

Gabriel, Katrin

## Videobasierte Erfassung von Unterrichtsqualität im Anfangsunterricht der Grundschule. Klassenführung und Unterrichtsklima in Deutsch und Mathematik

Kassel : kassel university press 2014, 333 S. - (Dissertation, Universität Kassel, 2013)



Quellenangabe/ Reference:

Gabriel, Katrin: Videobasierte Erfassung von Unterrichtsqualität im Anfangsunterricht der Grundschule. Klassenführung und Unterrichtsklima in Deutsch und Mathematik. Kassel : kassel university press 2014, 333 S. - (Dissertation, Universität Kassel, 2013) - URN: urn:nbn:de:0111-pedocs-306501 - DOI: 10.25656/01:30650

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-306501>

<https://doi.org/10.25656/01:30650>

in Kooperation mit / in cooperation with:



<http://kup.uni-kassel.de>

### Nutzungsbedingungen

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

### Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

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

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

### Kontakt / Contact:

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

Mitglied der

  
Leibniz-Gemeinschaft

KATRIN GABRIEL

# VIDEOBASIERTE ERFASSUNG VON UNTERRICHTSQUALITÄT IM ANFANGSUNTERRICHT DER GRUNDSCHULE

KLASSENFÜHRUNG UND UNTERRICHTSKLIMA  
IN DEUTSCH UND MATHEMATIK



Katrin Gabriel

# **Videobasierte Erfassung von Unterrichtsqualität im Anfangsunterricht der Grundschule**

Klassenführung und Unterrichtsklima  
in Deutsch und Mathematik

Die vorliegende Arbeit wurde vom Fachbereich Humanwissenschaften der Universität Kassel als Dissertation zur Erlangung des akademischen Grades einer Doktorin der Philosophie (Dr. phil.) angenommen.

Erster Gutachter: Prof. Dr. Frank Lipowsky

Zweite Gutachterin: Prof. Dr. Diemut Kucharz

Tag der mündlichen Prüfung

19. Juli 2013

Bibliografische Information der Deutschen Nationalbibliothek  
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar

Zugl.: Kassel, Univ., Diss. 2013  
ISBN 978-3-86219-694-4 (print)  
ISBN 978-3-86219-695-1 (e-book)  
URN: <http://nbn-resolving.de/urn:nbn:de:0002-36957>

© 2014, kassel university press GmbH, Kassel  
[www.upress.uni-kassel.de/](http://www.upress.uni-kassel.de/)

Druck und Verarbeitung: Print Managment Logistics Solutions, Kassel  
Printed in Germany

# Inhalt

I	EINLEITUNG .....	7
II	THEORETISCHER UND EMPIRISCHER HINTERGRUND .....	13
<b>1</b>	<b>Modellierung von Unterrichtsqualität .....</b>	<b>13</b>
1.1	Ältere und neuere Konzepte der Unterrichtsqualität .....	15
1.2	Das Konzept der drei Basisdimensionen der Unterrichtsqualität ..	19
1.2.1	Weiterentwicklungen des Konzepts der drei Basisdimensionen ...	21
1.2.2	Die Wechselwirkung der Klassenführung und des Unterrichtsklimas .....	24
1.2.3	Ist das Konzept auf alle Fächer sowie auf die Grundschule übertragbar? .....	26
<b>2</b>	<b>Unterrichtsqualität in der Grundschule .....</b>	<b>28</b>
2.1	Studien und Konzepte von Unterrichtsqualität in der Grundschule .....	29
2.2	Videobasierte Studien zur Unterrichtsqualität in der Grundschule .....	40
2.3	Befunde zur Ausprägung der Unterrichtsqualität im Anfangsunterricht .....	47
<b>3</b>	<b>Klassenführung als Basisdimension der Unterrichtsqualität .....</b>	<b>50</b>
3.1	Klassenführung – eine Begriffsklärung .....	50
3.2	Merkmale der Klassenführung in der Grundschule .....	54
3.2.1	Zeitmanagement / effiziente Zeitnutzung .....	54
3.2.2	Übergangmanagement (Managing Transition) .....	57
3.2.3	Regelklarheit bzw. -verwendung .....	59
3.2.4	Störungsfreiheit / effektiver Umgang mit auftretenden Unterrichtsstörungen .....	62
3.2.5	Vorbereitung des Klassenraums .....	65
<b>4</b>	<b>Unterrichtsklima als Basisdimension der Unterrichtsqualität .....</b>	<b>66</b>
4.1	Der Klimabegriff – eine langjährige Debatte .....	67
4.1.1	Unterstützendes Unterrichtsklima versus Schülerorientierung .....	69
4.1.2	Verwendung des Klimabegriffs in der vorliegenden Arbeit .....	71

4		Inhalt
4.2	Merkmale des Unterrichtsklimas in der Grundschule .....	72
4.2.1	Gegenseitige Wertschätzung / Anerkennung .....	75
4.2.2	Positiver Umgang mit Fehlern im Unterricht / positive Fehlerkultur .....	76
4.2.3	Lehrerfürsorglichkeit als Merkmal einer positiven Lehrer-Schüler-Beziehung .....	78
4.2.4	Positiv unterstützende (Lob / Ermutigung) und sachlich-konstruktive Rückmeldungen .....	81
4.2.5	Humorvolle Lernatmosphäre als Aspekt des Unterrichtsklimas ...	85
4.2.6	Wettbewerbs- und Leistungsdruck als Merkmal des Unterrichtsklimas .....	86
<b>5</b>	<b>Unterrichtsqualitätsmerkmale und multikriteriale Effekte ....</b>	<b>88</b>
5.1	Klassenführung und multikriteriale Effekte .....	93
5.1.1	Klassenführung und Leistung .....	93
5.1.2	Klassenführung und Selbstkonzept .....	99
5.2	Unterrichtsklima und multikriteriale Effekte .....	100
5.2.1	Unterrichtsklima und Leistungen .....	101
5.2.2	Unterrichtsklima und Selbstkonzept .....	104
<b>6</b>	<b>Die Erfassung von Unterrichtsmerkmalen durch Videoanalyse .....</b>	<b>111</b>
6.1	Videoanalysen und ihre Herausforderungen in der Grundschule ..	111
6.2	Hoch inferente Ratings zur Beurteilung der Unterrichtsqualität ...	113
6.2.1	Hoch inferente Ratings versus niedrig inferente Kodierungen ....	113
6.2.2	Bestandteile eines Ratingverfahrens nach Langer & Schulz von Thun (2007) .....	115
6.3	Gütekriterien zur Qualitätskontrolle bei Videoanalysen .....	117
6.3.1	Mögliche Urteilsfehler bei dem Einsatz des Ratingsystems .....	119
6.3.2	Generalisierbarkeitsstudien zur Berechnung der Inter-Rater-Reliabilität .....	120
<b>7</b>	<b>Forschungsfragen der Arbeit .....</b>	<b>123</b>
7.1	Inter-Rater-Reliabilität .....	123
7.2	Ausprägung der Unterrichtsqualitätsmerkmale (Studie 1) .....	126
7.3	Struktur der Unterrichtsqualitätsmerkmale (Studie 2) .....	128
7.4	Effekte der Unterrichtsqualitätsmerkmale auf multikriteriale Schülermerkmale (Studie 3) .....	130
<b>III</b>	<b>METHODISCHES VORGEHEN .....</b>	<b>135</b>
<b>8</b>	<b>Kontext der Untersuchung – Die PERLE-Studie .....</b>	<b>135</b>
<b>9</b>	<b>Die PERLE-Videostudien Deutsch und Mathematik .....</b>	<b>138</b>

9.1	Die Videostudie Deutsch – Vorgaben und Ablauf .....	138
9.2	Die Videostudie Mathematik – Vorgaben und Ablauf .....	140
9.3	Bedeutung der Vorgaben für den Unterrichtsverlauf .....	142
<b>10</b>	<b>Das Vorgehen bei der Videoerhebung .....</b>	<b>143</b>
<b>11</b>	<b>Stichprobenbeschreibung .....</b>	<b>145</b>
11.1	Die Lehrerstichprobe .....	147
11.2	Die Schülerstichprobe .....	150
11.3	Repräsentativität des gezeigten Verhaltens in den Videostudien ..	151
11.4	Hierarchische Datenstruktur .....	155
11.5	Umgang mit fehlenden Werten .....	157
<b>12</b>	<b>Operationalisierung und Erfassung der Unterrichtsqualitäts- und Schülermerkmale .....</b>	<b>158</b>
12.1	Das hoch inferente Ratingsystem .....	159
12.1.1	Die Entwicklung des hoch inferenten Ratingsystems .....	159
12.1.2	Der Aufbau des hoch inferenten Ratingsystems .....	162
12.1.3	Die Analyseeinheit .....	165
12.1.4	Die Ratingprozedur .....	166
12.2	Die Erhebung der Zielvariablen in Mathematik und Rechtschreiben .....	172
12.2.1	Leistungen in Mathematik .....	172
12.2.2	Leistungen im Rechtschreiben .....	173
12.2.3	Selbstkonzepte im Rechnen und Schreiben .....	175
<b>IV</b>	<b>ERGEBNISSE .....</b>	<b>179</b>
<b>13</b>	<b>Ergebnisse der Inter-Rater-Reliabilität .....</b>	<b>179</b>
13.1	Fachunspezifische Berechnung der Inter-Rater-Reliabilität .....	181
13.2	Fachspezifische Berechnung der Inter-Rater-Reliabilität .....	184
<b>14</b>	<b>Ergebnisse der Studie 1 – Ausprägung der Merkmale .....</b>	<b>188</b>
<b>15</b>	<b>Ergebnisse der Studie 2 – Überprüfung der Konstruktvalidität .....</b>	<b>192</b>
15.1	Konfirmatorische Faktorenanalysen (CFA) .....	196
15.1.1	Wahl der Schätzmethode .....	196
15.1.2	Theoretische Herleitung der Modelle .....	196
15.1.3	Ergebnis des Modellvergleichs .....	201
15.2	Fachspezifische explorative Faktorenanalyse (EFA) .....	204
15.3	Skalenbildung .....	210

<b>16</b>	<b>Ergebnisse der Studie 3 – Multikriteriale Effekte der Klassenführung und des Unterrichtsklimas .....</b>	<b>212</b>
16.1	Die Bedeutung der Unterrichtsmerkmale im Fach Deutsch (1. SJ) .....	218
16.2	Die Bedeutung der Unterrichtsmerkmale im Fach Mathematik (2. SJ) .....	220
V	ZUSAMMENFASSUNG UND DISKUSSION .....	225
<b>17</b>	<b>Inter-Rater-Reliabilität .....</b>	<b>226</b>
17.1	Ergebnisse der fachunspezifischen Inter-Rater-Reliabilität .....	226
17.2	Ergebnisse der fachspezifischen Inter-Rater-Reliabilität .....	228
17.3	Methodische Einschränkungen und weitere Forschungsfragen .....	234
<b>18</b>	<b>Teilstudien (Studie 1 bis 3) .....</b>	<b>238</b>
18.1	Ausprägung der Merkmale (Studie 1) .....	238
18.2	Überprüfung der Konstruktvalidität (Studie 2) .....	243
18.3	Multikriteriale Effekte der Klassenführung und des Unterrichtsklimas (Studie 3) .....	246
VI	FAZIT .....	257
VII	LITERATUR .....	259
VIII	ABBILDUNGSVERZEICHNIS .....	287
IX	TABELLENVERZEICHNIS .....	289
X	ANHANG .....	291
<b>19</b>	<b>Das hoch inferente Ratingsystem .....</b>	<b>291</b>
19.1	Merkmale der Klassenführung .....	291
19.2	Merkmale des Unterrichtsklimas .....	306
<b>20</b>	<b>Auswertungsbogen für die Rater .....</b>	<b>320</b>
<b>21</b>	<b>Deskriptive Statistiken .....</b>	<b>321</b>
<b>22</b>	<b>Ergebnis des Modellvergleichs – Darstellung der Modelle .....</b>	<b>325</b>
<b>23</b>	<b>Dokumentation der hoch inferenten Skalen .....</b>	<b>327</b>

# I Einleitung<sup>1</sup>

Mit den TIMSS<sup>2</sup>-Videostudien aus den Jahren 1995 und 1999 (vgl. Stigler, Gonzales, Kawanaka, Knoll & Serrano, 1999; Hiebert et al., 2003) wurde weltweit ein „Schub videobasierter Unterrichtsstudien ausgelöst“ (Pauli & Reusser, 2006, S. 775). Seitdem hat sich die Nutzung von Videodaten zur Untersuchung von Lehr-Lern-Prozessen als ein wichtiges Forschungsinstrument etabliert (vgl. Aufschnaiter & Welzel, 2001; Ulewicz & Beatty, 2001). Dies liegt unter anderem daran, dass Unterrichtsvideos die Möglichkeit bieten, verschiedene Merkmale des Unterrichts zeitgleich und unterrichtsnah zu erfassen, und somit der Unterricht hinsichtlich seiner Qualität sowie Effektivität untersucht werden kann.

