Health*, *19*(1), 51–69. doi:10.1080/0887044031000118456
- Schwarzer, R., & Hallum, S. (2008). Perceived teacher self-efficacy as a predictor of job stress and burnout: Mediation Analyses. *Applied Psychology*, *57*, 152–171. doi:10.1111/j.1464-0597.2008.00359.x
- Sibthorp, J., Paisley, K., Gookin, J., & Ward, P. (2007). Addressing response-shift bias: Retrospective pretests in recreation research and evaluation. *Journal of Leisure Research*, *39*(2), 295–315.

- Sprangers, M. A. G., & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: A theoretical model. *Social Science & Medicine*, 48(11), 1507–1515.
- Sörbom, D. (1975). Detection of correlated errors in longitudinal data. *British Journal of Mathematical and Statistical Psychology*, 28(2), 138-151.
- Tanaka, J. (1987). “How big is big enough?” Sample size and goodness of fit in structural equation models with latent variables. *Child Development*, 58(1), 134-146.
- Thiel, F. & Blüthmann, I. (2009). *Ergebnisse der Evaluation der lehrerbildenden Studiengänge an der Freien Universität Berlin. Sommersemester 2009.*]. Results of the evaluation of preservice teacher education at Freie Universität Berlin, summer semester 2009.] Berlin.
- Thiel, F., Ophardt, D. & Piwowar, V. (2013). Abschlussbericht des Projekts „Kompetenzen des Klassenmanagements (KODEK). Entwicklung und Evaluation eines Fortbildungsprogramms für Lehrkräfte zum Klassenmanagement.“ [Report of the research project „Classroom management competencies. Development and evaluation of a teacher training programme in classroom management.“]. Retrieved from http://www.ewi-psy.fu-berlin.de/einrichtungen/arbeitsbereiche/schulentwicklungsforschung/downloads/Lehrmasterbefragung_2009.pdf?1310986750
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70. doi:10.1177/109442810031002
- Visser, M. R. M., Oort, F. J., & Sprangers, M. A. G. (2005). Methods to detect response shift in quality of life data: A convergent validity study. *Quality of Life Research*, 14(3), 629–639. doi:10.1007/s11136-004-2577-x

Gesamtdiskussion

5. Gesamtdiskussion

In unterschiedlich paradigmatisch, methodisch und konzeptuell ausgerichteten Modellen sowie empirischen Studien und Metaanalysen wird seit nunmehr einigen Jahrzehnten die zentrale Bedeutung von Klassenmanagement für das Lernen von Schülerinnen und Schülern sowie die Lehrergesundheit herausgestellt (Brophy, 2006; Evertson & Harris, 1992; Hattie, 2009; Kounin, 1976; Simonsen, Fairbanks, Briesch, Myers & Sugai, 2008; Wang, Haertel & Walberg, 1993). Während Klassenmanagement international, insbesondere in den USA, ein eigenes Forschungsfeld markiert, blieb es in Deutschland lange Zeit weitestgehend unbeachtet (Ophardt & Thiel, 2012). Entsprechend wurden in Kapitel 1 Desiderata bezüglich der Konzeptualisierung des Konstrukts, der Verfügbarkeit nutzbarer empirischer Verfahren zur Erfassung sowie der systematischen Entwicklung dieser Kompetenzfacette aufgezeigt.

Daran anknüpfend wurde in der vorliegenden Arbeit eine Taxonomie zur Konzeptualisierung von Klassenmanagement entworfen und darauf aufbauend ein Schüler- und ein Beobachterfragebogen zur Erfassung von Kompetenzen im Klassenmanagement (KODEK-S und KODEK-B, Teilstudie 1) sowie ein Instrument zur Selbsteinschätzung von Wissen im Klassenmanagement entwickelt (SEWIK). Diese Instrumente kamen bei der Evaluation eines neu entwickelten Trainings zum Klassenmanagement für Lehrkräfte der Sekundarstufe 1 zum Einsatz (Teilstudie 2). Schließlich wurde anhand des SEWIK im Rahmen einer Interventionsstudie mit Studierenden des Lehramts das Phänomen des Response Shift untersucht (Teilstudie 3).

Nachfolgend werden die zentralen Ergebnisse der durchgeführten Studien anhand der in Kapitel 1 formulierten Fragestellungen dargestellt und integrierend diskutiert (Kapitel 5.1), Grenzen und Stärken der Arbeit herausgearbeitet (Kapitel 5.2) sowie abschließend in einem Ausblick Implikationen für Forschung und Praxis abgeleitet (Kapitel 5.3).

5.1. Zusammenfassung der Ergebnisse

5.1.1. Was konstituiert Klassenmanagement und wodurch ist gelingendes bzw. erfolgreiches Klassenmanagement gekennzeichnet?

Im deutschsprachigen Raum ist Klassenmanagement zum einen noch ein relativ „junges Konstrukt“: Es liegen nur wenige deutschsprachige Studien vor, die genuin Klassenmanagement oder Subaspekte von Klassenmanagement systematisch untersucht haben (Helmke, Schneider & Weinert, 1986; Kunter, Baumert & Köller, 2007; Mayr, 2006; Neuenschwander,

2006; Schönbächler, 2006; Thiel, Richter & Ophardt, 2013). Zum anderen wird unter Klassenmanagement sehr unterschiedliches subsumiert: Es finden sich eher enge, auf den Umgang mit Störungen und Zeitnutzung orientierte Auffassungen (z. B. Blömeke, Felbrich & Müller, 2008), bis hin zu sehr weiten Konzeptualisierungen, die auch die Diagnosekompetenz und das Lehrerehos mit einbeziehen (Neuenschwander, 2006, vgl. Kapitel 1.3.2). Forschungsbefunde lassen sich damit nur schwer integrieren und müssen vor dem Hintergrund der zu Grunde gelegten Definition bzw. empirischen Operationalisierung von Klassenmanagement interpretiert werden. So wäre es zum Beispiel interessant zu prüfen, ob sich die von Kunter und Voss (2011) berichteten Ergebnisse, dass Klassenführung ein bedeutsamer positiver Prädiktor für Leistungsängstlichkeit im Fach Mathematik ist (erfasst über die Skalen *Klassenführung*, *Zeitverschwendung* und *Störungen*), mit dem hier vertretenen Ansatz von Klassenmanagement replizieren ließen.

Angelehnt an die Arbeiten von Ophardt und Thiel (2007, 2008, 2013) und den internationalen Forschungsstand (Doyle, 2006; Emmer, Evertson & Worsham, 2003; Evertson & Harris, 1992; Evertson & Weinstein, 2006; Weinstein, 2007) wurden auf Basis einer umfassenden Literaturrecherche und Integration verschiedener Forschungsparadigmen herausgearbeitet, welche zentralen Strategien und Dimensionen Klassenmanagement konstituieren. Diese wurden in einem taxonomischen Modell systematisiert: Die unterschiedlichen Strategien von Klassenmanagement sind danach entweder auf die Regulation von Schülerverhalten (Regelarbeit, Störungsintervention, Monitoring–Management des Schülerverhaltens), die Instruktion (Prozeduren, Gruppenmobilisierung, Zeitmanagement, Klarheit des Handlungsprogramms–Management der Instruktion) oder das Verhältnis von Lehrkraft und Schülerinnen und Schülern (Arbeitsbündnis, Umgang mit Konflikten unter Schülern–Management der Lehrer–Schüler-Beziehung) bezogen. Das Modell unterscheidet sich damit deutlich von Auffassungen anderer Arbeitsgruppen, die ausschließlich die erste und vereinzelt Aspekte der zweiten Dimension unter Klassenmanagement fassen (z. B. Gold & Holodynski, 201; Helmke, 2007; Kunter et al., 2007).

Dieses umfassendere Begriffsverständnis wurde allen drei Teilstudien zu Grunde gelegt: Das Modell bildete den Ausgangspunkt für die in Teilstudie 1 entwickelten Fragebogenverfahren KODEK-S und KODEK-B (Fremdbericht) sowie das Selbsteinschätzungsinstrument SEWIK. Dass sich das entwickelte Modell empirisch bewährt, konnte daher zum einen über Faktorenanalysen (postulierte Multidimensionalität) und zum anderen über Zusammenhänge mit hypothetisch relevanten Kriterien (Bedeutsamkeit der Strategien) im Rahmen von Teilstudie 1 gezeigt werden.

In Bezug auf den ersten Aspekt (postulierte Multidimensionalität) werden die Ergebnisse der explorativen Mehrebenen-Faktorenanalyse des Schülerfragebogens KODEK-S berichtet und diskutiert. Erstens konnten sowohl die drei postulierten Dimensionen als auch die darunter gefassten Subdimensionen bzw. Strategien (überwiegend) empirisch identifiziert werden, einschließlich der Zuordnung von Strategien zu den jeweiligen Dimensionen. Da für die Faktorenanalyse ein exploratives und kein konfirmatives Vorgehen gewählt wurde, wurden die Dimensionen und Subdimensionen nicht als Faktoren erster und zweiter Ordnung extrahiert, sondern auf Ebene der *geteilten* (Dimensionen) und der *individuellen* Wahrnehmung (Subdimensionen). Während im Individualurteil also einzelne Strategien von Klassenmanagement unterschieden werden, erlaubt das aggregierte Schülerurteil lediglich eine Unterscheidung zwischen den drei allgemeineren Dimensionen Management des Schülerverhaltens, Management der Instruktion und Management der Lehrer-Schüler-Beziehung. Im Vorgriff auf die zweite Fragestellung (siehe Kapitel 5.1.2) sind diese Ergebnisse von zentraler Bedeutung für die empirische Analyse von Klassenmanagement auf der Basis von Schülerbefragungen: Einerseits wird häufig argumentiert, dass nur aggregierte Schülerurteile zuverlässig genug sind, um zu einer annähernd objektiven bzw. allgemeingültigen Evaluation von Unterrichtsmerkmalen im Sinne einer geteilten Wahrnehmung zu gelangen (Gärtner, 2010; Lüdtke, Trautwein, Kunter & Baumert, 2006, vergleiche auch die ICC[1] als Indikatoren der Reliabilität eines Einzelurteils). Andererseits weisen die vorliegenden Ergebnisse darauf hin, dass mit der Aggregation ein erheblicher Informationsverlust verbunden sein kann und die Konstruktschärfe gegebenenfalls verloren geht. Eine erhöhte Reliabilität geht damit zu Lasten der Validität der Messung. Diese Frage wurde bereits im Rahmen anderer Studien aufgegriffen, blieb bislang jedoch unbeantwortet (Muthén & Asparouhov, 2011).

Zweitens zeigen die Ergebnisse der *within*-Faktorenanalyse auf Individualebene, dass die instruktionalen Klassenmanagementstrategien mit der bisherigen Operationalisierung noch nicht trennscharf abgebildet oder dass sie von Schülerinnen und Schülern nicht getrennt voneinander bewertet werden können—Klarheit und Zeitmanagement sowie Prozeduren und Gruppenmobilisierung bilden gemeinsame Faktoren. Theoretisch wäre eher eine schlechte Trennung von Zeitmanagement und Prozeduren plausibel gewesen, da Prozeduren vor allem über die Zeitnutzung an Übergängen operationalisiert wurden (vgl. auch Doyle, 2006; Helmke, 2007 sowie die Skala *effektives Zeit- und Übergangsmanagement*, Lotz & Gabriel, 2012). Mit Blick auf die Items der jeweiligen Skalen kann jedoch eine mögliche Interpretation dieser Daten sein, dass Klarheit und Zeitmanagement deshalb nicht getrennt wahrgenommen werden, weil die Beurteilung von Zeitmanagement über die Wahrnehmung von Klarheit mode-

riert wird: Durch die Mühe, die eine Lehrkraft in ein klares Handlungsprogramm investiert, vermittelt sie den Schülerinnen und Schülern, dass sie die Stunde so strukturiert, dass diese die zur Verfügung stehende Lernzeit optimal nutzen können—was wiederum ein wesentliches Charakteristikum des Zeitmanagements ist. Gruppenmobilisierung und Prozeduren könnten demgegenüber deshalb nicht trennscharf sein, weil auch bei den Items der Skala Prozeduren indirekt Auskunft über den Grad der gleichzeitigen Aktivierung aller Schülerinnen und Schüler (Beschäftigungsradius) gegeben wird—eigentlich ein definitorisches Merkmal von Gruppenmobilisierung. Da auch die interne Konsistenz der eben angesprochenen Skalen an der unteren Grenze ist, sollte eine erneute Itemrevision durchgeführt werden, obwohl sich die inhaltlichen Überschneidungen bzw. Wechselwirkungen der Konstrukte damit nicht abschließend ausschließen lassen. Die übrigen Skalen hingegen können sehr gut abgebildet werden. Ob die theoretisch postulierte Mehrdimensionalität auch im Rahmen des SEWIK identifiziert werden kann, müsste noch geprüft werden.

Drittens erlauben die Ergebnisse der *between*-Faktorenanalyse auf Aggregatebene eine Aussage über die Zusammenhänge der Strategien untereinander: Die Items des Arbeitsbündnisses weisen hohe Ladungen sowohl auf der zweiten (Instruktion) wie der dritten Dimension (Lehrer-Schüler-Beziehung) auf. Damit wird der bereits in anderen Studien nachgewiesene Stellenwert des Arbeitsbündnisses für die instruktionale Steuerung des Schülerverhaltens untermauert (Bergin & Bergin, 1999; Freiberg, 1999; Weinstein, 1999). Auch auf Individual-ebene ist die Skala Arbeitsbündnis jene, die—mit Ausnahme von Prozeduren / Gruppenmobilisierung—mit allen anderen Faktoren bzw. Skalen am höchsten korreliert.

In Bezug auf den zweiten Aspekt (Bedeutsamkeit der Strategien) zeigen die Zusammenhänge zum off-task Verhalten und der Beteiligung der Schülerinnen und Schüler mit den durch KODEK-S und KODEK-B erfassten Kompetenzen im Klassenmanagement, dass alle Strategien mit den genannten Schülermerkmalen assoziiert sind und dass diese Zusammenhänge theoriekonform unterschiedlich stark sind: Distalere Skalen (Klarheit des Handlungsprogramms sowie Umgang mit Konflikten unter Schülern) waren am geringsten mit off-task Verhalten und Beteiligung assoziiert.

Zusätzliche Hinweise auf die Bedeutsamkeit und Trennschärfe der Skalen geben die hypothetisch postulierten Veränderungen, die in Teilstudie 2 (inferenzstatistisch), insbesondere jedoch in Teilstudie 3 (deskriptivstatistisch) aufgezeigt werden konnten (hinsichtlich des Wissens über effektive Störungsintervention, vgl. Kapitel 5.1.3).

Mit der Absicht, ein differenziertes, aber sparsames Modell zu entwickeln blieben einige Aspekte von Klassenmanagement in der vorliegenden Arbeit ausgeklammert, wie das phy-

sische Arrangement der Lernumgebung (Briesch & Chafouleas, 2009; Doyle, 2006; Fifer, 1986; Simonsen et al., 2008; Weinstein, 2007), Elternarbeit (Eichhorn, 2008; Freiberg & Lapointe, 2006; Marzano, 2003) oder Steuerungsprozesse außerhalb der Klassensituation (wie Schulregeln oder Umgang mit schwierigem Schülerverhalten als Angelegenheit von Teams, Jahrgangsstufen oder der Schule, Brophy, 2006; Gottfredson, Gottfredson & Hybl, 1993; Marzano, 2003; Ophardt & Thiel, 2008), welche hier nur mittelbar über die Skala Konflikte abgebildet wurden. Im Zuge einer Weiterentwicklung des Modells sollte jedoch eine Erweiterung um diese Aspekte in Betracht gezogen werden.

5.1.2. Wie können Kompetenzen im Klassenmanagement empirisch fassbar gemacht werden? Liegen entsprechende Verfahren vor, die eine standardisierte, reliable, valide und auch ökonomische Erfassung von Klassenmanagement ermöglichen?

Obwohl zahlreiche Befunde über Klassenmanagement und die Wirksamkeit spezifischer Trainings vorliegen gibt es doch erstaunlich wenig Instrumente, die sich für die wissenschaftliche Untersuchung von Klassenmanagement etabliert haben. In der Unterrichtsforschung werden klassischerweise Schüler- und Beobachterratings eingesetzt (Clausen, 2002); in der Kompetenzforschung gab es gerade in jüngerer Zeit Bemühungen, berufliche Kompetenzen durch Tests mittels advokatorischen Zugangs oder Multiple-Choice Fragen erfassen zu können (also unabhängig vom Unterricht der Lehrkraft, vgl. z. B. Gold, Förster & Holodynski, 2013; König & Blömeke, 2009; Oser, Curcio & Düggele, 2007; Voss, Kunter & Baumert, 2011). Aufgrund der noch eingeschränkten Nutzbarkeit (vgl. Kapitel 1.3.2) für die Veränderungsmessung wurde der Fokus in der vorliegenden Arbeit jedoch auf die fragebogengestützte Erfassung von Klassenmanagement gelegt.

Es liegen einige wenige, gut konstruierte Fragebogenverfahren zur Erfassung von Klassenmanagement vor (Lotz, Berner, Gabriel, Post, Faust & Lipowsky, 2011; Mayr, Eder & Fartacek, 2005; Schönbachler, 2004a, 2004b), diesen liegt jedoch nur zum Teil das hier dargestellte Verständnis von Klassenmanagement zugrunde oder sie weisen hinsichtlich der psychometrischen Gütekriterien noch Entwicklungsbedarf auf. Selbsteinschätzverfahren zum Wissen in Klassenmanagement liegen im deutschsprachigen Raum bislang nicht vor. Basierend auf bereits existierenden Instrumenten wurden daher drei neue Instrumente entwickelt (KODEK-S, KODEK-B und SEWIK) und in mehreren Pilotierungsschritten optimiert. Die Inhaltsvalidität der genannten Instrumente wurde durch die modellbasierte Fragebogenkonstruktion und durchgeführte Expertendiskussionen des ersten Itempools abgesichert. Die entwickelten Instrumente werden nachfolgend unter Rückgriff auf die Ergebnisse der drei Teil-

studien diskutiert (Teilstudien 1 und 2 für KODEK-S und KODEK-B, Teilstudien 2 und 3 für SEWIK).

KODEK-S weist–mit Ausnahme der schlechten Trennbarkeit der Skalen des instruktionalen Managements–sehr gute Testeigenschaften auf. Die Items bzw. Skalen klären zwischen den Klassen bis zu 39% (Itemebene) bzw. 48% (Skalenebene) der Varianz auf, erfassen also eine geteilte, intersubjektive Wahrnehmung der Schülerinnen und Schüler und deuten auf eine gute Differenzierungsfähigkeit zwischen Lehrkräften hin. Dies ist nicht für alle Merkmale der Unterrichtsqualität selbstverständlich (vgl. Kunter & Baumert, 2006). Dementsprechend hoch ist die Interraterreliabilität ICC(2): Das aggregierte Schülerurteil ist ein sehr zuverlässiges Maß zur Beurteilung der operationalisierten Strategien des Klassenmanagements, auch in Bezug auf die Retestreliaibilität. Die internen Konsistenzen der Skalen auf Individualebene sind mindestens akzeptabel oder gut. Erwartungskonforme Zusammenhänge zu Drittvariablen (Störverhalten und Schülerbeteiligung sowie Lehrerselbstwirksamkeit) konnten erste Hinweise auf die kriteriale Validität des Instruments liefern. Auf zum Teil paradoxe Zusammenhänge mit der Lehrerbelastung wird nachfolgend noch gesondert eingegangen. Als Instrument zur Veränderungsmessung in Teilstudie 2 konnte sich KODEK-S darüber hinaus ebenfalls bewähren: Die auf Basis des Schülerurteils nachgewiesenen Veränderungen des Klassenmanagements der Lehrkräfte hatten zum Teil eine Entsprechung im tatsächlichen Verhalten der Schülerinnen und Schüler: Schülerbeteiligung *und* die von Schülerinnen und Schülern beurteilten Kompetenzen im Klassenmanagement nahmen–in Bezug auf die Mittelwerte–in der Kontrollgruppe eher ab und in der Interventionsgruppe eher zu.