Im Zuge der videobasierten Forschungstradition in Deutschland hat sich für die Sekundarstufe zur Erklärung der Wirksamkeit naturwissenschaftlichen und mathematischen Unterrichts das Modell der drei Basisdimensionen guten Unterrichts besonders bewährt. Als bedeutsam für den Lernerfolg erwiesen sich eine effektive Unterrichts- und Klassenführung, Schülerorientierung / Unterrichtsklima sowie kognitive Aktivierung (z. B. Klieme, Schümer & Knoll, 2001; Klieme, 2002). Diese drei Basisdimensionen scheinen besonders gut geeignet, um bisherige theoretische Konzepte der empirischen Unterrichtsforschung und deren Wirkungen zu integrieren (Klieme, Lipowsky, Rakoczy & Ratzka, 2006). Für sprachliche Fächer findet jedoch häufig eine Übertragung der Ergebnisse aus dem naturwissenschaftlichen und mathematischen Bereich statt, ohne dass überprüft wird, ob den drei Basisdimensionen dieselbe Bedeutung zukommt. Aus Gründen der Fachspezifität stellt sich in diesem Zusammenhang die Frage, ob das Konzept auf den Deutschunterricht grundsätzlich übertragbar ist. Klieme, Schümer und Knoll (2001) argumentieren in diesem Zusammenhang, dass im Prinzip jeder Unterricht die oben genannten drei Basisdimensionen erfüllen muss (mit unterschiedlicher Schwerpunktsetzung und Qualität), dies gilt demnach auch für den Anfangs-

- 
- 1 Die Dissertation wurde im Kontext der PERLE-Studie (**P**ersönlichkeits- und **L**ernentwicklung von Grundschulkindern) realisiert. Mein besonderer Dank gilt in diesem Zusammenhang meinem Doktorvater Prof. Dr. Frank Lipowsky sowie meinen Kolleginnen und Kollegen der Arbeitsgruppe an der Universität Kassel, die mir in den letzten Jahren nicht nur bei methodischen Fragen zur Seite standen. Frau Prof. Dr. Diemut Kucharz danke ich für die intensive Unterstützung bei der endgültigen Fertigstellung dieser Arbeit. Auch allen weiteren Unterstützern sei an dieser Stelle herzlich gedankt. Ein ganz besonderer Dank geht an meinen Freund Hardy, der gleichzeitig Unterstützer und Motivator war.
  - 2 TIMSS steht für **T**hird **I**nternational **M**athematics and **S**cience Study, seit 2003 für **T**rends in **I**nternational **M**athematics and **S**cience Study (vgl. Baumert et al., 1997; Bos, Bonsel, Baumert, Prenzel, Selter & Walther, 2008).

unterricht<sup>3</sup>. Für den Grundschulbereich mangelt es jedoch an großangelegten und curricular vergleichbaren Videostudien und somit an Erkenntnissen zur Ausprägung sowie Wirkung der Basisdimensionen der Unterrichtsqualität. An diesem Punkt setzt die PERLE-Studie an, die die **Persönlichkeits- und Lernentwicklung** von Grundschulern<sup>4</sup> über die gesamte Grundschulzeit in den Fächern Deutsch, Kunst und Mathematik untersucht (vgl. Greb, Faust & Lipowsky, 2007; Lipowsky, Faust & Greb, 2009). Im Rahmen der PERLE-Studie konnten prozessbezogene quantitative Daten mittels Videografie im Anfangsunterricht unter curricular vergleichbaren Bedingungen erhoben werden. Bei der Auswertung der Unterrichtsvideos orientiert sich diese Studie an der aktuellen Unterrichtsqualitätsforschung aus der Sekundarstufe, die – wie eingangs erwähnt – allgemeine, fachübergreifend gültige Basisdimensionen guten Unterrichts identifizieren konnte.<sup>5</sup> Die vorliegende Arbeit befasst sich in diesem Zusammenhang mit den beiden Basisdimensionen Klassenführung und Unterrichtsklima sowie deren Erfassung im Rahmen der PERLE-Videostudien.

Im Kontext videobasierter Studien kommen zur Erfassung der Unterrichtsqualität häufig hoch inferente Ratings zum Einsatz. Diesen Einschätzungen wird eine hohe Reliabilität und Validität zugeschrieben (zusf. Clausen, 2002). Allerdings erfordern hoch inferente Einschätzungen seitens des Beobachters auch verstärkt interpretative Schlussfolgerungen. Die Integration und gleichzeitige Bewertung von komplexen und zusammenhängenden Unterrichtsereignissen – wie dies bei der Erfassung der Unterrichtsqualität der Fall ist – führt dazu, dass hoch inferente Ratings potenziell anfälliger für systematische und unsystematische Beurteilungsfehler sind (vgl. Clausen, Reusser & Klieme, 2003). Aus diesem Grund ist eine zuverlässige und valide Messung von großer Bedeutung, denn nur so können die Ergebnisse inhaltlich sinnvoll interpretiert werden. Bislang mangelt es an Studien, die erforschen, wie zuverlässig die Einschätzungen durch externe Beobachter tatsächlich sind (zusf. Praetorius, Lenske & Helmke, 2012).

Eine große Herausforderung besteht zudem darin, ein einmal entwickeltes Ratingsystem auch in anderen Studien bzw. Fächern einzusetzen. Ein solches Vorgehen wird bislang kaum praktiziert, da Ratingsysteme je nach Forschungsinteresse häufig nur in einem spezifischen Fach, z. B. Mathematik,

---

3 Der Begriff Anfangsunterricht bezeichnet im Rahmen dieser Studie die ersten beiden Grundschuljahre (vgl. Martschinke, 2006). In der Literatur werden teilweise die Begriffe Schulanfang, Erstunterricht, Anfangsphase bzw. Schuleingangsphase synonym verwendet (vgl. Toman, 2007).

4 In der Arbeit wird aus Gründen der Lesbarkeit auf die weibliche Form verzichtet. Die maskuline Form steht selbstverständlich für beide Geschlechter.

5 „Fachübergreifend gültig“ bedeutet, dass die Merkmale über Fächergrenzen, aber auch über Altersstufen hinweg Gültigkeit besitzen. Deswegen werden diese auch als allgemeine, generische Merkmale der Unterrichtsqualität bzw. als Basisdimensionen guten Unterrichts bezeichnet.

eingesetzt werden. Ob mithilfe ein und desselben Ratingsystems über verschiedene Fächer hinweg valide und reliable Daten erfasst werden können und somit eine gewisse Vergleichbarkeit angenommen werden kann, wurde bislang nicht überprüft. An diesem Punkt setzt die vorliegende Arbeit an, deren **primäres Ziel** die Entwicklung und Erprobung eines *fachunspezifischen hoch inferenten Ratingsystems*<sup>6</sup> zur Erfassung der beiden Unterrichtsqualitätsmerkmale Klassenführung und Unterrichtsklima im Anfangsunterricht in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) ist.

Da solche allgemeinen Unterrichtsqualitätsmerkmale nicht direkt beobachtbar sind, sondern es sich vielmehr um Abstraktionen bzw. Konstrukte handelt (vgl. Helmke & Schrader, 2008), wurden in einem **ersten Schritt** beobachtbare Indikatoren identifiziert, die sich auf bestimmte Regelmäßigkeiten im Verhalten der teilnehmenden Akteure im videografierten Unterricht beziehen, unabhängig vom Fach. Dazu musste im Vorfeld eine begriffliche Eingrenzung der zwei relevanten Konstrukte Klassenführung und Unterrichtsklima erfolgen, um sie der videobasierten Erfassung im Anfangsunterricht zugänglich zu machen. Durch die begriffliche Eingrenzung konnten beobachtbare Indikatoren für den Anfangsunterricht sowohl deduktiv (auf Grundlage der Theorie) als auch induktiv (anhand des Videomaterials aus den drei Fächern) abgeleitet werden. Für die Erfassung der Effektivität der Klassenführung wurden die folgenden Aspekte im Rahmen des hoch inferenten Ratingsystems operationalisiert: Allgegenwärtigkeit der Lehrperson<sup>7</sup>, Gruppenfokus, Zeitmanagement / effiziente Zeitnutzung, Übergangsmanagement (Managing Transition), Regelklarheit bzw. -verwendung, Störungsfreiheit, effektiver Umgang mit auftretenden Unterrichtsstörungen sowie Vorbereitung des Klassenraums. Zu den untersuchten Merkmalen des Unterrichtsklimas zählen: humorvolle Lernatmosphäre, Anerkennung der Schüler durch die Lehrperson, gegenseitige Anerkennung der Schüler, Fürsorglichkeit (Herzlichkeit und Wärme), Einsatz von Lob, Ermutigung durch die Lehrperson, sachlich-konstruktive Rückmeldungen, positiver Umgang der Lehrperson mit Fehlern, positives Fehlerklima zwischen den Schülern sowie Wettbewerb und Leistungsdruck. Aufgrund der Komplexität der Unterrichtsqualität, wurden lediglich zwei allgemeine, fachunspezifische Basisdimensionen fokussiert und somit beispielsweise bewusst fachdidaktische Unterrichtsmerkmale, aber auch die kognitive Aktivierung, vernachlässigt, da diese in der Operationalisierung fachspezifische Ergänzungen benötigen und somit je nach Fach bzw. Schuljahr unterschiedlich operationalisiert werden müssten.

---

6 Im Rahmen der Arbeit werden die fachübergreifend gültigen Merkmale der Unterrichtsqualität als „fachunspezifisch“ bezeichnet, da im Vergleich zu fach- bzw. lernbereichsspezifischen Merkmalen die Fachspezifität bei der Operationalisierung der Merkmale keine Rolle spielt.

7 Aus Gründen der Lesbarkeit wird im Rahmen der Arbeit nicht von Lehrerinnen und Lehrern gesprochen, sondern die neutrale Bezeichnung „Lehrperson“ gewählt. Diese steht für beide Geschlechter gleichermaßen.

In einem **zweiten Schritt** wurde auf Grundlage der Daten aus dem hoch inferenten Rating untersucht, ob die *fachunspezifisch* operationalisierten Merkmale der Klassenführung und des Unterrichtsklimas reliabel durch zwei Rater eingeschätzt werden konnten. Die Datengrundlage bildeten hier alle Videos aus den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr). Zur Überprüfung der Beobachterübereinstimmung wurde in Anlehnung an das bisherige Vorgehen im Kontext videobasierter Studien auf den Ansatz der Generalisierbarkeitstheorie zurückgegriffen (Cronbach, Gleser, Nanda & Rajaratnam, 1972; Shavelson & Webb, 1991). Als Maß für die Güte der Messung bzw. Übereinstimmung der beiden Rater wurde der relative Generalisierbarkeitskoeffizient (G-Koeffizient) herangezogen. Im Rahmen dieser Generalisierbarkeitsstudien konnte zusätzlich die Frage beantwortet werden, inwieweit die hoch inferenten Einschätzungen der Merkmale durch Rater-Bias verzerrt sind. In einem **weiteren Schritt** erfolgte die Berechnung fachspezifischer Generalisierbarkeitsstudien. Durch diese zusätzliche fachspezifische Analyse konnte überprüft werden, ob sich Unterschiede in der Höhe der relativen G-Koeffizienten in Abhängigkeit vom Fach zeigen.

Neben der Überprüfung der Beobachterübereinstimmung (bzw. Inter-Rater-Reliabilität) und somit der Messgüte des neu entwickelten hoch inferenten Ratingsystems ist ein **weiteres Ziel dieser Arbeit**, die aus dem hoch inferenten Rating gewonnenen Daten für eine deskriptive Beschreibung des Unterrichts in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) zu nutzen. Bislang existieren nur wenige Studien, die vergleichbare Ergebnisse zur Ausprägung der zwei Unterrichtsmerkmale im Anfangsunterricht auf Basis von videobasierter Beobachtung liefern.

Da im Rahmen der vorliegenden Arbeit sowohl die Klassenführung als auch das Unterrichtsklima als mehrdimensionale Merkmale guten Unterrichts erfasst wurden, d. h. durch verschiedene hoch inferente Merkmale, wurde in einem **weiteren Schritt** überprüft, ob die theoretisch formulierten hoch inferenten Merkmale auch empirisch auf die zugrunde liegenden Daten passen. In diesem Zusammenhang wurde der Frage nachgegangen, inwieweit die erfassten Merkmale tatsächlich die postulierte 2-Faktoren-Struktur aufweisen (Überprüfung der Konstruktvalidität). Zur Erfassung der Konstruktvalidität der Daten wurden schrittweise konfirmatorische und explorative Faktorenanalysen berechnet, die unter anderem Aussagen über Gemeinsamkeiten und / oder Unterschiede in der Faktorenstruktur zwischen den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) ermöglichen.

In einem **letzten Schritt** wurde überprüft, welche Bedeutungen eine effektive Klassenführung sowie ein positives Unterrichtsklima für den Lernerfolg der Schüler haben (= prognostische Validität). Dabei wird der Lernerfolg im Rahmen der vorliegenden Studie nicht nur im Sinne einer positiven Leistungsentwicklung der Schüler verstanden. Unter Berücksichtigung der Multikriterialität von Grundschulunterricht wird zusätzlich die Bedeutung der beiden Unterrichtsqualitätsmerkmale für die Entwicklung affektiv-

motivationaler Schülermerkmale (in diesem Fall dem domänenspezifischen Selbstkonzept) untersucht. Die Analyse erfolgt differenziert für die Fächer Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr).

*Theoretischer Teil:* Der Theorieteil umfasst die Kapitel 1 bis 7. Zu Beginn gibt Kapitel 1 einen Überblick über ausgewählte ältere und neue Modelle von Unterrichtsqualität, wobei das Konzept der drei Basisdimensionen als Ausgangspunkt für die vorliegende Arbeit eine zentrale Stellung einnimmt (Kapitel 1.2). In Kapitel 2 wird der aktuelle Forschungsstand zu Unterrichtsqualität in der Grundschule zusammenfassend dargestellt. Hier werden Ergebnisse von bereits existierenden Beobachtungs- und Videostudien herangezogen (Kapitel 2.1 und 2.2), um Aussagen über die Ausgestaltung von Unterrichtsqualitätsmerkmalen in der Grundschule zu generieren (Kapitel 2.3). Im Vordergrund des Kapitels 3 steht die Beschreibung des aktuellen Forschungsstands zur effektiven Klassenführung als eine Basisdimension guten Unterrichts. Ausgehend von einer Definition des Begriffs Klassenführung (Kapitel 3.1) werden die Merkmale einer effektiven Klassenführung ausführlich theoretisch hergeleitet und beschrieben (Kapitel 3.2). Ähnlich findet in Kapitel 4 eine theoretische Annäherung an den Forschungsstand zum Unterrichtsklima statt, beginnend mit der Definition (Kapitel 4.1). Dem schließt sich die theoretische Herleitung und inhaltliche Beschreibung der Merkmale des Unterrichtsklimas an (Kapitel 4.2). Kapitel 5 befasst sich mit den postulierten Wirkungen der beiden Unterrichtsqualitätsmerkmale auf die Leistungs- und Selbstkonzeptentwicklung der Schüler. In Kapitel 6 wird allgemein auf die Methode der Videoanalyse eingegangen, dabei erfolgt in Kapitel 6.1 ein kurzer Exkurs zu den Herausforderungen von Videoanalysen in der Grundschule. Anschließend werden in Kapitel 6.2 hoch inferente Ratings als Möglichkeit der Beschreibung der Unterrichtsqualität genauer betrachtet und deren Bestandteile erläutert. Da auch im Zusammenhang mit Videoanalysen verschiedene Güteprüfungen durchgeführt werden müssen, erfolgt in Kapitel 6.3 eine Darstellung der Gütekriterien zur Qualitätskontrolle bei Videoanalysen. Hier wird vor allem ein Bezug zur Generalisierbarkeitstheorie hergestellt (Kapitel 6.3.2). Zentrale Fragestellungen und Hypothesen, die auf Basis des theoretischen Hintergrunds abgeleitet wurden, sind in Kapitel 7 wiedergegeben und leiten zum methodischen Teil der Arbeit über.

*Methodischer Teil:* Zu Beginn des methodischen Teils widmet sich Kapitel 8 in allgemeiner Hinsicht dem Design der PERLE-Studie. Dem schließt sich in Kapitel 9 die spezifische Darstellung der Vorgaben und des Ablaufs der PERLE-Videostudien in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) an. Das Vorgehen bei der Videoaufzeichnung wird in Kapitel 10 und die zugrunde liegende Stichprobe in Kapitel 11 beschrieben. In Kapitel 12.1 werden die Entwicklung des hoch inferenten Ratingsystems und die Ratingprozedur dargestellt. Die abhängigen Variablen, die für die Untersuchung der Wirkungen einer effektiven Klassenführung und eines positiven Unterrichtsklimas in Studie 3 herangezogen werden, sind in Kapitel 12.2 be-

schrieben. Die Darstellung der Ergebnisse gliedert sich anhand der Fragestellungen in verschiedene Teilstudien. Nachdem die Ergebnisse der fachunspezifischen und fachspezifischen Beobachterübereinstimmung berichtet werden (vgl. Kapitel 13), sind in einer ersten Teilstudie (Kapitel 14) die Ergebnisse der deskriptiven Analysen für die zwei Fächer dargestellt und miteinander verglichen. In einer zweiten Studie (Kapitel 15) erfolgt die Darstellung der Ergebnisse der Faktorenanalysen, die auf Basis der hoch inferenten Daten sowohl fachunspezifisch (Kapitel 15.1) als auch fachspezifisch (Kapitel 15.2) berechnet wurden. In Kapitel 16 werden die Ergebnisse der Mehrebenenregressionsanalysen zu den Wirkungen einer effektiven Klassenführung und eines positiven Unterrichtsklimas auf multikriteriale Schülermerkmale dargestellt. Auch hier erfolgt die Analyse erneut fachspezifisch, d. h. getrennt nach den Fächern Deutsch (Kapitel 16.1) und Mathematik (Kapitel 16.2).

*Diskussion:* Abschließend werden in den Kapiteln 17 und 18 die zentralen Ergebnisse der vorliegenden Arbeit zusammengefasst und diskutiert. Dabei werden Schlussfolgerungen zur Analyse der Klassenführung und des Unterrichtsklimas im Anfangsunterricht mithilfe von hoch inferenten Ratings gezogen. Für jede der Teilstudien werden daran anknüpfend die methodischen Grenzen diskutiert und ein Ausblick auf weiterführende Fragestellungen gegeben.

## II Theoretischer und empirischer Hintergrund

### 1 Modellierung von Unterrichtsqualität

Anhand ausgewählter Modelle der Unterrichtsqualität wird im Folgenden die Entwicklung der empirischen Unterrichtsqualitätsforschung dargestellt.<sup>8</sup> Zwar existieren auch in der didaktischen Forschungstradition verschiedene Ansätze zur Beschreibung der Qualität von Unterricht (vgl. Lipowsky, 2009); im Rahmen der vorliegenden Arbeit wird allerdings aufgrund der Fülle an theoretischen Ansätzen, die ihrerseits „Eigenständigkeit beanspruchen, ohne jedoch hinreichend integrative Modellbildung zu betreiben“ (Arnold, Koch-Priewe & Lin-Klitzing, 2007, S. 21), nicht auf die (fach-)didaktische Forschungstradition zur Unterrichtsqualität eingegangen.<sup>9</sup>

Zunächst werden ältere Modelle von Unterrichtsqualität näher dargestellt, beginnend mit Carroll (1963) und seinem Modell schulischen Lernens bis hin zum Produktivitätsmodell schulischen Lernens von Walberg (1981) (vgl. Kapitel 1.1). In Kapitel 1.2 wird relativ ausführlich auf das Konzept der drei Basisdimensionen guten Unterrichts von Klieme, Schümer und Knoll (2001) eingegangen. Zusätzlich werden in diesem Kapitel weitere, aktuelle Modelle von Unterrichtsqualität vorgestellt, die aufbauend auf dem Konzept der drei Basisdimensionen eine hierarchische Stufung der Basisdimensionen guten Unterrichts empirisch bestätigen konnten (vgl. Klieme, 2010, zit. n. Einsiedler, 2012; Pietsch, 2010).

Anhand der schematischen Darstellung in Abbildung 1 können ältere und neuere Konzepte der Unterrichtsqualität zeitlich eingeordnet werden. Ausgehend von einer gemeinsamen Entwicklungslinie lassen sich im Verlauf der letzten Jahrzehnte, je nach Schulstufe, unterschiedliche Entwicklungsverläufe abbilden (vgl. Einsiedler, 2012). Während sich in der empirischen For-

8 In ähnlicher Weise erfolgte dies bereits in der Arbeit von Gruehn (2000). Deswegen wird hier nur kurz darauf eingegangen.

9 Zudem erfolgt die Auseinandersetzung mit dem Konstrukt der Unterrichtsqualität in der Allgemeinen Didaktik anders als in der empirischen Unterrichtsforschung, in deren Forschungstradition diese Arbeit verortet werden kann. Unterrichtsqualität wird in der Allgemeinen Didaktik zusammenfassend als die „*Reflexion des Unterrichts*“ verstanden [...], wobei der Schwerpunkt entweder auf der Ziel-, Inhalts- und Methodendimension, der Strukturdimension oder der Beziehungsdimension liegt“ (Ditton, 2002, S. 199, Hervorh. i. Orig.). Allen didaktischen Theorien gemeinsam ist, dass die Interdependenz, in der Ziele, Inhalte und Methoden (sowie Medien) zueinander stehen, im Zentrum steht. Aus „der Stimmigkeit der Ziel-, Inhalts- und Methodenentscheidungen und aus der Konsequenz ihrer Umsetzung“ (Jank & Meyer, 2009, S. 58) erwächst die Qualität des Unterrichts.

schungstradition für die Sekundarstufe das Konzept der drei Basisdimensionen etabliert hat und vielfach empirisch bestätigt werden konnte, kann für die Grundschule attestiert werden, dass bislang keine einheitliche Konzeptualisierung bzw. Definition von Unterrichtsqualität existiert. Aus diesem Grund erfolgt in Kapitel 2.1 eine ausführliche Darstellung des aktuellen Forschungsstands zur Unterrichtsqualität in der Grundschule.

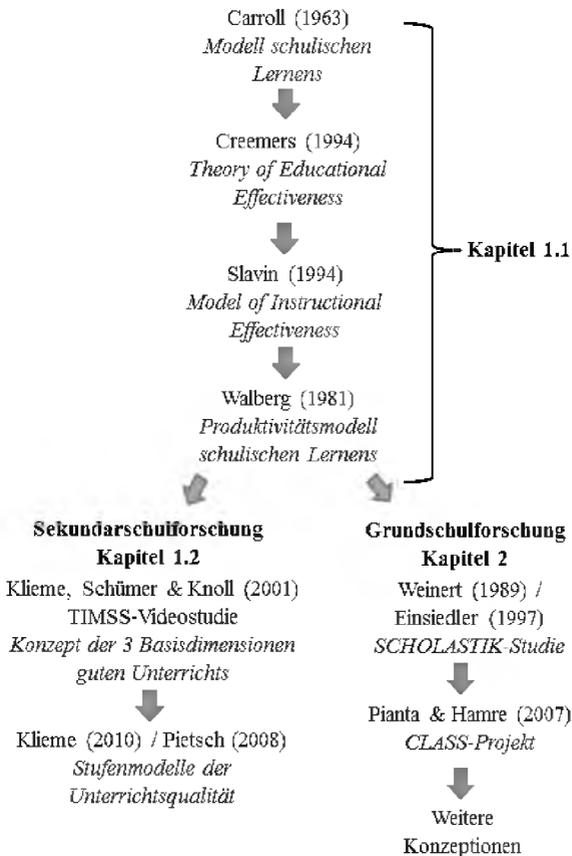


Abbildung 1: Ausgewählte Entwicklungslinien der Modelle und Konzepte zur Unterrichtsqualität (adaptiert nach Einsiedler, 2012, S. 21)

## 1.1 Ältere und neuere Konzepte der Unterrichtsqualität

Als Begründer des Konzepts *Unterrichtsqualität* kann Carroll (1963) angesehen werden. Sein *Modell schulischen Lernens* beruht auf der Grundannahme, dass das Ausmaß des Lernerfolgs durch das Verhältnis von aufgewendeter zu benötigter Zeit bestimmt wird, wobei sich beide Zeitvariablen lediglich auf das aktive Lernen<sup>10</sup> beziehen. Die individuell benötigte Lernzeit eines Schülers wird nach Carroll unter anderem durch die Unterrichtsqualität und die zugestandene Lernzeit, d. h. der Lerngelegenheit (Opportunity to Learn) beeinflusst. Als Merkmale der Unterrichtsqualität nennt Carroll (1963) (zit. n. Harnischfeger & Wiley, 1977, S. 209f): die klare Darstellung der Aufgabenanforderung, die adäquate Darbietung der Aufgabe, eine angemessene Lernschrittfolge und die Berücksichtigung der speziellen Bedürfnisse und Charakteristika des Schülers. Aufgrund der eher ungenauen Definition des Begriffs Unterrichtsqualität bzw. des nachgeordneten Stellenwertes von Unterrichtsqualität trug Carrolls Modell eher zur Weiterentwicklung des Konzepts Unterrichtsquantität / Lernzeit bei. Jedoch ging bereits Carroll davon aus, dass sowohl die Unterrichtsqualität als auch das Bereitstellen und Nutzen von mehr Unterrichtszeit durch administrative oder auch didaktische Maßnahmen von der Lehrperson beeinflusst werden können. Bei Carroll hat die Lehrperson jedoch nur indirekt über die Wirkung der zugestandenen Lernzeit und Unterrichtsqualität Einfluss auf die tatsächlich benötigte und aufgewendete Lernzeit der einzelnen Schüler.