KODEK-B weist ebenfalls sehr gute Testeigenschaften auf, jedoch mit einigen Einschränkungen. Die Interraterreliabilität war hoch und auch für einzelne Beurteiler noch zufriedenstellend. Allerdings weisen die niedrigen Korrelationen bei einer Messwiederholung im Abstand von 3 Monaten darauf hin, dass zwar situationsgenaue Beurteilungen vorgenommen werden können, dass diese jedoch stark durch den Kontext determiniert sind. Dies kann zum einen auf die untersuchte Stichprobe zurückzuführen sein: Es handelte sich um 21 Referendare, deren Performanz möglicherweise stark schwankt (auch wegen variierender Anforderungen) bzw. deren Kompetenzen ungleichmäßig zugenommen haben können. Gleichzeitig kann dies aber auch als Hinweis darauf gesehen werden, dass KODEK-B unter bestimmten Bedingungen (z. B. bei Berufsanfängern, am Beginn des Schuljahres oder kurz vor Zeugnisvergabe, vgl. Doyle, 2006), mindestens zwei verschiedene Unterrichtsstunden erfassen sollte, um den Einfluss potentiell relevanter Drittvariablen möglichst gering zu halten. KODEK-B

wies ebenfalls die postulierten Zusammenhänge zu relevanten Drittvariablen (off-task Verhalten und Schülerbeteiligung sowie Lehrerselbstwirksamkeit) auf.

KODEK-S und KODEK-B weisen mittlere bis hohe relative Übereinstimmung auf. Damit sind die Einschätzungen im Vergleich zu anderen Studien relativ hoch (Clausen, 2002; Desimone, 2009). Es wird angenommen, dass diese Ergebnisse auf die perspektivenspezifische Itemformulierung zurückzuführen sind. Erwartungswidrige Ergebnisse zeigten sich hinsichtlich des Zusammenhangs von Klassenmanagement und der beruflichen Belastung. Post-hoc Analysen haben gezeigt, dass die Kompetenzen im Klassenmanagement von Referendaren im Mittel eher positiv ($Md_r = .16$ und $.18$, KODEK-S und KODEK-B) und von Lehrkräften im Mittel negativ ($Md_r = -.17$ und $-.10$, KODEK-S und KODEK-B) mit der beruflichen Belastung assoziiert sind. Bisherige Befunde wiesen jedoch übereinstimmend einen negativen Zusammenhang von Belastung und Klassenmanagement nach (Clunies-Ross, Little & Kienhuis, 2008; Friedman, 2006). Diese Effekte sollten genauer untersucht werden, auch unter Berücksichtigung potentieller Prädiktoren dieser Zusammenhänge sowie die Entwicklung über die Zeit (vgl. Klusmann, Kunter, Voss & Baumert, 2012).

Die Instrumente KODEK-S und KODEK-B wurden sorgfältig konstruiert und auf der Grundlage verschiedener statistischer Kriterien weiterentwickelt und optimiert. Nach erfolgter Itemoptimierung des Schülerfragebogens folgt damit nach Lienert und Raatz (1998, vgl. Kapitel 1.3.3) als nächstes die Normierung an einer repräsentativen Stichprobe, so dass qualitative Aussagen über Mittelwerte getroffen werden können. Problematisch an der Entwicklung von Normtabellen ist zunächst jedoch die Auswahl der sozialen Norm bzw. der zu vergleichenden Normierungsgruppen. Während die klassische Gruppierungsvariable Alter hier (je nach Fragestellung) möglicherweise genau *nicht* berücksichtigt werden sollte, ist fraglich, welche Referenzgruppe stattdessen aussagekräftig ist: Sollen Deutschlehrkräfte nur mit Deutschlehrkräften verglichen werden? Gymnasiallehrkräfte nur mit Gymnasiallehrkräften? Kann die Klassengröße, die Zusammensetzung der Schülerinnen und Schüler, das Leistungsniveau einer Klasse berücksichtigt werden? Da KODEK-S und KODEK-B eine Beurteilung auf Ebene der Performanz vornehmen, ist die professionelle Kompetenz stets konfundiert mit den genannten Merkmalen bzw. Anforderungen (vgl. Ophardt & Thiel, 2007), also situations- und kontextabhängig (Kaufhold, 2006; Klieme & Hartig, 2007). Als besonders relevantes Kontextmerkmal hat sich im Zusammenhang mit Klassenmanagement jedoch die Schulform (Kunter & Voss, 2011; Ophardt & Thiel, 2008) und das Alter der Schülerinnen und Schüler erwiesen (Brophy, 1988; Doyle, 2006; Emmer et al., 2003; Weinstein, 2007), so dass eine Normierungsstichprobe repräsentativ hinsichtlich verschiedener Schulformen und Klassenstu-

fen sein sollte. Weiterhin können Ergebnisse von Unterrichtsevaluation auf Basis der Instrumente KODEK-S und KODEK-B interindividuell relativ zu einer bekannten Gruppe (z. B. das eigene Kollegium) oder intraindividuell durch die Identifikation von Stärken und Schwächen oder Veränderungen über die Zeit interpretiert werden (vgl. Helmke, 2009). Es wäre darüber hinaus interessant zu untersuchen, wie stark Kompetenzen im Klassenmanagement einer Lehrkraft kontextabhängig beurteilt werden, z. B. über verschiedene Klassen (Klasse 8a vs. Klasse 8b) oder Fächer (Mathematik in Klasse 8a vs. Geografie in Klasse 8a) hinweg.

Abschließend wird auf den entwickelten Fragebogen SEWIK eingegangen, der in keinem eigenen Beitrag dargestellt, jedoch als Evaluationsinstrument in Teilstudie 2 und 3 eingesetzt und dort beschrieben wurde. Die klassischen Gütemerkmale wurden hierbei nur zum Teil direkt überprüft, so dass vor allem indirekt aus dessen Einsatz in den Teilstudien Rückschlüsse über Objektivität, Reliabilität und Validität gezogen werden. Die Entwicklung des Fragebogens SEWIK beruhte auf den gleichen Konstruktionsprinzipien wie KODEK-S und KODEK-B. Das Instrument ist vor dem Hintergrund entstanden, dass ein komplexes Evaluationsdesign (mit Evaluationskriterien auf verschiedenen Ebenen, Kirkpatrick, 1998, Teilstudie 2), in dem verschiedene Fragebogenverfahren implementiert werden sollten, zeitliche sowie ressourcenbedingte Grenzen aufwies: Der Einsatz eines Tests zur Erfassung von Wissen über einen advokatorischen Zugang wäre im Rahmen der durchgeführten Studie vor allem für die Lehrkräfte kaum zumutbar gewesen (vgl. Nebengütekriterien nach Lienert & Raatz, 1998 sowie Fries & Souvignier, 2009); ein Instrument mit Multiple-Choice Fragen zu Wissen im Klassenmanagement ist jedoch noch nicht verfügbar, da das einzige existierende Instrument nicht frei zugänglich ist (Blömeke et al., 2008). Bezüglich der internen Konsistenz ist SEWIK, obwohl jede Skala nur durch 3 Items repräsentiert ist, sehr zufriedenstellend. Weitere Reliabilitätsnachweise, z. B. bezüglich der Stabilität, stehen noch aus. Hinweise auf die Validität des Instruments gibt der Vergleich des mittleren Wissensniveaus bei der Prämessung in den Teilstudien 2 (Lehrkräfte) und 3 (Studierende): Studierende schätzen ihr Wissen erwartungskonform niedriger ein als Lehrkräfte. In einer post-hoc Analyse der Lehrkräfte aus Teilstudie 2 schätzen Lehrkräfte im Vorbereitungsdienst ihr Wissen geringer ein als Lehrkräfte im Beruf und höher als Studierende (vgl. Thiel, Ophardt & Piwowar, 2013). Dies spricht für die Qualität der Skalierung der Items, da die verwendete Skala zwischen verschiedenen Wissensstufen und nicht nur an einem Ende der Skala zu differenzieren erlaubt. In beiden Interventionsstudien wies das Instrument erwartete Veränderungen nach, wenn diese auch konfundiert sein können mit gewünschter Veränderung (Konfirmationseffekt, vgl. Elliot & Busse, 2004). Dass methodenspezifische Verzerrungseffekte beim Einsatz des SEWIK möglich sind, zeigen

Ergebnisse aus Teilstudie 3. Der Einsatz des Instruments zur retrospektiven Wissens einschätzung repliziert Ergebnisse anderer Studien, nach denen die Retrognose höhere interne Konsistenz aufweist als bei der Statusdiagnostik (also die konventionelle Prämessung, Moore & Tananis, 2009). Dies kann darauf hinweisen, dass bei retrospektiven Beurteilungen konsistenter geantwortet wird, zum Beispiel auf Grund von Lerneffekten, Erinnerungsfehlern oder Halo-Effekten (Elliot & Busse, 2004). Die Ergebnisse der Messinvarianzprüfung zeigten, dass weder die konventionelle noch die retrospektive Prämessung invariant zur Postmessung sind, folglich also hohe situative Anteile die Selbsteinschätzung beeinflussen (genauer, siehe Kapitel 5.1). Zur Erhöhung der Validität solcher Selbsteinschätzungen ist daher vor allem die Vorgabe eines geeigneten Referenzrahmens (z. B. im Vergleich zum Ideallehrer–Idealnorm; im Vergleich zu meinen Kolleginnen und Kollegen/Kommilitoninnen und Kommilitonen–soziale Norm) von zentraler Bedeutung. Der nächste wichtige Schritt ist die Überprüfung der Validität des SEWIK anhand seiner Zusammenhänge mit Kompetenzen im Klassenmanagement und über Tests erfassten Wissens im Klassenmanagement sowie bezüglich seiner Unterscheidbarkeit von Selbstwirksamkeitsüberzeugungen (Cramer, 2010).

5.1.3. Wie können Kompetenzen im Klassenmanagement gezielt weiterentwickelt werden?

Gibt es ausreichend bzw. angemessene Aus- und Weiterbildungsmöglichkeiten für Studierende bzw. Lehrkräfte in unterschiedlichen Entwicklungsstadien?

In Deutschland ist die gezielte Entwicklung von Kompetenzen im Klassenmanagement weder in der Ausbildung systematisch verankert, noch gibt es ein flächendeckendes Angebot an Trainings zur Fort- und Weiterbildung von Lehrkräften im Beruf (Havers, 2010; Ophardt & Thiel, 2013). Gleichzeitig ist die Fortbildungsbereitschaft von Lehrkräften eher gering und an bereits vorliegenden Kompetenzen orientiert (Richter, Kunter, Klusmann, Lüdtke & Baumert, 2011). Dass dennoch ein Bedarf an systematischer Entwicklung von Kompetenzen im Klassenmanagement–insbesondere für Lehrkräfte der Sekundarstufe 1 (Freiberg & Lapointe, 2006)–besteht, zeigen nicht nur die Befragungen von Studierenden, Lehrkräften und Auszubildenden (Jones, 2006), sondern auch das breite Angebot an Praxishilfen und Ratgeberbüchern zu diesem Thema (u. a. Eichhorn, 2008; Rogers & Kühn, 2013; Toman, 2007).