Aufbauend auf dem Carroll'schen Modell erfolgte in den darauffolgenden Jahrzehnten eine zunehmende Ausdifferenzierung des Konstrukts Unterrichtsqualität (vgl. Bloom, 1976; Creemers, 1994a; Fraser, Walberg, Welch & Hattie, 1987; Harnischfeger & Wiley, 1977; Slavin, 1987; Slavin, 1994; Walberg, 1981). Im Zusammenhang mit der Bedeutung einer effektiven Klassenführung als Merkmal der Unterrichtsqualität (vgl. Kapitel 3) wird an dieser Stelle detaillierter auf das schulische Effektivitätsmodell (Theory of Educational Effectiveness) von Creemers (1994a) eingegangen, da dieses als einer der ersten das Lehrerverhalten als Qualitätsmerkmal mit berücksichtigt. Im Unterschied zu Carroll (1963), der die Lernzeit in den Mittelpunkt seiner Betrachtungen stellt, integriert Creemers (1994a) in seinem schulischen Effektivitätsmodell zahlreiche empirische Befunde und beschreibt die Qualität von Unterricht durch folgende drei Merkmale: (1) die Qualität der Lehrpläne und deren Umsetzung im Unterricht (Curriculum), (2) Maßnahmen zur inneren Differenzierung (Grouping Procedures)<sup>11</sup> und (3) Merkmale des Leh-

---

10 Das aktive Lernen ist die Zeit, die tatsächlich für den Lernakt verwendet wird (vgl. Harnischfeger & Wiley 1977).

11 Unter (1) fasst Creemers (1994a) einerseits die Strukturierung und Klarheit des Lernstoffs, das Anwenden sogenannter ‚Advance Organizers‘ sowie Feedbackmaßnahmen zusammen, andererseits gehört hierzu vor allem die Quantität des zu lernenden Stoffes bzw. des

rerverhaltens (Teacher Behaviour). Basierend auf Forschungen zum Zusammenhang zwischen Lehrerverhalten auf der einen Seite und dem Lernen der Schüler andererseits, bezieht Creemers (1994a) unter (3) erstmals ein effektives Klassenmanagement, das Störungen frühzeitig unterbindet und somit genügend aktive Unterrichtszeit zur Verfügung stellt, mit in seine Überlegungen ein. Zwar spricht bereits Bloom (1973) davon, dass die aktive Lernzeit durch ein besonderes Managementverhalten und ‚Monitoring‘ der Lehrperson beeinflusst werden kann, jedoch liegt der Fokus bei Bloom (1973) eher auf der Beteiligung der Schüler am Unterricht. Als ebenfalls besonders effektiv für das Lernen gelten im Modell von Creemers (1994a) folgende Merkmale des Lehrerverhaltens (zit. n. Gruehn, 2000, S. 10): wohldosierte Hausaufgaben, klare Zielsetzung und Strukturierung der Lerninhalte gemäß den hierarchisch geordneten Lernzielen oder die Klarheit der Stoffpräsentation.

Neben Creemers (1994a) trägt auch Slavin (1994) mit seinem ‚Model of Instructional Effectiveness‘ zur Weiterentwicklung der Modelle schulischen Lernens und des Begriffs Unterrichtsqualität bei. Er postuliert für die folgenden vier Faktoren einen Zusammenhang mit schulischen Leistungen: Unterrichtsqualität (Quality of Instruction), gemessenes Leistungsniveau (Appropriate Levels of Instruction), Anreizwert (Incentives to Learn, z. B. Lob) und zugestandene Lernzeit (Time). Die Unterrichtsqualität umfasst in dem sogenannten QuAIT-Modell die Qualität des Curriculums, des Materials, der Aufgaben und der Darstellung und Erklärungen (vgl. Ophardt, 2008). Eine zusammenfassende Beschreibung des Modells erfolgt bei Gruehn (2000). Diese hebt vor allem das Verdienst der Weiterentwicklungen von Creemers (1994a) und Slavin (1994) hervor, indem sie darauf verweist, dass die Modelle einerseits widersprüchliche Befunde der Unterrichtsforschung integrieren und systematisch miteinander verknüpfen, andererseits postulieren die zwei Modelle einen additiven oder möglicherweise multiplikativen Zusammenhang der einzelnen Modellkomponenten. Unterricht wird somit nicht mehr als ein aus einzelnen Elementen bestehendes Phänomen betrachtet, sondern als komplexes System „mit mehr oder weniger effektiven Kombinationen unterschiedlicher Unterrichtsmethoden und Lehrerverhaltensweisen“ (Gruehn, 2000, S. 12). Jedoch kritisiert Gruehn (2000) in diesem Zusammenhang, dass in beiden Modellen außerschulische Faktoren und Einflussfaktoren der Schüler nicht berücksichtigt wurden. Diese fehlenden Faktoren werden in dem Produktivitätsmodell von Walberg (1981; vgl. auch Fraser et al., 1987) berücksichtigt.

Einschränkend muss nach Gruehn (2000) jedoch auch für dieses Modell festgehalten werden, dass die Integration des Klassenklimas (Classroom Environment) als Determinante des schulischen Lernens additiv erfolgt. Während der Klassenführung als wesentlicher Komponente der Unterrichtsqua-

---

Unterrichts an sich. Die Maßnahmen zur inneren Differenzierung bzw. Gruppierungsmethoden (2) beinhalten währenddessen unter anderem den Einsatz kooperativer Lernformen sowie die angemessene Anwendung sogenannter Mastery-Learning-Konzepte.

lität bereits durch Creemers (1994a) eine hohe Bedeutung zugeschrieben wurde, kann auch mithilfe des Modells von Walberg (1981) keine eindeutige Aussage über die Rolle des Klassenklimas gemacht werden. Gruehn (2000) merkt in diesem Zusammenhang kritisch an, dass „nicht geklärt [ist], inwieweit das Klassenklima eine additive Komponente oder ein Mediationskonstrukt zwischen unterrichtlichem Lehrerhandeln und schulischem Lernen darstellt“ (S. 14). Neuere Modelle setzen die Sicherung eines lernförderlichen Klimas und der pädagogischen Strukturen voraus, um eine Klasse effektiv zu führen (vgl. Kapitel 1.2.1). Eine effektive Klassenführung und ein positives Unterrichtsklima stellen dabei sich gegenseitig unterstützende Faktoren der Unterrichtsqualität dar (vgl. Kapitel 1.2.2).

#### *Neuere Konzepte der Unterrichtsqualität:*

In den früheren Konzeptualisierungen wurde der Begriff Unterrichtsqualität häufig sehr eng gefasst. Neuere Erkenntnisse stammen vor allem aus Überblicksartikeln, die die Merkmale von Unterrichtsqualität in Form von katalogartigen Auflistungen darstellen (vgl. Brophy, 2000; Ditton, 2000; Helmke, 2009; Meyer, 2004; Seidel & Shavelson, 2007; Wang, Haertel & Walberg, 1990). Clausen, Schnabel und Schröder (2002) sprechen aktuell von mehr als hundert existierenden Konstrukten. So beschreiben Helmke und Schrader (2008) die Merkmale der Unterrichtsqualität als

Aspekte, mittels derer sich Unterricht und Lehrer-Schüler-Interaktionen beschreiben lassen. Diese Merkmale sind nicht direkt beobachtbar, sondern Abstraktionen. Es handelt sich um „Konstrukte“, d. h. gedankliche Ordnungsleistungen, die sich auf bestimmte Regelmäßigkeiten im beobachtbaren Verhalten beziehen. Solche Beschreibungen können mit hohem oder niedrigem Allgemeingrad vorgenommen werden (zum einen allgemein, abstrakt und situationsübergreifend, zum anderen konkret, situations- und verhaltensbezogen) (S. 27).

Je nach Autor resultieren aus solchen Katalogen oder Auflistungen unterschiedliche Merkmale guten Unterrichts, wobei sich vor allem Differenzen in der Terminologie und im Abstraktionsniveau der einzelnen Konstrukte erkennen lassen. Die Inkonsistenz zwischen den Merkmalslisten und die einfache additive Aneinanderreihung der Merkmale werden zunehmend kritisiert (z. B. Gruehn, 2000). Allerdings existieren auch Überlappungszonen bzw. werden einzelne Konstrukte von verschiedenen Autoren genannt (z. B. Supportive Classroom Climate / lernförderliches Klima). Abbildung 2 verdeutlicht diese Problematik beispielhaft anhand der Übersichten von Brophy (2000), Meyer (2004) und Helmke (2009). Die Zusammenstellung von Unterrichtsmerkmalen von Brophy (2000) gilt international als die bekannteste.

<b>Brophy (2000)</b>	<b>Meyer (2004)</b>	<b>Helmke (2009)</b>
Supportive Classroom Climate	Klare Strukturierung des Unterrichts	Klassenführung
Opportunity to Learn	Hoher Anteil echter Lernzeit	Klarheit und Strukturiertheit
Curricular Alignment	Lernförderliches Klima	Konsolidierung und Sicherung
Establishing Learning Orientations	Inhaltliche Klarheit	Aktivierung
Coherent Content	Sinnstiftendes Kommunizieren	Motivierung
Practice and Application Activities	Methodenvielfalt	Lernförderliches Klima
Thoughtful Discourse	Individuelles Fördern	Schülerorientierung
Scaffolding Students' Task Engagement	Intelligentes Üben	Kompetenzorientierung
Strategy Teaching	Transparente Leistungserwartungen	Umgang mit Heterogenität
Co-Operative Learning	Vorbereitende Umgebung	Angebotsvariation
Goal-Oriented Assessment		

Abbildung 2: Gegenüberstellung der Merkmalskataloge von Brophy (2000), Meyer (2004) und Helmke (2009)

Nach Klieme et al. (2006) sind Listen von Produktivitätsfaktoren oder auch Metaanalysen „bestenfalls Vorstufen für systematische theoretische Konzeptualisierungen“ (S. 127). Kritisiert wird im Kontext der empirischen Unterrichtsforschung das eher induktive als theoriegeleitete Vorgehen, was nach Klieme et al. (2006) dem funktionalen Verständnis der Unterrichtsqualität geschuldet ist. Der Begriff Unterrichtsqualität umschließt jedes Merkmal einer Lernumgebung, das im Sinne des Angebots-Nutzungs-Paradigmas positive Auswirkungen auf den Lernerfolg der Schüler hat.

Auch aus theoretischer Perspektive scheint die Darstellung isolierter Einzelmerkmale von Unterrichtsqualität problematisch (vgl. Pietsch, 2010), denn neuere Studien liefern vermehrt Belege dafür, dass solche Einzelmerkmale in der Regel zusammenhängen bzw. nicht isoliert voneinander betrachtet werden können, da ihre Wirksamkeit begrenzt ist (vgl. Brophy, 2000; Fraser et al., 1987a; Helmke, 2006; Pietsch, 2010; Seidel & Shavelson, 2007). So wird unter anderem den Merkmalen effektive Klassenführung und positives Unterrichtsklima ein wechselseitiger Einfluss attestiert, auf den in Kapitel 1.2.2 ausführlicher eingegangen wird.

Aufgrund der Kritik erfolgte im Kontext der TIMSS-Videostudie ein erster Versuch der theoriegeleiteten Zusammenstellung von Unterrichtsqualitätsmerkmalen in Form des Modells der drei *Basisdimensionen guten Unterrichts*. Diese drei Basisdimensionen erscheinen nach Klieme et al. (2006) besonders geeignet, die bisherigen theoretischen Konzepte und Wirkungen der empirischen Unterrichtsforschung zu integrieren. Anschließend wird auf dieses Konzept ausführlich eingegangen.

## 1.2 Das Konzept der drei Basisdimensionen der Unterrichtsqualität

Eine erste theoriegeleitete Zusammenstellung von Unterrichtsqualitätsmerkmalen sowie der Integration von Annahmen zum Zusammenhang der Lehreraktivität mit der Verarbeitung bei den Schülern erfolgte im Rahmen der TIMSS-Studie (vgl. Klieme, Schümer & Knoll, 2001; Klieme, 2002). Die Autoren konnten mittels Faktorenanalysen 21 Urteilsdimensionen videografiert Unterrichtslektionen zu folgenden drei Basisdimensionen guten (Mathematik-)Unterrichts zusammenfassen: Unterrichts- und Klassenführung, Schülerorientierung und kognitive Aktivierung (vgl. Abbildung 3).<sup>12</sup>

(1) Unterrichts- und Klassenführung	(2) Schülerorientierung	(3) Kognitive Aktivierung
Effektive Behandlung von Unterrichtsstörungen	Sozialorientierung	Genetisch-sokratisches Vorgehen
Häufigkeit von Unterrichtsstörungen (-)	Individuelle Bezugsnormorientierung	Anspruchsvolles Üben
Zeitverschwendung (-)	Diagnostische Kompetenz des Lehrers im Sozialbereich	Repetitives Üben (-)
Sprunghaftigkeit der Lehrperson (-)	Interaktionstempo (-)	Motivierungsfähigkeit des Lehrers
Regelklarheit	Leistungsdruck (-)	
Klarheit und Strukturiertheit des Unterrichts		
Monitoring		
Time on Task		

Abbildung 3: Basisdimensionen der Unterrichtsqualität (nach Klieme, Schümer & Knoll 2001, S. 51)

Die *Unterrichts- und Klassenführung* (1) fasst in der TIMSS-Videostudie die klassischen Merkmale der klaren, strukturierten und gut organisierten Instruktion zusammen (effektive Behandlung von Unterrichtsstörungen, Häufigkeit von Unterrichtsstörungen (invertiert), Zeitverschwendung (invertiert), Sprunghaftigkeit der Lehrperson (invertiert), Regelklarheit, Klarheit und Strukturiertheit des Unterrichts, Monitoring, Time on Task) und stellt eine Voraussetzung für die beiden anderen Dimensionen dar. Der Faktor *Schülerorientierung* (2) beschreibt hingegen das Eingehen auf individuelle Lernpo-

12 Clausen, Reusser und Klieme (2003) identifizierten in einer späteren Studie einen vierten Faktor „Klarheit und Strukturierung“. Klieme und Kollegen (2001) ordnen diese Aspekte hingegen dem Faktor Klassen- und Unterrichtsführung zu. In der vorliegenden Arbeit sind in Anlehnung an Clausen und Kollegen (2003) Aspekte der „Klarheit und Strukturierung“ nicht als Merkmal Klassenführung operationalisiert. Aus diesem Grund werden sie im Kontext der vorliegenden Arbeit nicht weiter betrachtet.

tenziale und Bedürfnisse der einzelnen Schüler und beinhaltet neben dem geringen Leistungsdruck („der Unterricht geht so schnell weiter, dass viele Schüler Schwierigkeiten haben mitzukommen“, invertiert) die Aspekte Sozialorientierung der Lehrperson („kümmert sich um die Probleme der Schüler“), die individuelle Bezugsnormorientierung der Lehrperson („lobt auch die Schüler, wenn er merkt, dass sie sich verbessern“), die diagnostische Kompetenz des Lehrers im Sozialbereich („merkt ziemlich schnell, wenn jemand Kummer hat“) sowie ein niedriges Interaktionstempo („geht gleich zum nächsten, wenn man nicht sofort antwortet“, invertiert) (Klieme et al., 2001). Clausen et al. (2003) ergänzen die Schülerorientierung später (TIMSS-R<sup>13</sup> Videostudie) um den Aspekt positive Fehlerkultur, der auch im Rahmen der vorliegenden Arbeit relevant ist (vgl. Kapitel 4.2). In Anbindung an die Selbstbestimmungstheorie subsumiert Rakoczy (2008) unter dem Konstrukt Schülerorientierung Maßnahmen, die dazu beitragen, das Autonomie- und Kompetenzerleben sowie das Erleben sozialer Eingebundenheit zu fördern. Dazu zählen einerseits die grundlegende Wertschätzung sowie Freiräume, aber auch die Aspekte Bewertungsstruktur (im Sinne von Feedback bzw. sachlich-konstruktiven Rückmeldungen) und die inhaltliche Relevanz des Lerngegenstands. Vor allem wertschätzende Beziehungen werden im Rahmen der Selbstbestimmungstheorie als wichtige Bedingung für motiviertes Lernen genannt (vgl. Kapitel 4.2).

Die *kognitive Aktivierung* (3) spiegelt hingegen die Komplexität der Aufgabenstellungen und Argumentationen sowie die Intensität des fachlichen Lernens wider und beinhaltet ein genetisch-sokratisches Vorgehen („lässt die Schüler auch einmal mit ihren eigenen Vermutungen in die Irre gehen, bis sie es selbst merken“), das anspruchsvolle Üben bzw. das repetitive Üben (invertiert) und die Motivierungsfähigkeit der Lehrperson (vgl. Klieme et al., 2001). Im Vergleich zu den beiden anderen Basisdimensionen ist diese Dimension stärker fachbezogen.<sup>14</sup> Eine kognitiv herausfordernde Aufgabe im Fach Mathematik muss anders operationalisiert werden als beispielsweise im Fach Deutsch oder Englisch (vgl. Klieme, 2006).

Die drei Basisdimensionen sind direkt oder indirekt auch in den zuvor erwähnten Merkmalskatalogen zu finden (vgl. Kapitel 1.1). Während Helmke (2009) im Wortlaut Klieme et al. (2001) stark ähnelt – er spricht von Klassen-

13 Die TIMSS-Repeat-Videostudie wurde als Folgestudie zu TIMSS im Jahr 1999 initiiert. Deutschland nahm an dieser Schulleistungsstudie, die durch Videostudien ergänzt wurde, nicht teil (vgl. Clausen et al., 2003).

14 Nach Klieme und Rakoczy (2008) muss sowohl von Fach zu Fach als auch vermutlich in Abhängigkeit von der Jahrgangsstufe die kognitive Aktivierung unterschiedlich konkretisiert werden. Das bedeutet, dass im Rahmen der Operationalisierung z. B. in Videostudien die kognitive Aktivierung am ehesten um fachspezifische Aspekte ergänzt werden muss. Da auch im Rahmen der PERLE-Studie die kognitive Aktivierung je nach Fach unterschiedlich operationalisiert wird, wird im weiteren Verlauf die kognitive Aktivierung als dritte Basisdimension nicht näher betrachtet.

führung, lernförderlichem Klima und Aktivierung –, tauchen zwei der von Klieme et al. (2001) beschriebenen Basisdimensionen bei Brophy (2000) nur indirekt auf. So greift Brophy (2000) in der zweiten Variable ‚Opportunity to Learn‘ (dt. Lerngelegenheit, Unterrichtszeit) den Aspekt der Unterrichtsführung und in der sechsten Variable ‚Thoughtful Discourse‘ (dt. aktive Gespräche) den Aspekt der kognitiven Aktivierung indirekt mit auf. Das ‚Supportive Classroom Climate‘ (dt. unterstützendes Klassenklima) ähnelt wiederum stark dem schülerorientierten Sozialklima. Auch Meyer (2004) nennt das lernförderliche Klima direkt. Aspekte der Unterrichts- und Klassenführung entfallen bei Meyer (2004) einerseits auf das Merkmal „hoher Anteil echter Lernzeit“ und andererseits auf das Merkmal „klare Strukturierung des Unterrichts“. Kognitive Aktivierung wird in seiner Auflistung eher dem Aspekt des intelligenten Übens zugeordnet.

Das Konzept der drei Basisdimensionen wurde in den letzten Jahren unter Verwendung einer Vielzahl von Variationen in den Begrifflichkeiten und mit unterschiedlichen Schwerpunktsetzungen immer wieder aufgegriffen, häufig auch, um den Unterricht mithilfe von Videodaten zu beschreiben (vgl. Clausen et al., 2003; Klieme, 2002; Klieme et al., 2006; Kunter et al., 2006; Kunter et al., 2007; Kunter & Voss, 2011; Lipowsky, Rakoczy, Pauli, Drollinger-Vetter, Klieme & Reusser, 2009). So wurden beispielsweise im Rahmen der Studie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“ (kurz: Pythagorasstudie) sowohl allgemeine Unterrichtsmerkmale wie das ‚Classroom Management‘ und Disziplinprobleme als auch Aspekte des Unterrichtsklimas und der Motivationsunterstützung mittels hoch inferenter Ratings erhoben (vgl. Rakoczy & Pauli, 2006) und faktoranalytisch geprüft. Auch in dieser Studie konnte die 3-Faktoren-Struktur bestätigt werden (vgl. Lipowsky et al., 2009). Waldis, Grob, Pauli und Reusser (2010b) ließen im Rahmen der internationalen und schweizerischen Videostudie zum Mathematikunterricht für das achte Schuljahr die Merkmale Klarheit und Strukturiertheit, Instruktionseffizienz, Schülerorientierung und kognitive Aktivierung hoch inferent einschätzen. Der Begriff Instruktionseffizienz wird hier synonym für die Effektivität der Klassenführung verwendet und beinhaltet neben der Regelklarheit unter anderem das Ausmaß an Unterrichtsstörungen und die Zeitnutzung / Zeitverschwendung.

### *1.2.1 Weiterentwicklungen des Konzepts der drei Basisdimensionen*

Neuere Studien bescheinigen den Basisdimensionen guten Unterrichts einen hierarchischen und kumulativen Charakter (vgl. Pietsch, 2010; Klieme, 2010, zit. n. Einsiedler, 2012). In Anlehnung an Klieme et al. (2001), die davon ausgehen, dass ein direktes, störungspräventives Verhalten der Lehrperson unter anderem die Grundvoraussetzung ist, auf der beispielsweise kognitiv aktivierende Instruktionsprozesse aufbauen müssen, entwickelte Pietsch (2008,

Pietsch, 2010; Pietsch, Schnack & Schulze, 2009) im Rahmen der Schulinspektion Hamburg<sup>15</sup> ein mehrdimensionales Modell von Unterrichtsqualität, dass unter anderem zum Ziel hat, „den Ist-Zustand der Unterrichtsqualität an Hamburger Schulen widerzuspiegeln und [...] darüber hinaus Perspektiven für die Unterrichtsentwicklung aufzuzeigen“ (Pietsch et al., 2009, S. 39). Mithilfe eines Bewertungsbogens wurden insgesamt 2.240 Unterrichtssequenzen aus allen Schulformen (Grundschulen, Haupt- und Realschulen, Gesamtschulen und Gymnasien) beurteilt. Der Bewertungsbogen (30 Merkmale guten Unterrichts) umfasst in Anlehnung an Helmke (2006) folgende sechs Kategorienbeschreibungen: Klassenmanagement und Klassenklima, Unterricht strukturieren und Methoden variieren, Motivieren, intelligentes Üben und aktives Lernen, Schülerorientierung und Unterstützung, individuelle Förderung sowie Lernerfolgssicherung. Die Arbeitsgruppe um Pietsch geht davon aus und konnte dies auch empirisch bestätigen, dass die von ihnen entwickelten 30 Beobachtungselemente unterschiedliche Ansprüche an die Qualität von Unterricht stellen; einige Items sind leicht, andere wiederum schwer zu erfüllen. Folglich kann Unterrichtsqualität auf verschiedenen Niveaus gemessen werden bzw. Lehrpersonen können die vier Stufen in unterschiedlichem Ausmaß erreichen.

Unterricht, dessen Qualität auf *Niveaustufe 1* liegt (vgl. Abbildung 4), zeichnet sich durch die Sicherung eines lernförderlichen Klimas, klare Strukturen, die Einhaltung von Regeln und durch überwiegend klare, präzise und an der Schülerschaft orientierte Arbeitsaufträge aus. Diese Stufe beinhaltet zusätzlich das Vorhandensein eines respektvollen Umgangs der Schüler untereinander. *Niveaustufe 2* erreicht ein Unterricht, der nicht nur das Lernklima und die pädagogischen Strukturen sichert, sondern darüber hinaus können die Lehrpersonen ihre Klasse effektiv führen (durch Optimierung der aktiven Lernzeit) und eine Variation in den Methoden gewährleisten. Unterricht, dessen Qualität die *Niveaustufe 3* erreicht, hebt sich durch eine vielfältige Motivation hervor. Unterricht dieser Niveaustufe ermöglicht des Weiteren aktives Lernen und Wissenstransfer für die Schüler. Unterricht auf der höchsten *Niveaustufe 4* gelingt es nicht nur, die drei darunterliegenden Gelingensbedingungen guten Unterrichts zu gewährleisten, sondern er erfüllt zusätzlich hohe Anforderungen im Hinblick auf die Schülerorientierung, Binnendifferenzierung sowie Individualisierung des Lernens. Hinzu kommt, dass die Reflexion des eigenen Lernprozesses ein wichtiger Bestandteil im Unterricht ist (vgl. zusf. Pietsch, 2010; Pietsch et al., 2009).

In dem abgestuften Unterrichtsqualitätsmodell von Pietsch (2010; vgl. auch Pietsch et al., 2009) verhalten sich die ermittelten vier Abstufungen zueinander „probabilistisch [...] d. h., dass bei Erreichen einer Abstufung da-

---

15 Die Schulinspektion Hamburg wurde im ersten Halbjahr 2008 durchgeführt. Inzwischen werden die Schulen zum zweiten Mal inspiziert ([www.schulinspektion.hamburg.de](http://www.schulinspektion.hamburg.de), Zugriff 05.12.2012).

runter liegende Merkmale mit höherer und darüber liegende Merkmale mit niedrigerer Wahrscheinlichkeit zu beobachten sind“ (Pietsch, 2010, S. 139). Die Sicherung des Lernklimas und der pädagogischen Strukturen führt nicht automatisch zu einer effektiven Klassenführung, erhöht jedoch die Wahrscheinlichkeit, dass diese gelingt. Damit untermauert Pietsch (2010) die gegenwärtig existierenden Befunde, die davon ausgehen, dass guter Unterricht anhand differenzierter Facetten beschrieben werden sollte, wobei diese nicht unabhängig voneinander betrachtet werden können (z. B. Klieme et al., 2001).

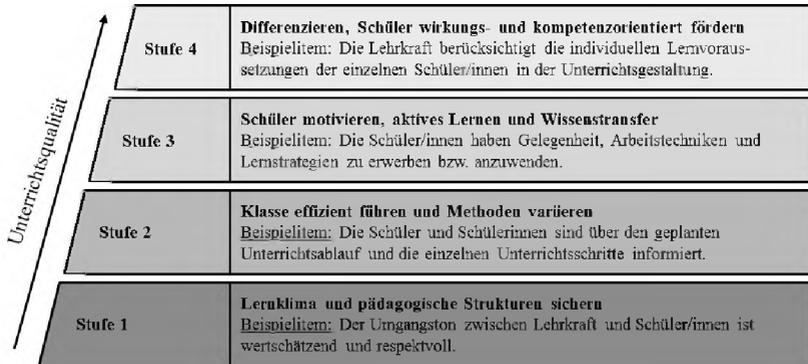


Abbildung 4: Das mehrdimensionale Stufenmodell (in Anlehnung an Pietsch et al., 2009, S. 39)

Obwohl das mehrdimensionale Modell der Unterrichtsqualität im Vergleich zu einem eindimensionalen Modell besser auf die zugrunde liegenden Daten der Schulevaluation Hamburg passte, zeigten sich sehr hohe Zusammenhänge zwischen den einzelnen Subdimensionen des Stufenmodells. Die Einhaltung von Regeln als Aspekt der Klassenführung steht bei Pietsch (2010) in einem engen Zusammenhang mit der Sicherung eines lernförderlichen Unterrichtsklimas (Niveaustufe 1), während vorausplanendes Handeln der Lehrperson bzw. die Optimierung der aktiven Lernzeit und Aspekte der Schülerorientierung erst auf der Niveaustufe 2 als Gelingensbedingung guten Unterrichts angesehen werden. Die Einhaltung von Regeln sowie die Sicherstellung eines lernförderlichen Klimas haben demnach einen geringeren Schwierigkeitsgrad für die Lehrperson als das vorausplanende Handeln, die Optimierung der aktiven Lernzeit oder die Schülerorientierung und Unterstützung. Dies legt den Schluss nahe, dass Aspekte der Klassenführung und des Unterrichtsklimas / Schülerorientierung für sich hierarchisch nach ihrer Gelingenswahrscheinlichkeit gegliedert sind. Die Sicherstellung eines lernförderlichen Klimas und pädagogischer Strukturen (Niveaustufe 1) ist somit grundlegend notwendig, um auf der nächsthöheren Niveaustufe eine optimale Zeitnutzung oder positive Schülerorientierung zu erreichen. Klieme (2010, zit. n. Einsiedler, 2012) bestätigt anhand der Daten aus der Pythagorasstudie die Rolle der

Klassenführung als Basis-Qualitätsmerkmal, das Voraussetzung für die zwei höheren Stufen ist. Darauf folgen auf Stufe 2 das unterstützende Unterrichtsklima und auf Stufe 3 die kognitive Aktivierung.