In Teilstudien 2 und 3 wurden daher neu entwickelte Trainings zum Klassenmanagement vorgestellt und evaluiert–ein Training für Lehrkräfte (KODEK, Teilstudie 2) und ein Training für Studierende (Teilstudie 3). Es folgen eine kurze Darstellung der entwickelten Trainings sowie die Ergebnisse der implementierten Evaluationsstudien. Darüber hinaus werden durch die Integration der Ergebnisse aus den beiden Teilstudien Rückschlüsse über

den relativen Wert verschiedener Lerngelegenheiten sowie Aspekte der Passung dieser mit unterschiedlich erfahrenen Teilnehmerinnen und Teilnehmern diskutiert.

Das für Lehrkräfte entwickelte KODEK-Training basiert auf den Prinzipien des *professional learning* (Simons & Ruijters, 2004), die spezifisch auf das Lernen von Professionellen abgestimmt sind: Elaboration, Expansion und Externalisation. Diese Prinzipien erlauben es, hoch prozeduralisiertes Wissen und routinisiertes Handeln weiterzuentwickeln sowie das Handlungsrepertoire der Lehrkräfte zu erweitern. Orientiert an diesem Dreischritt wurde ein Training mit drei Modulen entwickelt. In Modul 1 wird auf Basis direkter Instruktion in einem Wechsel von Vortrag, videografierten Praxisbeispielen und Videoanalyse in Kleingruppen eine gemeinsame Wissensbasis erarbeitet und das Repertoire deklarativ verfügbarer Strategien erweitert. Dies entspricht der Kompetenzstufe der Akkumulation (bzw. Rekonzeptualisierung) von Wissensbeständen nach Schraw (2006, vgl. Kapitel 1.4.1). In Modul 2 werden im Rahmen von Microteaching und Rollenspielen implizite Routinen reflektiert, neu erlernte Strategien in gestellten Situationen ausprobiert und beobachtungs-basiertes Feedback erprobt. Bezogen auf Schraw (2006) werden damit Prozesse der Integration in und Anpassung an vorhandene Schemata angeregt. In Modul 3 steht die Reflexion der eigenen Unterrichtspraxis im Fokus: Auf Basis einer videografierten Unterrichtsstunde, in der eine erste Implementation neuer Strategien erprobt wurde, wird die eigene Praxis im Rahmen von angeleiteten Videozirkeln reflektiert, evaluiert und diskutiert (*reflection on action*, Schön, 1983). Dabei werden Metastrategien erlernt, um den eigenen Unterricht–auch durch kollegiales Feedback–systematisch analysieren und reflektieren zu können.

In einem multikriterialen Prä-Post Design wurde das KODEK-Training durch den Vergleich mit einer Kontrollgruppe evaluiert. Die Kontrollgruppe nahm lediglich an Modul 1 teil (Fokus: Elaboration), das als traditionelles Trainingsformat betrachtet wurde. Nach dem Training konnten in der Interventionsgruppe positive Effekte auf das Wissen und Können der teilnehmenden Lehrkräfte sowie auf die Schülerbeteiligung nachgewiesen werden. Obwohl das Wissen über Klassenmanagement nach dem Training auch in der Kontrollgruppe höher eingeschätzt wurde, war der Wissenszuwachs bei den Lehrkräften der Interventionsgruppe stärker ausgeprägt. Dies verdeutlicht, dass–wie auch Simons und Ruijters (2004) postulieren–alle drei Prozesse (Elaboration, Expansion und Externalisation) wichtig sind, um sich neues Wissen anzueignen, und dass diese Wissensbasis sich nicht isoliert, sondern in Rückkopplung mit Praxiserfahrungen entwickelt wird. Bezogen auf Schraws (2006) Konzeptualisierung von Kompetenzentwicklung deuten diese Ergebnisse darauf hin, dass sein in Stufen konzipiertes

Modell transaktionale Anteile enthält (Petermann, Niebank & Scheithauer, 2004; Sameroff, 1995), die Kompetenzstufen höherer Ordnung also auch zurückwirken auf vorangehende Stufen: Lehrkräfte, die die Effektivität des Ignorierens erleben, werden dies eher als potentielle Handlungsstrategie in ihr deklaratives Wissen integrieren (vgl. Guskey, 1986). Dieser transaktionale Prozess führt dazu, dass sich die Kompetenz auf allen genannten Stufen immer weiter entwickelt bzw. verfeinert (Sameroff, 1995, vgl. dazu auch Ergebnisse der Expertenforschung, Berliner, 1994, 2001; Bromme, 1992). Zwar hält auch Schraw (2006) fest, dass die Stufen Akkumulation, Integration und Automatisierung immer wieder neu durchlaufen werden; Wechselwirkungen hat er jedoch nicht beschrieben—ein stärkerer Wissenszuwachs in der Interventionsgruppe wurde daher vorab nicht postuliert. Die Reaktionen auf das Training waren konsistent positiv: Die Akzeptanz des Trainings war hoch. Ein Vergleich der subjektiven Bewertung von Reflexionsmöglichkeiten und Transfer der Trainingsinhalte auf die eigene Praxis über die Module hinweg verdeutlicht, dass die implementierten Lerngelegenheiten unterschiedlich bewertet wurden: Die Lehrkräfte berichteten ein jeweils ansteigendes Ausmaß an Reflexion und Transfer über die Module hinweg; das Ausmaß der Reflexion erhöhte sich am deutlichsten zwischen Modul 1 und 2, das Ausmaß des Transfers erhöhte sich am deutlichsten zwischen Modul 2 und 3. Dies bestätigt die theoretisch angenommenen Ziele der jeweiligen Lerngelegenheiten, nach denen bereits in Simulationsübungen (Modul 2) reflection on action stattfinden kann, dass aber erst die Auseinandersetzung mit der realen Unterrichtssituation im Videozirkel die Praxistauglichkeit der neuen Handlungsstrategien überprüft bzw. erweist. Dennoch kann daraus nicht der Schluss gezogen werden, dass die implementierten Videozirkel den anderen Lerngelegenheiten per se überlegen sind—vielmehr wird postuliert, dass das spezielle Lehr-Lernarrangement, also die Komposition der verschiedenen Lerngelegenheiten und der sukzessive Aufbau von Reflexions-, Beurteilungs- und Feedbackkompetenzen, den Erfolg des entwickelten Trainings wesentlich determiniert haben. Interessant wäre z. B. zu untersuchen, wie wirksam Videozirkel ohne entsprechende vorangehende Module sind bzw. sein können sowie über entsprechende Designs den relativen Wert einzelner Lerngelegenheiten zu bestimmen (vgl. Desimone, 2009; Slavin, 2008). Wichtig wäre es auch zu prüfen, ob die nachgewiesenen Verbesserungen bzw. Veränderungen nach dem KODEK-Training nachhaltig sind, also langfristig beibehalten werden können. Obwohl eine follow-up Erhebung im Studiendesign angelegt war, konnte diese aufgrund der geringen Beteiligung (14 Lehrkräfte) nicht in die statistischen Analysen einbezogen werden. Darüber hinaus steht noch der Nachweis aus, dass die gefundenen Effekte sich mit anderen Teilnehmerinnen und

Teilnehmern replizieren lassen–die Gruppen waren hinsichtlich ihrer Berufserfahrung nicht vergleichbar (die Kontrollgruppe hatte im Mittel fast neun Jahre weniger Berufserfahrung).

Das KODEK-Training wurde für fortgeschrittene Studierende angepasst und im Rahmen eines Seminars durchgeführt (vgl. Teilstudie 3). Dabei wurde Modul 1 sowie das Microteaching aus Modul 2 implementiert; inhaltlich wurde insbesondere der Umgang mit störendem bzw. problematischem Schülerverhalten fokussiert. Damit ist das Training spezifisch auf die nach Berliner (1994) beschriebenen Stadien von Novizen und möglicherweise ersten Fortgeschrittenen Anfängern zugeschnitten; es werden vor allem Wissensakkumulation und -integration adressiert (vgl. Schraw, 2006). Infolgedessen wurde hier ein anderes Evaluationsdesign als in Teilstudie 2 gewählt bzw. beschränkten sich die Evaluationskriterien auf die abhängige Variable Wissen im Klassenmanagement. Die Kontrollgruppe erhielt ein Alternativtraining, in dem bezogen auf Klassenmanagement lediglich die Kombination aus Vortrag und Videobeispielen aus Modul 1 enthalten war. Der zweite Teil des Trainings wurde nach der Postmessung durchgeführt. Obwohl in Teilstudie 3 keine inferenzstatistische Evaluation vorgenommen wurde, so zeigen die deskriptiven Statistiken, dass das Wissen im Vergleich von Semesterbeginn und -ende in beiden Gruppen zugenommen hat. Dieser Wissenszuwachs ist wiederum stärker ausgeprägt in der Interventionsgruppe–insbesondere bezüglich Störungsintervention und Regeln. Damit gehört in der Interventionsgruppe der Bereich Störungsintervention, über den das Wissen am Semesterbeginn am geringsten eingeschätzt wurde, zu den Strategien, die am Semesterende am höchsten beurteilt wurden. Die Ergebnisse der Prüfung des Response Shift mittels Messinvarianzprüfung weisen darauf hin, dass sich das Verständnis der Konstrukte in der Interventionsgruppe stark verändert hat (Rekonzeptualisierung). Dies kann ein Effekt der Intervention sein (es war beabsichtigt, die Konstrukte zu rekonzeptualisieren), sollte jedoch in einem entsprechenden Design mit einer unbehandelten Kontrollgruppe hypothesengeleitet überprüft werden. Das Design erlaubte ebenfalls keine Prüfung langfristiger Effekte. Insbesondere der Erfolg im Praktikum bzw. im Referendariat, auch bezüglich des Stressempfindens, könnte sich deutlich unterscheiden.