Dass sich vor allem eine effektive Klassenführung und ein positives Unterrichtsklima gegenseitig beeinflussen, wird nun noch einmal kurz anhand des Wirkungsgeflechts der Klassenführung (vgl. Helmke, 2007) dargestellt.

### *1.2.2 Die Wechselwirkung der Klassenführung und des Unterrichtsklimas*

Es konnte gezeigt werden, dass es für den Unterricht wichtig ist, dass Lehrpersonen für eine Ausbalancierung der Basisdimensionen sorgen (vgl. Klieme et al., 2006). Dies bedeutet, dass die Lehrpersonen einerseits ihren Unterricht strukturiert und effektiv führen, gleichzeitig aber auch ein unterstützendes Unterrichtsklima schaffen und die Schüler kognitiv aktivieren sollten. Vor allem eine effektive Klassenführung und ein positives Unterrichtsklima stellen sich gegenseitig unterstützende Faktoren des Unterrichts dar, was mithilfe des theoretischen Rahmenmodells von Helmke (2007) im Anschluss verdeutlicht wird (vgl. Abbildung 5). In Anlehnung an das Angebots-Nutzungs-Modell der Unterrichtswirksamkeit von Helmke hebt das Wirkungsgeflecht der Klassenführung (vgl. Helmke, 2007) auf der einen Seite die bedeutende Rolle der Klassenführung und ihre wechselseitige Beeinflussung mit gutem Unterricht hervor (in Abbildung 5 dargestellt durch einen Doppelpfeil). Gestaltet die Lehrperson den Unterricht motivierend, fordert sie die Schüler angemessen heraus und sind die Schüler aktiv an der Unterrichtsgestaltung beteiligt, dann wirft auch die Klassenführung für die Lehrperson wenige Probleme auf. In diesem Zusammenhang wird häufig argumentiert, dass Techniken der Klassenführung in ein unterstützendes und vertrauensvolles Klima eingebettet sein sollten, d. h., dass beispielsweise Regeln nicht nur aufgestellt, sondern auch gerecht und konsequent umgesetzt und eingefordert werden sollten. Die verlässliche Einhaltung von Regeln wirkt sich förderlich auf das Klima einer Klasse aus (vgl. Meyer, 2004). Für Eder (2011) ist unter anderem eine „durch Regelklarheit, Aufgabenorientierung und Disziplin geprägte Klassenführung“ (S. 115) Ausdruck eines positiven Klimas.

Andererseits wird in dem Modell vor allem „auf die entscheidende Rolle eines durch Unterstützung, Freundlichkeit und wechselseitigen Respekt charakterisierten Lernklimas und der Aktivierung der Schüler hingewiesen“ (Helmke, 2007, S. 45). In Klassen mit einem positiven Unterrichtsklima arbeiten die Schüler intensiver mit und stören seltener bzw. zeigen weniger abweichendes Verhalten (vgl. Holtappels, 1985). Aus diesem Grund gilt für Hennemann und Hillenbrand (2010; vgl. auch Evertson et al., 2006) theoretisch das Schaffen eines positiven (Lern-)Klimas im Klassenraum als proaktives, d. h. vorausplanendes Kriterium des Classroom Managements (vgl. Ka-

pitel 3.1) und somit als ein wichtiges Merkmal zur Vermeidung von Fehlverhalten aufseiten der Schüler.

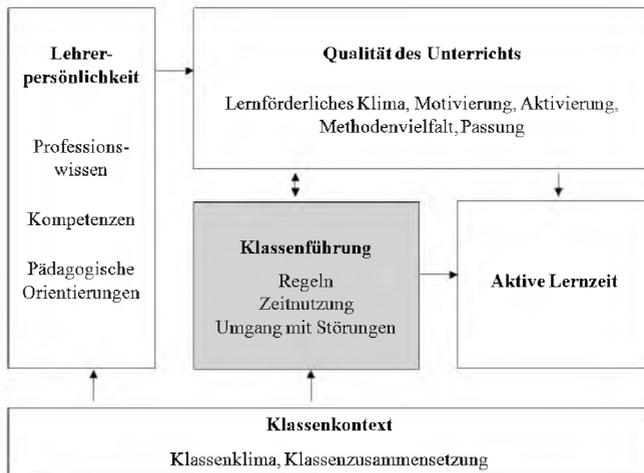


Abbildung 5: Wirkungsgeflecht der Klassenführung (vgl. Helmke, 2007, S. 45)

Auch Gräsel und Mandl (2009) gehen davon aus, dass „ein wichtiger Einflussfaktor auf die aktive Lernzeit ist, inwieweit in der Klasse ein lernförderliches Klima besteht“ (S. 253). Jones (1996) nennt als Hauptziel der Klassenführung die Schaffung einer positiven Lehrer-Schüler-Beziehung. Es kann zudem angenommen werden, dass in Klassen, die sich durch eine positive Lehrer-Schüler-Beziehung auszeichnen, die Schüler ohne Weiteres Regeln und Konsequenzen für Regelbrüche akzeptieren (vgl. Marzano, 2003).

Auch in zahlreichen gängigen Verfahren zur Messung des Unterrichtsklimas werden neben Aspekten der Lehrer-Schüler-Beziehung und der Schüler-Schüler-Beziehung Merkmale des Unterrichts erfasst (zusf. Gruehn, 2000, S. 73f). Somit enthalten viele Klimafragebögen zusätzlich Komponenten, die unmittelbar mit Klassenführungsmerkmalen in Verbindung gesetzt werden können. Beispielsweise werden mit dem CES (Classroom Environment Scale, Moos & Trickett, 1974) unter anderem die Ordnung und Organisation sowie die Regelklarheit in der Klasse untersucht, während mit den Landauer Skalen zum Sozialklima (LASSO, v. Saldern & Littig, 1985) zusätzlich Merkmale von Disziplin und Ordnung erfasst werden.

Bislang zeigen die Ergebnisse auf Basis von Beobachtungsstudien eher niedrige Zusammenhänge zwischen einer effektiven Klassenführung und einem positiven Unterrichtsklima. So konnten Helmke und Renkl (1993) für das zweite, dritte und vierte Schuljahr einen Zusammenhang zwischen der hoch inferent erfassten Klassenführung und dem Sozialklima bestätigen ( $r$  zwischen .13 und .37), wobei dieser nur im dritten Schuljahr signifikant wird und

hier auch am höchsten ausfällt. Dies hängt vermutlich mit dem obligatorischen Lehrerwechsel nach dem zweiten Schuljahr zusammen. Eine effektive Klassenführung und die gleichzeitige Schaffung eines positiven Sozialklimas scheinen vor allem direkt nach der Übernahme einer neuen Klasse von Bedeutung zu sein. Die Beobachtung der einzelnen Unterrichtsmerkmale erfolgte in der Studie von Helmke und Renkl (1993) während des regulären Unterrichts in den Hauptfächern. Eine fachspezifische Analyse zum Zusammenhang der beiden Unterrichtsqualitätsmerkmale erfolgte nicht. Für den Rechtschreibunterricht im vierten Schuljahr konnte Hofmann (2007) im Rahmen ihrer Dissertation keinen Zusammenhang zwischen dem Arbeitsklima und der Klassenführung bestätigen.

Auf Ebene einzelner Merkmale der Klassenführung und des Unterrichtsklimas können hingegen höhere Zusammenhänge erwartet werden. In dem Stufenmodell der Unterrichtsqualität von Pietsch (2010; vgl. Kapitel 1.2.1) zeigten sich hohe Zusammenhänge zwischen der Einhaltung von Regeln als Aspekt der Klassenführung und der Sicherung eines lernförderlichen Unterrichtsklimas. In der Studie von Pianta, La Paro und Hamre (2008) konnte ein positiver Zusammenhang zwischen einem positiven Klima und den Merkmalen Verhaltensmanagement ( $r = .51$ ) und Produktivität ( $r = .38$ ) bestätigt werden. In Klassen, die sich durch eine positive Lehrer-Schüler- bzw. Schüler-Schüler-Interaktion sowie durch Wärme und Respekt auszeichnen, zeichnet sich die Lehrperson demnach durch besondere Fähigkeiten im Umgang mit Unterrichtsstörungen und durch ein effizientes Zeitmanagement aus. Auch andere Merkmale des ‚Emotional Supports‘ und der ‚Classroom Organization‘ korrelierten im Rahmen der Validierung des CLAS-Systems (vgl. Pianta et al., 2008; Hamre & Pianta, 2007) in der dritten Jahrgangsstufe teilweise sehr hoch mit bis zu  $r = .66$ . Auch auf Skalenebene können deutlich positive Zusammenhänge der drei Subskalen des CLASS bestätigt werden (vgl. Cadima, Leal & Burchinal, 2010). So kommen Cadima et al. (2010) im Rahmen ihrer Studie zu dem Ergebnis: „teachers who were observed to be warmer and consistently responsive to students also tended to be more proactive, managed the activities and student behavior more efficiently, and provided activities that encourage higher-order thinking“ (S. 475).

### *1.2.3 Ist das Konzept auf alle Fächer sowie auf die Grundschule übertragbar?*

Obwohl ein Großteil der Studien zur Unterrichtsqualität bzw. zu den drei Basisdimensionen sich eher am naturwissenschaftlichen bzw. mathematischen Bereich orientiert, findet häufig eine Übertragung der Ergebnisse für andere Fächer wie beispielsweise Deutsch statt. Helmke und Klieme (2008) merken in diesem Zusammenhang jedoch an, dass sich das Modell der drei Basisdimensionen zwar bei der Erklärung der Wirksamkeit naturwissenschaft-

lichen Unterrichts bewährt hat, es „[...] kann allerdings – aus Gründen der Fachspezifität – nicht ohne weiteres auf den Deutsch- und Englischunterricht übertragen werden“ (S. 306). Im Rahmen der DESI-Studie<sup>16</sup> zeigte sich beispielsweise für das Fach Englisch, dass im Vergleich zu anderen Merkmalen das Merkmal „Strukturierung“ in seiner Bedeutung für den Kompetenzerwerb zurücktritt (vgl. Helmke, Helmke, Schrader, Wagner, Nold & Schröder, 2008). Demnach kann grundsätzlich die Frage gestellt werden, ob den drei Basisdimensionen in sprachlichen Fächern dieselbe Bedeutung zukommt. Hier mangelt es an entsprechenden Studien. Bisherige Studien beispielsweise zum Schriftspracherwerb berücksichtigen eher selten unterrichtliche Einflussmerkmale, sondern beziehen allenfalls Kontext- bzw. sozioökonomische Merkmale in entsprechende Analysen ein (zusf. Roos & Schöler, 2009).

Die wenigen Ergebnisse für den Bereich Deutsch / Sprache deuten jedoch darauf hin, dass auch hier angenommen werden kann, dass sich Unterrichtsqualität durch verschiedene Grunddimensionen auszeichnet, obwohl es an einem einheitlichen Qualitätsverständnis mangelt (vgl. Eikenbusch, 2001; Wildemann, 2011; Klieme et al., 2008). Beispielsweise konnten Klieme et al. (2008) für das Fach Deutsch (Sekundarstufe) im Rahmen der DESI-Studie (DESI-Konsortium, 2008) unter anderem die Grunddimensionen schülerorientiertes Unterrichtsklima und effektive Klassenführung faktoranalytisch bestätigen. Grundlagen bildeten hier jedoch Angaben zur Unterrichtsqualität aus Schülersicht. Für das Fach Englisch liegen hingegen Videoaufzeichnungen zur Gewinnung verhaltensnaher Indikatoren der Unterrichtsqualität vor. Mithilfe einer hoch inferenten Einschätzung der Unterrichtsvideos konnten auch für das Fach Englisch unter anderem die Klassenführung, das Unterrichtsklima und das Fehlerklima<sup>17</sup> als wesentliche Aspekte der Unterrichtsqualität identifiziert werden (vgl. Helmke et al., 2008).