In beiden Interventionsstudien zeigten sich zum Teil unerwünschte Ergebnisse in den Kontrollgruppen, die Hinweise auf die Effekte der spezifischen Lerngelegenheiten geben. In der Kontrollgruppe aus Teilstudie 1 (Lehrkräfte) gab es Verschlechterungen in einigen Bereichen des Wissens, des Könnens und in der Schülerbeteiligung in Bezug auf den Vergleich der Mittelwerte. Auch wenn es sich dabei um keine statistisch abgesicherten Effekte handelt (die verwendeten Verfahren ANCOVA bzw. ANOVA mit Messwiederholung lassen

darüber keine Aussagen zu), so könnten diese Tendenzen darauf hinweisen, dass die reine Wissensakkumulation und -integration Lehrkräfte in ihrem praktischen Handeln zunächst verunsichern können. Der Rückgang im Wissen (Wissen wird selten “weniger”) deutet gleichzeitig auf messmethodische Probleme hin, die in Teilstudie 3 gesondert überprüft wurden. Demnach kann dem Rückgang im Wissen möglicherweise auch ein Response Shift zu Grunde liegen: Das Wissen ist nach dem Training damit nicht geringer, sondern wird anders bewertet–z. B. in dem die Grundgesamtheit des potentiell möglichen Wissens größer, der Anteil des eigenen Wissens folglich geringer bewertet wird (Rekalibrierung). Dass eine solche Verschiebung durchaus wahrscheinlich ist, zeigen die Ergebnisse aus Teilstudie 3: Sowohl in einem varianzanalytischen, manifesten Verfahren (Vergleich von konventionellem und retrospektivem Prätest) als auch eine latente, faktorenanalytische Prüfung (Vergleich von konventionellem Prätest und Posttest) wurde Response Shift identifiziert–durch eine veränderte interne Metrik (Rekalibrierung) oder sogar eine veränderte Auffassung des Konstrukts (Rekonzeptualisierung). Auch in anderen Studien wurde wiederholt nachgewiesen, dass Probanden Konstrukte retrospektiv anders bewerten–in der Regel besser (im Sinne einer hohen Merkmalsausprägung) als zur echten Prämessung (Cantrell, 2003; Hoogstraten, 1985; Moore & Tananis, 2009). In dieser Studie zeigten sich auf Basis der varianzanalytischen Prüfung jedoch drei *negative* Effekte: Nach dem Training schätzen die Teilnehmerinnen und Teilnehmer der Kontrollgruppe ihr Wissen zum Zeitpunkt vor dem Training in drei Bereichen deutlich höher ein als vor dem Training in der konventionellen Prämessung. Die Teilnahme an der Intervention hat die Studierenden also möglicherweise dazu veranlasst, Teile ihrer Wissensbasis retrospektiv „nach oben“ zu korrigieren. Es könnte also–genau umgekehrt–das potentiell verfügbare Wissen nach dem Kontrolltraining geringer, das eigene Wissen damit höher bewertet werden. Mit Blick auf das Phänomen des sogenannten “Praxischocks” (Lipowsky, 2003; Müller-Fohrbrodt, Cloetta & Dann, 1978, vgl. auch Woolfolk Hoy & Burke Spero, 2005) sind diese Ergebnisse nicht nur methodisch problematisch, sondern ebenso inhaltlich relevant, da die angehenden Lehrkräfte möglicherweise mit falschen oder ungünstigen Vorstellungen in den Beruf gehen. Auch wenn es sich bei den dargestellten Befunden bezüglich der jeweiligen Kontrollgruppen um relativ kurzfristige Effekte handelt, die sich möglicherweise mit der Zeit neutralisieren, so weisen die Ergebnisse auch darauf hin, dass nicht intendierte Nebeneffekte möglich sind, die die ethische Legitimierbarkeit von Interventionsstudien betreffen (vgl. Mittag & Hager, 2000) und in der Konstruktion solcher Studien berücksichtigt werden sollten.

5.1.4. Wie kann wissenschaftlich fundiert überprüft werden, ob Ergebnisse von Lehreraus- und -weiterbildung auf implementierte Interventionen zurückzuführen sind?

Nach den Ergebnissen einer Metaanalyse von 319 Metaanalysen zur Effektivität von Interventionen gehen lediglich 25% der Varianz der beobachteten Trainingseffekte auf Merkmale der Intervention zurück (z. B. Art der Intervention, Dauer und Intensität)–ungefähr der gleiche Anteil wird jeweils der Zusammensetzung der Stichprobe sowie den implementierten Methoden (Design, Art der Vergleichsgruppe, Stichprobengröße, Operationalisierung der abhängigen Variablen) zugewiesen (Wilson & Lipsey, 2001). Der (valide) Nachweis der Wirksamkeit eines Trainings beinhaltet daher neben dessen sorgfältiger Konzeption ein ebenso gut durchdachtes Evaluationsdesign, um die Einflüsse anderer, interventionsunspezifischer Merkmale so gering wie möglich zu halten. Diesen Nachweis zu erbringen ist nicht trivial. Bedeutsam sind unter anderem ein den Fragestellungen entsprechender Versuchsplan, zuverlässige Instrumente und valide Verfahren zur statistischen Überprüfung der erhobenen Daten.

In Teilstudie 3 wurde ein Aspekt herausgegriffen und empirisch untersucht, der die interne Validität von Evaluationsstudien in besonderem Maße einschränkt: Das Phänomen des Response Shift. Response Shift ist dann wahrscheinlich, wenn Selbsteinschätzverfahren zur wiederholten Messung eingesetzt werden. Während Response Shift zunächst als allgemeine Veränderung der internalen Metrik aufgefasst wurde (Howard & Dailey, 1979), so spezifizierten Sprangers & Schwartz (1999) Verschiebungen auf drei mögliche Weisen: Ein veränderter Beurteilungsmaßstab (Rekalibrierung), eine veränderte Bedeutung von einzelnen Merkmalen des Konstrukts (Repriorisierung) sowie eine veränderte Bedeutung des Konstrukts an sich (Neudefinition bzw. Rekonzeptualisierung). Diese veränderten Bedeutungen lassen sich mit Hilfe von Verfahren zur Überprüfung der Messinvarianz prüfen, wurden bislang jedoch fast ausschließlich auf das (formative) Konstrukt des gesundheitlichen Wohlbefindens angewendet (Ahmed, Mayo, Corbiere, Wood-Dauphinee, Hanley & Cohen, 2005; Gandhi, Ried, Huang, Kimberlin & Kauf, 2013; Oort, 2005). Ein globales Maß zur Identifikation von Response Shift und (interdisziplinär) weit häufiger angewendet ist der Vergleich von konventioneller und retrospektiver Prämessung mittels Mittelwertvergleich (Hill & Betz, 2005; Sibthorp, Paisley, Gookin & Ward, 2007). Unter der–bislang weitestgehend ungeprüften–Prämisse, dass retrospektive Prämessung und Postmessung auf der gleichen internalen Metrik basieren, wird angenommen, dass der Vergleich dieser Messungen Aussagen über die Vergleichbarkeit von konventioneller Prämessung und Postmessung erlaubt. In der vorliegenden Studie identifizierten beide Verfahren (Mittelwertvergleich und Messinvarianzprüfung) Response Shift, die Prüfung mittels Messinvarianz identifizierte diesen jedoch insgesamt auf

mehr Skalen und auch gruppenspezifisch: In der Interventionsgruppe lag häufiger Response Shift vor und auch stärkerer (nach der Nomenklatur Sprangers' und Schwartz', 1999). Gleichzeitig wurde nachgewiesen, dass die implizite Annahme, retrospektiver Prämessung und Postmessung läge die gleiche Metrik zu Grunde, empirisch nicht bestätigt werden konnte: In beiden Gruppen wiesen drei von acht Skalen weniger als starke Invarianz auf. Zwar wird dies nicht zwingend als Response Shift interpretiert, belegt jedoch, dass die beiden Maße nicht per se messinvariant sind. Der retrospektive Prätest stellt sich damit nicht als uneingeschränkte Alternative zum konventionellen Prätest dar und sollte damit nur dann eingesetzt werden, wenn der Ausschluss von Prätesteffekten relevant ist oder wenn Veränderungen vom Verständnis des Konstrukts durch Teilnahme an einer Intervention wahrscheinlich sind (ggf. zusätzlich zum konventionellen Prätest). Dass jedoch auch der Einsatz des konventionellen Prätests nicht unproblematisch ist, indizieren die faktorenanalytischen Analysen zur Prüfung der Messinvarianz: In der Kontrollgruppe wurden zwei, in der Interventionsgruppe sogar sechs von acht Konstrukten rekonzeptualisiert. Ein Vergleich der Prä- und Postmessung ist damit nicht möglich, da offenbar ein anderes Konstrukt gemessen wurde. Das varianzanalytische Verfahren konnte diese Bedeutungsverschiebungen nicht aufdecken. Da Evaluation häufig auf den Einsatz von Selbstberichten angewiesen ist (aus forschungspragmatischen Gründen, jedoch häufig auch aus Mangel an valideren Alternativen) bleibt zu fragen, wie mit diesen Ergebnisse umgegangen werden kann. Stichproben sind nicht immer groß genug (vgl. Teilstudie 2), um potentielle Verzerrungen mittels ausgewählter Parameterrestriktionen ex post facto zu kontrollieren. Wichtig wäre daher, z. B. auf Basis eines Solomon-Vier-Gruppen-Plans zu prüfen: Gibt es unterschiedliche Effekte, je nach dem, ob auch ein Prätest beantwortet wurde? Welchen Einfluss hat die Teilnahme an einem Training per se? Auch der Einsatz von objektiven Tests oder die Durchführung von Interviews könnten aufschlussreich sein, um die hier aufgeworfenen Fragen tiefergehend zu untersuchen.

Während das Problem des Response Shift vor allem ein messmethodisches ist, wurden eingangs noch weitere Varianzquellen genannt, die das Ausmaß des gefundenen Effekts mit beeinflussen. So kamen in Teilstudie 1 Beobachterinnen und Beobachter bzw. Schülerinnen und Schüler–beide bewerteten das Können von Lehrkräften–zu leicht unterschiedlichen Beurteilungen und es zeigten sich in Teilstudie 2, je nach eingesetzter Methode, auch unterschiedliche Veränderungen. Obwohl der Einsatz multipler Instrumente wünschenswert ist und den genannten Methoden eine spezifische Validität zugesprochen wird, bleibt doch zu fragen, wie mit divergierenden Ergebnissen umzugehen ist. Häufig besteht die Gefahr im „fishing for significance“–also je nach Eignung bestimmte Ergebnisse herauszugreifen oder überhaupt zu

berichten. Einer hochwertigen Evaluation eines Trainings ist damit jedoch nicht gedient, da es möglicherweise trotz Bedarf nicht optimiert wird.