Folgt man der Argumentation von Klieme et al. (2001), muss im Prinzip jeder Unterricht die oben genannten drei Basisdimensionen erfüllen (mit unterschiedlicher Schwerpunktsetzung und Qualität). Bis zum jetzigen Zeitpunkt mangelt es in Deutschland jedoch an Studien aus dem Grundschulbereich (vgl. Kapitel 2.1) und somit an Belegen, ob sich diese Grunddimensionen auch im Grundschulunterricht reliabel – im vorliegenden Fall mithilfe von Unterrichtsvideos – erfassen und voneinander abgrenzen lassen. Zudem stellt sich die Frage, ob den drei Basisdimensionen in den Fächern Deutsch und Mathematik die gleiche Bedeutung zukommt. Auf den aktuellen For-

---

16 DESI steht als Abkürzung für **D**eutsch **E**nglisch **S**chülerleistungen **I**nternational.

17 Die Klassenführung beinhaltet Aspekte der Aufgabenorientierung, Unterrichtsplanung und Zeitmanagement sowie Störungsfreiheit. Das Unterrichtsklima wurde getrennt vom Fehlerklima erfasst, da eine inhaltliche Trennung erfolgt. Das Unterrichtsklima wurde spezifiziert über Wärme, Herzlichkeit und Humor, während das Fehlerklima den positiven versus negativen Umgang mit Schülerfehlern seitens der Lehrperson und seitens der Klasse beinhaltet (vgl. Helmke et al., 2008).

schungsstand zur Qualität des Anfangsunterrichts bzw. der Grundschule allgemein wird in Kapitel 2 ausführlich eingegangen.

Neben der Kritik an der Übertragbarkeit der Ergebnisse aus dem naturwissenschaftlich-mathematischen Bereich, stellt sich im Zusammenhang mit dem Konzept der drei Basisdimensionen zusätzlich die Frage, ob die Qualitätsmerkmale, die hauptsächlich auf die direkte Instruktion ausgerichtet sind, auch für Grundschulunterricht erhoben werden können. Wie Einsiedler (2000) anmerkt, ist die bisherige Forschung schwerpunktmäßig auf die Qualität des Klassenunterrichts (die sog. direkte Instruktion, vgl. Good, 1979; Grell, 2002) gerichtet. Wenig ist bislang bekannt über die Bewertung der Qualitätsmerkmale in einer eher offenen und selbstgesteuerten Unterrichtsgestaltung (Ausnahme: WOLFF-Projekt<sup>18</sup>, vgl. Kucharz, Schnebel & Helming, 2011; vgl. Kapitel 2.2), wie sie im Grundschulunterricht vermehrt aufzufinden ist. Hier kommen offenere Unterrichtsmethoden wie beispielsweise Stationenlernen oder Projektarbeit zum Einsatz, was unter anderem Auswirkungen auf die Rolle der Lehrperson und die Aspekte der Klassenführung hat (vgl. Bohl & Kucharz, 2010). Kammermeyer und Martschinke (2009) kritisieren in diesem Kontext, dass die klassische Unterrichtsqualitätsforschung dem Grundschulunterricht nicht gerecht wird. Aus diesem Grund legen sie im Rahmen der KILIA-Studie<sup>19</sup> ein sehr weites Verständnis von Unterrichtsqualität zugrunde und schlagen für die Grundschule eine theoretische Anbindung an die Selbstbestimmungstheorie vor.

## 2 Unterrichtsqualität in der Grundschule

Während Studien zur Unterrichtsqualität in der Sekundarstufe sich derzeit häufig an dem Konzept der drei Basisdimensionen orientieren und dieses zur Beschreibung von Unterrichtsqualität immer wieder aufgreifen, existieren für die Grundschule bislang nur wenige Studien zur Unterrichtsqualität – vor allem in Deutschland. Dabei kommt der Unterrichtsqualität aufgrund des Klassenlehrerprinzips vermutlich eine besondere Bedeutung zu. So konnten beispielsweise Babu und Mendro (2003) im Rahmen ihrer Längsschnittstudie zeigen, dass sich Unterrichtsqualität (Teacher Effectiveness) auf schwächere und jüngere Schüler auswirkt bzw. diese deutlicher von der Qualität der Lehrperson und des Unterrichts profitieren. Es zeigen sich ähnliche Befunde für Mathematik und Lesen (zusf. Lipowsky, 2007a). Des Weiteren belegten Babu und Mendro (2003) in ihrer Studie, dass der mittlere Leistungszuwachs in

---

18 WOLFF steht als Abkürzung für **Wirksamkeit offener Lehr-Lern-Formen**.

19 KILIA steht als Abkürzung für **Kooperationsprojekt Identitäts- und Leistungsentwicklung im Anfangsunterricht**.

den ersten Schuljahren größer ist als in den darauffolgenden Jahren. Daher kann angenommen werden, dass sich Verbesserungen der Unterrichtsqualität bzw. des Unterrichts allgemein besonders in den Anfangsjahren auszahlen. Die Frage nach der Übertragbarkeit der Ergebnisse für Deutschland bleibt jedoch bislang unbeantwortet.

Im Folgenden wird der Frage nachgegangen, wie Unterrichtsqualität in der Grundschule bislang operationalisiert wurde und ob sich die drei Basisdimensionen, deren Bedeutung für die Sekundarstufe vielfach bestätigt werden konnte, auch in Konzeptualisierungen von Unterrichtsqualität für die Grundschule wiederfinden lassen. Ein weiteres Ziel des folgenden Kapitels ist die vergleichende Betrachtung verschiedener Konstrukte der Unterrichtsqualität in der Grundschule.

## 2.1 Studien und Konzepte von Unterrichtsqualität in der Grundschule

Betrachtet man Studien aus dem Grundschulbereich, so kann festgestellt werden, dass je nach Studie das Konstrukt der Unterrichtsqualität im Hinblick auf verschiedene Aspekte operationalisiert wird (z. B. Hamre et al., 2007; Helmke & Schrader, 1997; Kammermeyer & Martschinke, 2009; NICHD ECCRN<sup>20</sup>, 2002; Roßbach, 2002a; b; Pianta et al., 2008; Perry, Donohue & Weinstein, 2007; Teilprojekt 2 der BiKS<sup>21</sup>-Forschergruppe, 2009). Im Vergleich zur Sekundarstufe gibt es bislang keinen einheitlichen Forschungsstand. Diese Problematik der unterschiedlichen Operationalisierung wird anhand von Tabelle 1 (S. 33) deutlich. Diese erhebt keinen Anspruch auf Vollständigkeit, sondern soll nur die Problematik der unterschiedlichen Operationalisierungen von Unterrichtsqualität dokumentieren.

In Deutschland liegen unter anderem Ergebnisse zur Unterrichtsqualität in der Grundschule aus der KILIA-Studie (vgl. Kammermeyer & Martschinke, 2009) und der SCHOLASTIK-Studie<sup>22</sup> (vgl. Helmke & Schrader, 1997) sowie für die deutsche Teilstichprobe der European Child Care and Education

---

20 NICHD ECCRN steht als Abkürzung für National Institute of Child Health and Human Development Early Child Care Research Network.

21 BiKS steht als Abkürzung für Bildungsprozesse, Kompetenzentwicklung und Formation von Selektionsentscheidungen im Vor- und Grundschulalter. Teilprojekt 2 der Längsschnittstudie BiKS 3–8 untersucht – unter der Leitung von Prof. Dr. Hans-Günther Roßbach (Universität Bamberg) – die längsschnittlichen Auswirkungen globaler und bereichsspezifischer Anregungsqualität in Familie, Kindergarten und Grundschule für die Kompetenzentwicklung von Kindern.

22 SCHOLASTIK steht als Abkürzung für Schulorganisierte Lernangebote und Sozialisati-on von Talenten, Interessen und Kompetenzen.

(ECCE<sup>23</sup>) Study (Roßbach, 2002a; b) vor. Zusätzlich werden derzeit im Rahmen des FiS-Projekts<sup>24</sup> an der Universität Münster (vgl. Hanke, 2007; Hein, Eckerth & Hanke, 2010; Eckerth, Hanke & Hein, 2012) Merkmale der Unterrichtsqualität mithilfe von Unterrichtsbeobachtungen erfasst. Auch im Rahmen der Schulinspektion Hamburg wurden Merkmale der Unterrichtsqualität in der Grundschule durch geschulte Beobachter erhoben (z. B. Pietsch, 2010).

Als am weitesten verbreitete Konzeptualisierung der Unterrichtsqualität gilt in Deutschland bis heute die Definition aus der SCHOLASTIK-Studie von Weinert, Schrader und Helmke (1989). Sie definieren das Metakonstrukt Unterrichtsqualität als „jedes stabile Muster von Instruktionsverhalten, das als Ganzes oder durch einzelne Komponenten die substantielle Vorhersage und / oder Erklärung von Schulleistung erlaubt“ (S. 899, zit. n. Klieme & Reusser, 2003, S. 194). Einsiedler (1997) (vgl. auch Einsiedler, 2000) bewertet jedoch die Definition von Weinert et al. (1989) als zu eng gefasst und schlüsselt im Sinne einer Weiterentwicklung das Instruktionsverhalten systematisch auf nach:

- Makromethoden des Unterrichts (z. B. entdecken-orientiertes Lernen),
- Mikromethoden des Unterrichts (z. B. Strukturierungshinweise, Feedback oder kognitives Niveau der Gesprächsführung),
- Klassenmanagement,
- Sozialformen (Klassenunterricht, Gruppenarbeit, Einzelarbeit) und
- Sozialklima.

Wie bereits Walberg (1981; vgl. Kapitel 1.1) nennt Einsiedler (1997) für die Grundschule neben dem Klassenmanagement das Sozialklima einer Klasse als wesentliches Merkmal der Unterrichtsqualität bzw. des Instruktionsverhaltens der Lehrperson. Dem Konstrukt Klassenführung liegen die Skalen Unterrichtsorganisation und Zeitnutzung zugrunde. Das soziale Klima setzt sich aus der Akzeptanz des Privatbereiches und dem Beziehungsaspekt zusammen. Trotz der Weiterentwicklung durch Einsiedler (1997) beschränkt sich die Definition von Unterrichtsqualität auf die Wirksamkeit in Bezug

23 Die Studie European Child Care and Education (kurz ECCE-Studie) untersucht in Deutschland, Österreich, Portugal und Spanien die Qualitäten institutioneller Gruppenbetreuungen für Kinder im Kindergartenalter und deren Auswirkungen auf die Kinder im Alter von vier bis acht Jahren. In einer ersten Phase (1993/1994) wurden insgesamt 1.244 vierjährige Kinder aus vorschulischen Einrichtungen in die Erhebung einbezogen. In einer zweiten Phase 1997/1998) konnten insgesamt 586 Kinder in Längsschnittdaten weiterverfolgt werden (vgl. Roßbach, Kluczniok & Isenmann, 2008). Roßbach (2002a; b) veröffentlicht Ergebnisse zur Unterrichtsqualität in Grundschulklassen für die deutsche Teilstichprobe (überwiegend 2. Schuljahr).

24 FiS steht als Abkürzung für Förderung der Lern- und Bildungsprozesse von Kindern in der Schuleingangsphase und wurde in den Jahren 2006 bis 2009 unter der Leitung von Prof. Dr. Petra Hanke und Dr. Anna Katharina Hein an der Universität Münster durchgeführt (vgl. Hanke, 2007).

auf Leistungskriterien. Dieser alleinige Bezug auf Instruktionsverhalten und Schulleistung wird kritisiert (z. B. Kleinbub, 2010). Da Schulen und Unterricht neben dem Bildungs- auch einen Erziehungsauftrag haben, ergeben sich unterschiedliche Zielkriterien (vgl. Kapitel 5).

Der KILIA-Studie liegt ein erweitertes Verständnis von Unterrichtsqualität zugrunde (vgl. Kammermeyer & Martschinke, 2009). Die Konzeption basiert auf der Annahme, dass sich guter Grundschulunterricht vor allem dadurch auszeichnet, „dass er grundlegende psychologische Bedürfnisse der Kinder nach Selbstbestimmung, Kompetenzerleben und sozialem Eingebundensein berücksichtigt“ (Kammermeyer & Martschinke, 2009, S. 43). Die Unterstützung dieser grundlegenden Bedürfnisse kann nach Ansicht der Autorinnen eben nicht durch Unterrichtsmerkmale erreicht werden, die sich nur auf den lehrergesteuerten Unterricht im Sinne direkter Instruktion beschränken. Aus diesem Grund liegen dem Unterrichtsqualitätskonzept von Kammermeyer und Martschinke (2009), im Vergleich zur SCHOLASTIK-Studie, folgende drei Merkmale zugrunde: Freiheitsspielräume, Kooperationsmöglichkeiten und konstruktivistischer Unterricht. Mithilfe ihres Unterrichtsqualitätskonzepts suchten Kammermeyer und Martschinke (2009) theoriegeleitet nach Zusammenhängen mit multikriterialen Zielen<sup>25</sup>. Allerdings konstatieren sie in ihrer Studie selbst, „dass auch mit der Orientierung an den grundlegenden Bedürfnissen nur ein Ausschnitt des komplexen Unterrichtsgeschehens erfasst wird [...]“ (ebd., S. 51). Vollkommen unberücksichtigt bleiben in ihrer Studie – aufgrund des weiten Verständnisses von Unterrichtsqualität und des Versuchs einer theoretischen Herleitung – z. B. die drei Basisdimensionen als die für die Sekundarstufe am häufigsten empirisch nachgewiesenen Merkmale guten Unterrichts (vgl. Kapitel 1.2).

Im Rahmen der European Child Care and Education (ECCE) Study identifizierte Roßbach (2002b) (vgl. auch Roßbach, 2002a) für die deutsche Teilstichprobe (überwiegend zweites Schuljahr) neben den Faktoren direktes Unterrichten, Anwendungsbezug und Unterstützung in Stillarbeit die Faktoren Klassenmanagement und schülerorientiertes sozial-emotionales Klima. Während ein positives Klassenmanagement einerseits eine Organisation des Unterrichts anspricht, bei dem beispielsweise die Übergänge fließend sind, klare Regeln existieren und eingehalten werden oder effektiv mit Störungen umgegangen wird, gehört bei Roßbach (2002a; b) auch zum Klassenmanagement dazu, dass sich die Lehrperson bemüht, ein Klima wechselseitigen Respekts und sozialer Unterstützung des Lernens zu generieren. Hier wird erneut der Zusammenhang zwischen dem Unterrichtsmerkmal Klassenmanagement und Unterrichtsklima deutlich (vgl. Kapitel 1.2.2). Ein positives schülerorien-

---

25 Als Kriterien der Unterrichtsqualität wurden im Rahmen der KILIA-Studie einerseits die Leistungen der Schüler und andererseits, im Hinblick auf die Persönlichkeitsentwicklung der Schüler, sowohl das fachspezifische Selbstkonzept als auch die fachspezifische Lernfreude in den Bereichen Schriftspracherwerb und Mathematik erhoben (vgl. Kammermeyer & Martschinke, 2009).

tiertes sozial-emotionales Klima indiziert bei Roßbach (2002a; b) vor allem, dass die Schüler als Person anerkannt werden bzw. die Lehrperson Raum für private Angelegenheiten zur Verfügung stellt sowie die Ablehnung eines leistungsorientierten und kontrollierenden Klimas in der Klasse.

Das FiS-Projekt (vgl. Eckerth et al., 2012; Hanke, 2007; Hein et al., 2010) untersucht aktuell, wie sich der Anfangsunterricht bezogen auf die Klassenführung, die individuelle Unterstützung sowie die inhaltliche Strukturierung und kognitive Aktivierung gestaltet. Im Rahmen des Projekts wird unter anderem der Frage nachgegangen, was den Unterricht in jahrgangshomogenen und jahrgangsübergreifenden Lerngruppen bezogen auf den schriftsprachlichen und mathematischen Anfangsunterricht kennzeichnet bzw. welche unterrichtlichen Bedingungen sich in den zwei verschiedenen Lerngruppen als besonders wirksam für die Förderung erweisen (vgl. Hein et al., 2010).

Die Forschergruppe des Teilprojekts 2 der BiKS 3–8 Längsschnittstudie orientiert sich bei der Operationalisierung der Unterrichtsqualität eng an dem Konzept der drei Basisdimensionen aus der Sekundarstufe (vgl. Teilprojekt 2 der BiKS-Forschergruppe, 2009). Hier werden verschiedene Aspekte der Klassenführung / Regelklarheit / Struktur, des unterstützenden Lernklimas (Lehrer-Schüler-Verhältnis) und der kognitiven Aktivierung erfasst. So beinhaltet die Klassenführung / Regelklarheit / Struktur die Aspekte Beeinträchtigung des Unterrichts durch Störungen, störungspräventives Verhalten, Umgang mit Unterrichtsstörungen, Struktur und Klarheit der Unterrichtseinheit, Unterrichtsorganisation / Zeitnutzung und sachlich-konstruktive Rückmeldung / positive Fehlerkultur. Unter dem Begriff des unterstützenden Lernklimas (Lehrer-Schüler-Verhältnis) werden die folgenden Aspekte subsumiert: Anerkennung der Schüler durch den Lehrer, Anerkennung des Lehrers durch die Schüler, gegenseitige Anerkennung der Schüler, Engagement des Lehrers, Engagement der Schüler, Beziehungsaspekt und Regelverwendung. Auch die kognitive Aktivierung im Unterricht wird im Rahmen des Teilprojekts mithilfe zahlreicher Merkmale erfasst. Hierzu zählen unter anderem die Herausforderung zum anspruchsvollen Denken, die Exploration von Vorwissen und Denkweisen der Schüler, das Gesprächsverhalten des Lehrers oder das Vorkommen herausfordernder Hilfestellung. Auf eine ausführliche inhaltliche Beschreibung der einzelnen Merkmale wird an dieser Stelle verzichtet.

Tabelle 1: Komponenten der Unterrichtsqualität in der Grundschule – Eine Gegenüberstellung verschiedener Studien

Studie & Autoren	Schuljahr	Operationalisierung der Unterrichtsqualität
<i>Deutschsprachige Studien</i>		
Schulinspektion Hamburg: z. B. Pietsch (2008; 2010)	Grundschule & Sekundarstufe	Klassenmanagement und Klassenklima, Unterricht strukturieren, Methoden variieren, motivieren, intelligent üben, aktiv lernen, Schülerorientierung und Unterstützung, individuelle Förderung, Lernerfolgssicherung
KILIA-Studie: Kammermeyer & Martschinke (2009)	1. SJ	Freiheitsspielräume, Kooperationsmöglichkeiten und konstruktivistischer Unterricht
ECCE-Study: Roßbach (2002a, b)	2. SJ	Klassenmanagement, schülerorientiertes sozial-emotionales Klima, direktes Unterrichten, Anwendungsbezug, Unterstützung in Stillarbeit
SCHOLASTIK-Studie: Helmke & Weinert (1997b)	3./4. SJ	Klassenführung, Strukturiertheit, individuelle fachliche Unterstützung, Förderungsorientierung, soziales Klima, Variabilität der Unterrichtsformen, Klarheit und Motivierungsqualität
FIS-Projekt: Hanke (2007)	Anfang 1. bis Ende 2. SJ	Klassenmanagement, individuelle Unterstützung, inhaltliche Strukturierung, kognitive Aktivierung
Teilprojekt 2 der BiKS-Forschergruppe (2009)	Kindergarten / 1.–2. SJ	Unterstützendes Lernklima, Klassenführung / Regelklarheit / Struktur, kognitive Aktivierung
<i>Internationale Studien</i> (dt. Übersetzungen erfolgen im Zuge der Beschreibung)		
Hamre & Pianta (2007) / Pianta et al. (2008)	Preschool / Kindergarten, 1.–5. SJ	<ul style="list-style-type: none"> <li>– Emotional Support (Positive Climate, Negative Climate, Teacher Sensitivity, Regard to Student Perspectives / Overcontrol)</li> <li>– Classroom Organization (Behavior Management, Productivity, Instructional Learning Formats)</li> <li>– Instructional Support (Concept Development, Quality of Feedback, Language Modeling)</li> </ul>
Stipek (1997, zit. n. Perry, Donohue & Weinstein, 2007)	Preschool bis 2. SJ	<ul style="list-style-type: none"> <li>– Instructional Climate (Learning Standards, Coherence of Instructional, Teaching Concepts, Instructional Conversation),</li> <li>– Social Climate (Teacher Warmth / Responsiveness, Support for Instructional Skills, Support for Communication, Relevance of Activities to Children's Experience, Student Engagement, Individualization of Learning),</li> <li>– Management (Management, Discipline, Child Responsibility)</li> </ul>
National Institute of Child Health and Human Development Early Child Care Research Network (NICHD ECCRN, 2002)	1. SJ	<ul style="list-style-type: none"> <li>– Emotional Support (Overcontrol, Intrusiveness, Detachment, Positive Emotional Climate, Negative Emotional Climate, Management, Sensitivity / Responsivity)</li> <li>– Instructional Support (Literacy Instruction, Evaluative Feedback, Instructional Conversation, Encourages Responsibility)</li> </ul>

Im Zusammenhang mit der Konzeption von Unterrichtsqualität in der Grundschule lohnt zusätzlich ein Blick über die Grenzen Deutschlands hinaus. Hier liefern vor allem die Arbeitsgruppe um R. C. Pianta von der University of Virginia sowie Studien des National Institute of Child Health and Human Development (NICHD) eine Fülle an Forschungsergebnissen (vgl. Hamre et al., 2007; NICHD ECCRN, 2002; Pianta & La Paro, 2003; Pianta et al., 2008; Stuhlman & Pianta, 2009). Die Instruktionsqualität der Lehrer-Schüler-Interaktionen werden auf der obersten Ebene in folgende drei Faktoren (Domänen) der ‚Classroom Quality‘ gruppiert (vgl. Pianta et al., 2008; Hamre et al., 2007; Pianta & Hamre, 2009): (1) emotionale Unterstützung (Emotional Support), (2) Organisation des Klassenraums (Classroom Organization) und (3) Anreizungsqualität (Instructional Support) (vgl. Abbildung 6).



Abbildung 6: Überblick über die Faktoren des CLASS-Beobachtungssystems und deren Merkmale (vgl. Pianta et al., 2008, S. 88)

Der Faktor ‚Emotional Support‘ beinhaltet vier Merkmale: positives Klima (Positive Climate), negatives Klima (Negative Climate) als Gegenteil, Sensibilität der Lehrperson (Teacher Sensitivity) und die Berücksichtigung der Schülerperspektiven (Regard to Student Perspectives). Ein positives Klima ist beispielsweise durch Wärme sowie positive Lehrer-Schüler-Beziehungen und Schüler-Schüler-Beziehungen charakterisiert. Das Merkmal Sensibilität der Lehrperson beinhaltet unter anderem, dass die Lehrperson die Schüler unterstützt und ermutigt und sensibel gegenüber den schulischen und emotionalen Belangen der Schüler ist. Die Dimension ‚Emotional Support‘ betont demnach vor allem die grundsätzliche Bedeutung von sozial-emotionaler Sicherheit für die kindliche Entwicklung, sowie die konstruktive Gestaltung der Sozialbeziehungen sowie die Autonomieermöglichung. Nach Hamre et al. (2007) haben vor allem zwei Theorien entscheidend dazu beigetragen, dass die emotionale Unterstützung im Rahmen des CLASS als Aspekt der Unterrichtsqualität bzw. der Lehrer-Schüler-Interaktionen angesehen wird: einerseits Theorien zur Bindung (z. B. Ainsworth, Blehar, Waters & Wall, 1987) und andererseits die Selbstbestimmungstheorie (vgl. Deci & Ryan, 1993; Ryan & Deci, 2000). Es existieren Belege dahingehend, dass Schüler von Lehrpersonen, die emotionale Unterstützung anbieten, selbstbewusster und risi-

kofreudiger sind (vgl. Hamre & Pianta, 2007). Die Selbstbestimmungstheorie geht hingegen davon aus, dass Schüler motiviert werden, indem die Lehrpersonen deren Bedürfnisse nach Kompetenzerleben, sozialer Eingebundenheit und Autonomieerleben unterstützen.

„Classroom Organization“ oder auch „Management“ umfasst im CLASS einerseits Aspekte der Verhaltenskontrolle (Behavior Management). Dies beinhaltet, inwieweit die Lehrperson effektive Methoden zur Prävention von Fehlverhalten aufseiten der Schüler einsetzt. Auf der anderen Seite wird „Classroom Organization“ durch das Merkmal der Produktivität (Productivity) erfasst. Diese beinhaltet unter anderem, inwieweit die Lehrperson in der Lage ist, das Maximum an Unterrichtszeit und somit Lerngelegenheiten zur Verfügung zu stellen bzw. inwieweit Regeln und Routinen existieren, die die Unterrichtsabläufe organisieren. Des Weiteren impliziert „Classroom Organization“ den Aspekt Lernformate (Instructional Learning Formats). Mithilfe dieses Merkmals kann die Fähigkeit der Lehrperson in Bezug auf die abwechslungsreiche Gestaltung des Unterrichts und des eingesetzten Materials eingeschätzt werden.

Der dritte Faktor „Instructional Support“ wird im CLASS differenziert in „general“ (allgemein) und „content-specific“ (inhaltspezifisch) (vgl. Hamre & Pianta, 2007). „General Instructional Support“ impliziert Aspekte, die durchweg über alle Schuljahre und inhaltliche Bereiche relevant und beobachtbar sind, wie die Förderung des Problemlösens durch die Lehrperson (Concept Development), die Qualität des Feedbacks und die Qualität der Sprache, die im Klassenraum benutzt wird (Language Modeling). Das Merkmal „Concept Development“ beinhaltet zahlreiche Facetten der kognitiven Aktivierung aus der TIMSS-Videostudie (Klieme et al., 2001; Klieme, 2002). So kann „Concept Development“ dahingehend verstanden werden, dass die Lehrperson vorhandene Konzepte / Präkonzepte aufgreift und weiterentwickelt.

Die Qualität des Feedbacks (Quality of Feedback) fokussiert im CLASS einerseits das sogenannte instruktionale Feedback (Rückmeldungen zum konkreten Verhalten der Schüler in der Klasse) und andererseits Feedback in Form von konkreten Rückmeldungen zu dem Inhalt einer Schüleräußerung oder dem Lernprozess eines Schülers (vgl. Hamre & Pianta, 2007; Pianta et al., 2008). „Content-specific Instructional Support“ beinhaltet hingegen verschiedene Lehrstrategien in den einzelnen Fächern, beispielsweise Deutsch und Mathematik (vgl. Hamre & Pianta, 2007). Je nachdem in welcher Schulstufe sich die Schüler befinden, unterscheiden sich diese Lehrstrategien. Nach Hamre und Pianta (2007) ist es tendenziell unwahrscheinlich, dass junge Schüler durchgehend einem qualitativ hohen „Content-specific Instructional Support“ ausgesetzt sind. Sie gehen in ihrem Paper unter anderem darauf ein, dass es den Lehrpersonen in der Grundschule häufig an Fachwissen (Content Knowledge) mangelt oder sie in Mathematik und Naturwissenschaften fachfremd unterrichten, was es fast unmöglich macht, diese Domänen effektiv (im Sinne des „Content-specific Instructional Support“) zu unter-

richten. Dies ist unter anderem ein Grund für die geringe Ausprägung dieses Merkmals in der Grundschule.

Im Rahmen der