5.2. Grenzen und Stärken der Arbeit

Während die Einschränkungen der durchgeführten Studien im vorangegangenen Kapitel bereits stellenweise ausgeführt wurden, werden abschließend die zentralen Limitationen zusammengefasst.

5.2.1. Grenzen

Auswahl der Stichproben. Obwohl die Instrumente KODEK-B und KODEK-S in mehreren Pilotierungsschritten entwickelt wurden, so erfolgte die Prüfung der Endformen der Verfahren nicht an einer Zufalls-, sondern an einer Gelegenheitsstichprobe. Gerade die Analyse von Zusammenhängen mit externen Kriterien, insbesondere in Bezug auf die berufliche Belastung, könnten damit schlechte Schätzer für Populationsmaße darstellen, aber auch potentiell bedeutsame Moderatoren (z. B. systematische Einflüsse der Bewertungen aufgrund des Schulfachs oder der Uhrzeit am Tag) konnten nicht untersucht werden. Auch die Stichprobe in Teilstudie 2 schränkt die externe und interne Validität der Untersuchung ein, da sich die Gruppen in zentralen Ausgangsbedingungen unterschieden (Berufserfahrung), ohne dass dies statistisch kontrolliert werden konnte (aufgrund des geringen Stichprobenumfangs).

Implementiertes Design. Hinsichtlich der Entwicklung von Kompetenzen im Klassenmanagement können die hier durchgeführten Studien lediglich Aussagen über relativ kurzfristige Entwicklungen treffen. Sowohl in der Stichprobe der Lehrkräfte, als auch in der Stichprobe der Studierenden ist aber von zentraler Bedeutung, ob es den Personen gelingt, ihr neu gewonnenes Wissen (und Können) beizubehalten und in ihrer beruflichen Praxis zu implementieren. Darüber hinaus konnte die Frage der Konfundierung von Response Shift und Trainingseffekten in Teilstudie 3 mit dem implementierten Design nicht hinreichend geprüft werden, da beide Untersuchungsgruppen an einer Intervention zum Evaluationskriterium teilgenommen haben.

Verwendete Variablen. Die Validierung der Fragebögen KODEK-S und KODEK-B anhand von Außenkriterien beschränkte sich auf vier Variablen (off-task Verhalten, Schülerbeteiligung, Lehrerselbstwirksamkeit und berufliche Belastung). Das Wissen wurde im Rahmen der Evaluationsstudien über einen Selbstbericht evaluiert. Es handelt sich damit um ein subjektives Maß; Konfundierungen mit berufsbezogenen Selbstwirksamkeitsüberzeugungen sind denkbar und wahrscheinlich, konnten jedoch nicht kontrolliert werden.

5.2.2. Stärken

Zunächst wurde ein theoretisch zum Teil diffuses Konstrukt (Klassenmanagement) systematisiert und ein spezifisch auf die Operationalisierung zugeschnittenes Modell abgeleitet, das einen Anschluss an die internationale Forschung ermöglicht sowie explizit auf andere Konzeptualisierungen–auch aus dem deutschen Sprachraum–verweist und diese berücksichtigt. Auf Grundlage dieses Modells wurden, entgegen dem aktuellen Trend, in mehreren Pilotierungsphasen Fragebogenverfahren entwickelt, die durch sorgfältige Konstruktion besondere methodische Schwächen dieser Verfahren berücksichtigen (u. a. perspektivenspezifische Itemgenerierung mit entsprechendem Inferenzgrad, Itemverständlichkeit und -validität in KODEK-S, zwischen Aktivitätsstrukturen differenzierende Indikatoren in KODEK-B). Darüber hinaus wurde ein neu entwickeltes Training für ein hoch relevantes Thema entwickelt und auf Basis eines multikriterialen Evaluationsdesigns überprüft. Es konnten positive Effekte auf allen relevanten Ebenen identifiziert werden (Reaktion, Lernen, Verhalten, Ergebnisse, Kirkpatrick, 1998), sogar im Verhalten der Schülerinnen und Schüler (vgl. Lipowsky, 2010). Schließlich wurde ein an der realen Forschungspraxis orientierter, kritischer Blick auf die Validität von Prä-Post Designs gelegt der notwendig ist, wenn Selbsteinschätzungsinstrumente zum Einsatz kommen. Durch den Vergleich verschiedener Verfahren wurde ausgeleuchtet, dass retrospektive Prämessungen nicht unreflektiert an Stelle von konventionellen Prämessungen eingesetzt werden sollten.

5.3. Ausblick

5.3.1. Implikationen für zukünftige Forschung

In der vorliegenden Arbeit wurde an verschiedenen Stellen das Spannungsverhältnis zwischen methodisch anspruchsvollen Instrumenten, Designs und Auswertungsmethoden und den tatsächlich realisierbaren Forschungsvorhaben kritisch beleuchtet. Sowohl das hergeleitete Modell zur theoretischen Systematisierung von Klassenmanagement als auch die darauf basierenden Instrumente zur Erfassung von Kompetenzen im Klassenmanagement können dazu beitragen, systematische Forschung auf Basis eines multidimensionalen Verständnisses von Klassenmanagement in Deutschland anzuregen. Denn nicht nur interindividuell, sondern auch interkulturell können Unterschiede hinsichtlich des adäquaten Einsatzes verschiedener Klassenmanagementstrategien vorliegen, die auch für die Aus- und Weiterbildung berücksichtigt werden sollten (Lewis, Romi, Qui & Katz, 2005; van de Grift, 2007): So wird Schülerverhalten von deutschen Lehrkräften sehr wahrscheinlich in anderer Weise positiv verstärkt als z. B. in Israel oder China (Lewis et al., 2005). Ohne dass nun alle internationalen Ergeb-

nisse mit deutschen Stichproben repliziert werden müssten, so gibt es doch spezifische Strategien, die genauer untersucht werden sollten (z. B. Gruppenmobilisierung). Dafür ist es wichtig, die entwickelten Instrumente weiter zu optimieren sowie weitere Nachweise bezüglich der Validität zu erbringen, z. B. Zusammenhänge mit sozio-emotionalen Kompetenzen oder Prozessen der kognitiven Verarbeitung von Schülerinnen und Schülern, insbesondere aber auch hinsichtlich des uneindeutigen Zusammenhangs mit der beruflichen Belastung von Lehrkräften.

In Bezug auf die professionelle Entwicklung von Lehrkräften wäre interessant herauszufinden, welche Strategien relativ früh (z. B. Klarheit des Handlungsprogramm) und welche möglicherweise erst sehr spät beherrscht werden (z. B. Monitoring). Damit einher geht die Frage nach der optimalen Passung von Lerngelegenheiten und dem Kompetenzniveau von Lehrkräften. So könnte auf Basis der hier ermittelten Befunde systematisch überprüft werden, welche Lerngelegenheiten für die verschiedenen Entwicklungsstadien (z. B. Fortgeschrittene Anfänger vs. Lehrkräfte im Kompetenz-Stadium) geeignet sind. Zudem sollten die hier berichteten Befunde bezüglich der Wirksamkeit der Trainings in anderen Stichproben, mit längeren bzw. mehr Evaluationsintervallen und mit weiteren Evaluationskriterien überprüft werden. Dabei könnte gleichzeitig eine weitere Analyse des Response Shifts erfolgen.

5.3.2. Implikationen für die pädagogische Praxis

Die Implikationen für die pädagogische Praxis beziehen sich vorrangig auf die entwickelten Trainings sowie die Befunde der implementierten Evaluationsstudien. Es konnten zwei Trainings entwickelt werden, die sich als wirksam zur Entwicklung von Kompetenzen im Klassenmanagement erwiesen haben. Das KODEK-Training könnte in der Hinsicht erweitert werden, dass die Videozirkel so vorbereitet werden, dass sie anschließend im Rahmen sogenannter *booster-sessions* weitergeführt werden können; möglicherweise auch ohne externe Moderatoren (Fries & Souvignier, 2009). Neben der weiteren systematischen Evaluation ist die Implementation dieser Trainings als fester Bestandteil in der Aus- und Weiterbildung von Lehrkräften der nächste sinnvolle Schritt (ebd.). Dafür müssen zunächst weitere Trainer geschult werden. Dies erfordert nicht nur eine hohe inhaltliche Qualifizierung in Bezug auf Klassenmanagement, sondern–für das Training für Lehrkräfte–auch Fähigkeiten des Coaching und der Moderation (vgl. Ophardt & Thiel, 2013). Darüber hinaus ist von zentraler Bedeutung, auch die institutionellen Rahmenbedingungen für die Teilnahme an solchen Fortbildungen zu schaffen. So sind es gerade extern durchgeführte und längerfristig angelegte Programme wie KODEK, die eine besonders hohe Wirksamkeit und Akzeptanz zeigen (Messner &

Reusser, 2000; Roelofs, Veenman & Raemaekers, 1994), Gleichzeitig tragen diese Merkmale dazu bei, dass sie von Lehrkräften seltener wahrgenommen werden, da dies für Lehrkräfte als zusätzliche Belastung zum Unterrichtsalltag wahrgenommen wird.

Schließlich können die entwickelten Instrumente KODEK-S und KODEK-B zur internen Unterrichtsevaluation eingesetzt werden und im Rahmen von Team-Teaching zum Einsatz kommen. Dafür wurden auf der projektbezogenen Homepage (www.erwiss.fu-berlin.de/kodek) nicht nur die Instrumente hinterlegt, sondern auch Eingabemasken, die die Auswertung für interessierte Lehrkräfte erleichtern.

5.4. Literaturverzeichnis

- Ahmed, S., Mayo, N. E., Corbiere, M., Wood-Dauphinee, S., Hanley, J. & Cohen, R. (2005). Change in quality of life of people with stroke over time: True change or response shift? *Quality of Life Research*, 14(3), 611–627.
- Bergin, C. & Bergin, D. A. (1999). Classroom discipline that promotes self-control. *Journal of Applied Developmental Psychology*, 20(2), 189–206.
- Berliner, D. C. (1994). The wonder of exemplary performances. In J. N. Mangieri & C. C. Block (Hrsg.), *Creating powerful thinking in teachers and students. Diverse perspectives* (S. 1–37). Fort Worth, TX: Holt, Rinehart & Winston.
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35 (5), 463–482.
- Blömeke, S., Felbrich, A. & Müller, C. (2008). Messung des erziehungswissenschaftlichen Wissens angehender Lehrkräfte. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare: Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (S. 171–193). Münster: Waxmann.
- Briesch, A. M. & Chafouleas, S. M. (2009). Review and analysis of literature on self-management interventions to promote appropriate classroom behaviors (1988-2008). *School Psychology Quarterly*, 24(2), 106–118.
- Bromme, R. (1992). *Der Lehrer als Experte: Zur Psychologie des professionellen Wissens*. Bern: Huber
- Brophy, J. E. (1988). Educating teachers about managing classrooms and students. *Teaching and Teacher Education*, 4(1), 1–18.
- Brophy, J. E. (2006). History of research on classroom management. In C. M. Evertson & C. S. Weinstein (Hrsg.), *Handbook of classroom management. Research, practice and contemporary issues* (S. 17–43). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cantrell, P. (2003). Traditional vs. retrospective pretests for measuring science teaching efficacy beliefs in preservice teachers. *School Science and Mathematics*, 103(4), 177–185.
- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive?* Münster: Waxmann.
- Clunies-Ross, P., Little, E. & Kienhuis, M. (2008). Self-reported and actual use of proactive and reactive classroom management strategies and their relationship with teacher stress and student behaviour. *Educational Psychology*, 28, 693–710.

- Cramer, C. (2010). Kompetenzerwartungen Lehramtsstudierender: Grenzen und Perspektiven selbsteingeschätzter Kompetenzen in der Lehrerbildungsforschung. In A. Gehrman, U. Hericks & M. Lüders. (Hrsg.), *Bildungsstandards und Kompetenzmodelle. Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht* (S. 85–97). Bad Heilbrunn: Klinkhardt.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181–199.
- Doyle, W. (2006). Ecological approaches to classroom management. In C. M. Evertson & C. S. Weinstein (Hrsg.), *Handbook of classroom management. Research, practice and contemporary issues* (S. 97–125). Mahwah, NJ: Lawrence Erlbaum Associates.
- Eichhorn, C. (2008). *Classroom-Management: Wie Lehrer, Eltern und Schüler guten Unterricht gestalten*. Stuttgart: Klett-Cotta.
- Elliot, S. & Busse, R. T. (2004). Assessment and evaluation of students' behavior and intervention outcomes. The utility of rating scale methods. In R. B. Rutherford (Hrsg.), *Handbook of research in emotional and behavioral disorders* (S. 123–142). New York, NY: Guilford Press.
- Emmer, E. T., Evertson, C. M. & Worsham, M. E. (2003). *Classroom management for secondary teachers* (6. Aufl.). Boston: Allyn and Bacon.
- Evertson, C. M. & Harris, A. H. (1992). What we know about managing classrooms. *Educational Leadership*, 49(7), 74–78.
- Evertson, C. M. & Weinstein, C. S. (Hrsg.). (2006). *Handbook of classroom management: Research, practice and contemporary issues*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fifer, F. L. (1986). Effective classroom management. *Academic Therapy*, 21, 401–410.
- Freiberg, H. J. (1999). Beyond behaviorism. In H. J. Freiberg & J. E. Brophy (Hrsg.), *Beyond behaviorism: Changing the classroom management paradigm* (S. 3–20). Boston: Allyn and Bacon.
- Freiberg, H. J. & Lapointe, J. M. (2006). Research-based programs for preventing and solving discipline problems. In C. M. Evertson & C. S. Weinstein (Hrsg.), *Handbook of classroom management. Research, practice and contemporary issues* (S. 735–786). Mahwah, NJ: Lawrence Erlbaum Associates.

- Friedman, I. A. (2006). Classroom management and teacher stress and burnout. In C. M. Evertson & C. S. Weinstein (Hrsg.), *Handbook of classroom management* (S. 925–945). Mahwah, NJ: Lawrence Erlbaum Associates.
- Fries, S. & Souvignier, E. (2009). Training. In E. Wild & K. Möller (Hrsg.), *Pädagogische Psychologie* (S. 405–428). Heidelberg: Springer.
- Gärtner, H. (2010). Wie Schülerinnen und Schüler ihre Lernumwelt wahrnehmen: Ein Vergleich verschiedener Maße zur Übereinstimmung von Schülerwahrnehmungen. *Zeitschrift für Pädagogische Psychologie*, 24(2), 111–122.
- Gandhi, P. K., Ried, L. D., Huang, I.-C., Kimberlin, C. L. & Kauf, T. L. (2013). Assessment of response shift using two structural equation modeling techniques. *Quality of Life Research*, 22(3), 461–471.
- Gold, B., Förster, S. & Holodynski, M. (2013). Evaluation eines videobasierten Trainingsseminars zur Förderung der professionellen Wahrnehmung von Klassenführung im Grundschulunterricht. *Zeitschrift für Pädagogische Psychologie*, 27(3), 141–155.
- Gold, B. & Holodynski, M. (2011). Klassenführung. In E. Kiel & K. Zierer (Hrsg.), *Basiswissen Unterrichtsgestaltung: Unterrichtsgestaltung als Gegenstand der Praxis* (S. 133–151). Hohengehren: Schneider.
- Gottfredson, D. C., Gottfredson, G. D. & Hybl, L. G. (1993). Managing adolescent behavior a multiyear, multischool study. *American Educational Research Journal*, 30(1), 179–215.
- Guskey, T.R. (1986). Staff development and the process of teacher change. *Educational Researcher*, 15, 5–12.
- Hattie, J. (2009). *Visible learning: A synthesis of meta-analyses relating to achievement*. London: Routledge.
- Havers, N. (2010). Lässt sich effiziente Klassenführung lehren? Das Potential des Lehrert Trainings. In J. Abel & G. Faust (Hrsg.), *Wirkt Lehrerbildung? Antworten aus der empirischen Forschung* (1. Aufl., S. 283–290). Münster: Waxmann.
- Helmke, A. (2007). Aktive Lernzeit optimieren: Was wissen wir über effiziente Klassenführung? *Pädagogik*, 59(5), 44–49.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität*. Seelze-Velber: Klett Kallmeyer.
- Helmke, A., Schneider, W. & Weinert, F. E. (1986). Quality of instruction and classroom learning outcomes: The German contribution to the IEA classroom environment study. *Teaching and Teacher Education*, 2(1), 1–18.

- Hill, L. G. & Betz, D. L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation*, 26(4), 501–517.
- Hoogstraten, J. (1985). Influence of objective measures on self-reports in a retrospective pretest-posttest design. *Journal of Experimental Education*, 53(4), 207–210.
- Howard, G. S. & Dailey, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology*, 64(2), 144–150.
- Jones, V. F. (2006). How do teachers learn to be effective classroom managers? In C. M. Evertson & C. S. Weinstein (Hrsg.), *Handbook of classroom management. Research, practice and contemporary issues* (S. 887–907). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kaufhold, M. (2006). *Kompetenz und Kompetenzerfassung: Analyse und Beurteilung von Verfahren der Kompetenzerfassung*. Wiesbaden: Verlag für Sozialwissenschaften.
- Kirkpatrick, D. L. (1998). *Evaluating training programs. The four levels* (2. Aufl.). San Francisco, CA: Berrett-Koehler.
- Klieme, E. & Hartig, J. (2007). Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. *Zeitschrift für Erziehungswissenschaft*, 10(8), 11–29.
- Klusmann, U., Kunter, M., Voss, T. & Baumert, J. (2012). Berufliche Beanspruchung angehender Lehrkräfte: Die Effekte von Persönlichkeit, pädagogischer Vorerfahrung und professioneller Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 26, 275–290.
- König, J. & Blömeke, S. (2009). Pädagogisches Wissen von angehenden Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 12(3), 499–527.
- Kounin, J. S. (1976). *Techniken der Klassenführung*. Bern: Huber.
- Kunter, M. & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, 9, 231–251.
- Kunter, M., Baumert, J. & Köller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction*, 17(5), 494–509.
- Kunter, M. & Voss, T. (2011). Das Modell der Unterrichtsqualität in COACTIV: Eine multi-kriteriale Analyse. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 85–113). Münster: Waxmann.
- Lewis, R., Romi, S., Qui, X. & Katz, Y. J. (2005). Teachers' classroom discipline and student misbehavior in Australia, China and Israel. *Teaching and Teacher Education*, 21(6), 729–741.

- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz.
- Lipowsky, F. (2003). *Wege von der Hochschule in den Beruf. Eine empirische Studie zum beruflichen Erfolg von Lehramtsabsolventen in der Berufseinstiegsphase*. Bad Heilbrunn: Klinkhardt.
- Lipowsky, F. (2010). Lernen im Beruf. Empirische Befunde zur Wirksamkeit von Lehrerfortbildung. In F. H. Müller (Hrsg.), *Lehrerinnen und Lehrer Lernen* (S. 51–70). Münster: Waxmann.
- Lotz, M. & Gabriel, K. (2012, März). *Niedrig und hoch inferente Verfahren zur Erfassung von Unterrichtsqualität in der Grundschule*. Vortrag im Rahmen des Symposiums „Was konstituiert Unterricht? Grenzgänge zwischen quantitativer und qualitativer Unterrichtsforschung“ (W. Helsper & E. Klieme) auf dem 23. Kongress der Deutschen Gesellschaft für Erziehungswissenschaft, Osnabrück, Deutschland.
- Lotz, M., Berner, N. E., Gabriel, K., Post, S., Faust, G. & Lipowsky, F. (2011). Unterrichtsbeobachtung im Projekt PERLE. In D. Kucharz, T. Irion & B. Reinoffer (Hrsg.), *Grundlegende Bildung ohne Brüche* (S. 183–194). Wiesbaden: Verlag für Sozialwissenschaften.
- Lüttke, O., Trautwein U., Kunter M. & Baumert J. (2006). Analyse von Lernumwelten. Ansätze zur Bestimmung der Reliabilität und Übereinstimmung von Schülerwahrnehmungen. *Zeitschrift für Pädagogische Psychologie*, 20(1/2), 85–96.
- Marzano, R. J. (2003). *What works in schools*. Alexandria, VA: ASCD.
- Mayr, J. (2006). Klassenführung auf der Sekundarstufe II: Strategien und Muster erfolgreichen Lehrerhandelns. *Schweizerische Zeitschrift für Bildungswissenschaften*, 28(2), 227–242.
- Mayr, J., Eder, F. & Fartacek, W. (2005). *Linzer Diagnosebogen zur Klassenführung (LDK)*. Zugriff am 10.09.2013 von <http://ius.uni-klu.ac.at/projekte/ldk/>.
- Messner, H. & Reusser K. (2000). Die berufliche Entwicklung von Lehrpersonen als lebenslanger Prozess. *Beiträge zur Lehrerbildung*, 18(2), 157–171.
- Mittag, W. & Hager, W. (2000). Ein Rahmenkonzept zur Evaluation psychologischer Interventionsmaßnahmen. In W. Hager & J. L. Patry (Hrsg.), *Handbuch Evaluation psychologischer Interventionsmaßnahmen* (S. 102–128). Bern: Huber.
- Moore, D. & Tananis, C. A. (2009). Measuring change in a short-term educational program using a retrospective pretest design. *American Journal of Evaluation*, 30(2), 189–202.
- Müller-Fohrbrodt, G., Cloetta, B. & Dann, H.-D. (1978). *Der Praxisschock bei jungen Lehrern: Formen, Ursachen, Folgerungen*. Stuttgart: Klett.

- Muthén, B. O. & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. Hox & J. K. Roberts (Hrsg.), *Handbook of advanced multilevel analysis* (S.15-40). New York, NY: Routledge.
- Neuenschwander, M. P. (2006). Überprüfung einer Typologie der Klassenführung. *Schweizerische Zeitschrift für Bildungswissenschaften*, 28(2), 243–257.
- Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of Life Research*, 14(3), 587–598.
- Ophardt, D. & Thiel, F. (2007). Klassenmanagement als professionelle Gestaltungsleistung. In D. Lemmermöhle, M. Rothnagel, S. Bögeholz, M. Hasselhorn & R. Watermann (Hrsg.), *Professionell Lehren. Erfolgreich Lernen* (S. 133–145). Münster: Waxmann.
- Ophardt, D. & Thiel, F. (2008). Klassenmanagement als Basisdimension der Unterrichtsqualität. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungsperspektiven und methodische Zugänge* (2. Aufl., S. 259–284). Wiesbaden: Verlag für Sozialwissenschaften.
- Ophardt, D. & Thiel, F. (2013). *Klassenmanagement – Ein Handbuch für Studium und Praxis*. Stuttgart: Kohlhammer.
- Oser, F., Curcio, G.-P. & Düggeli, A. (2007). Kompetenzmessung in der Lehrerbildung als Notwendigkeit – Fragen und Zugänge. *Beiträge zur Lehrerbildung*, 25(1), 14–26.
- Petermann, F., Niebank, K. & Scheithauer, H. (2004). *Entwicklungswissenschaft. Entwicklungspsychologie - Genetik - Neuropsychologie*. Berlin: Springer.
- Richter, D., Kunter, M., Klusmann, U., Lüdtke, O. & Baumert, J. (2011). Professional development across the teaching career: Teachers' uptake of formal and informal learning opportunities. *Teaching and Teacher Education*, 27(1), 116–126.
- Roelofs, E., Veenman, S. & Raemaekers, J. (1994). Improving instruction and classroom management behaviour in mixed-age classrooms: Results of two improvement studies. *Educational Studies*, 20(1), 105–126.
- Rogers, B. & Kühn, M. (2013). *Classroom Management: Das Praxisbuch*. Weinheim: Beltz.
- Sameroff, A. J. (1995). General systems theories and developmental psychopathology. In D. Cicchetti & D. U. Cohen (Hrsg.), *Developmental psychopathology*. Cambridge: Cambridge University Press.
- Schön, D. A. (1983). *The reflective practitioner. How professionals think in action*. London: Temple Smith.
- Schönbächler, M.-T. (2004a). *Fragebogen zum Unterricht auf der Primarstufe. Fragebogen für Schülerinnen und Schüler*. Bern: Institut für Pädagogik und Schulpädagogik.

- Schönbächler, M.-T. (2004b). *Klassenmanagement auf der Primarstufe. Fragebogen für Lehrpersonen*. Bern: Institut für Pädagogik und Schulpädagogik
- Schönbächler, M.-T. (2006). Inhalte von Regeln und Klassenmanagement. *Schweizerische Zeitschrift für Bildungswissenschaften*, 28(2), 259–271.
- Schraw, G. (2006). Knowledge: Structures and processes. In P. Alexander & P. Winne (Hrsg.), *Handbook of educational psychology* (2. Aufl., S. 245–264). San Diego, CA: Academic Press.
- Sibthorp, J., Paisley, K., Gookin, J. & Ward, P. (2007). Addressing response-shift bias: Retrospective pretests in recreation research and evaluation. *Journal of Leisure Research*, 39(2), 295–315.
- Simons, P. R. J. & Ruijters, M. C. P. (2004). Learning professionals: Towards an integrated model. In H. P. A. Boshuizen, R. Bromme & H. Gruber (Hrsg.), *Professional learning: Gaps and transitions on the way from novice to expert* (S. 207-229). Dordrecht: Kluwer.
- Simonsen, B., Fairbanks, S., Briesch, A., Myers, D. & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education and Treatment of Children*, 31(3), 351–380.
- Slavin, R. E. (2008). Perspectives on evidence-based research in education. What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 37(1), 5–14.
- Sprangers, M. A. G. & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: A theoretical model. *Social Science & Medicine*, 48(11), 1507–1515.
- Thiel, F., Ophardt, D. & Piwowar, V. (2013). Abschlussbericht des Projekts „Kompetenzen des Klassenmanagements (KODEK). Entwicklung und Evaluation eines Fortbildungsprogramms für Lehrkräfte zum Klassenmanagement.“ [Report of the research project „Classroom management competencies. Development and evaluation of a teacher training programme in classroom management.“]. Retrieved from http://www.ewi-psy.fu-berlin.de/einrichtungen/arbeitsbereiche/schulentwicklungsforschung/downloads/Lehramtsmasterbefragung_2009.pdf?1310986750
- Thiel, F., Richter, S. G. & Ophardt, D. (2012). Steuerung von Übergängen im Unterricht. Eine Experten-Novizen-Studie zum Klassenmanagement. *Zeitschrift für Erziehungswissenschaft*, 4, 727–752.
- Toman, H. (2007). *Classroom-Management. Praxishilfen für das Classroom-Management* (Basiswissen Grundschule, Bd. 25). Baltmannsweiler: Schneider.

- Van de Grift, W. (2007). Quality of teaching in four European countries: A review of the literature and application of an assessment instrument. *Educational Research*, 49(2), 127–152.
- Voss, T., Kunter, M. & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, 103(4), 952–969.
- Wang, M. C., Haertel, G. D. & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63(3), 249–294.
- Weinstein, C. S. (1999). Reflections on best practices and promising programs: Beyond assertive classroom discipline. In H. J. Freiberg & J. E. Brophy (Hrsg.), *Beyond behaviorism. Changing the classroom management paradigm* (S. 147–163). Boston: Allyn and Bacon.
- Weinstein, C. S. (2007). *Middle and secondary classroom management: Lessons from research and practice*. Boston: Mc Graw Hill.
- Wilson, D. B. & Lipsey, M. W. (2001). The role of method in treatment effectiveness research: Evidence from meta-analysis. *Psychological Methods*, 6(4), 413–429.
- Woolfolk, A., Hoy, A. W. & Burke Spero, R. (2005). Changes in teacher efficiency during the early years of teaching. *Teaching and Teacher Education*, 21(4), 343–356.

Danksagung

Die vorliegende Arbeit entstand während meiner Arbeit als wissenschaftliche Mitarbeiterin im Rahmen der Projekte KODEK und SPRINT am Arbeitsbereich Schulpädagogik/Entwicklungsforschung unter der Leitung von Felicitas Thiel und Diemut Ophardt und durch die Finanzierung des Bundesministeriums für Bildung und Forschung. Während des Arbeitsprozesses habe ich in ganz unterschiedlicher Weise Unterstützung von verschiedenen Menschen erfahren.

Allen voran möchte ich Felicitas Thiel danken, die mir stets Vertrauen entgegen gebracht hat und mich während der ganzen Zeit nicht nur wissenschaftlich, sondern auch persönlich begleitet, unterstützt und inspiriert hat. Ihre Rückmeldungen, Kritikpunkte und Anregungen trugen maßgeblich zum Gelingen der Arbeit bei und ich danke ihr dafür von Herzen.

Dank gilt auch meinen Kolleginnen und Kollegen am Arbeits- bzw. Fachbereich, die mich in dieser Zeit begleitet haben, insbesondere Diemut Ophardt, die das Projekt, in dem die Arbeit entanden ist, mit geleitet und zur Realisierung dieser Arbeit durch ihr Engagement, Interesse und ihre Unterstützung beigetragen hat. Außerdem danke ich Irmela Blüthmann, Anja Böhnke, Katharina Krysmanski und Sabine Richter für ihre hilfreichen Rückmeldungen und den anregenden Austausch. Franziska Pfitzner-Eden hat nicht nur mit größter Sorgfalt und Hingabe die englischen Publikationen korrigiert, sondern darüber hinaus wertvolles und stets konstruktives Feedback gegeben, wofür ich ihr sehr dankbar bin.

Prof. Bettina Hannover danke ich für ihr Interesse an dieser Arbeit sowie ihre Bereitschaft, diese zu begutachten.

Ein herzlicher Dank geht auch an die vielen studentischen Hilfskräfte, Praktikantinnen, TestleiterInnen und RaterInnen, ohne die die Realisierung dieser Arbeit nicht möglich gewesen wäre sowie an die Studierenden, Lehrkräfte, Schülerinnen und Schüler, Seminarleiter und Referandare, die sich an den empirischen Untersuchungen beteiligt haben. Persönlich bedanken möchte ich mich an dieser Stelle bei Gunnar Beyer, Belinda Anderschitz (geb. Frank), Gönke Bitter (geb. Kühl), Sina Gellermann, René Morgenstern, Nicole Persson, Christin Reimann, Lubina Rycer, Marco Splitt, Hannah Wallis und Lili Williams.

Abschließend möchte ich meiner Familie und Matthias Tenten danken, die mich während der ganzen Zeit mit viel Humor, Geduld und Liebe begleitet und motiviert haben.

Erklärung

Hiermit versichere ich, die Dissertationsarbeit „Konzeptualisierung, Erfassung und Entwicklung von Kompetenzen im Klassenmanagement“ selbstständig angefertigt zu haben. Sämtliche Hilfsmittel, die ich verwendet habe, sind angegeben. Sie ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden.

Berlin, im September 2013

Lebenslauf

Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten.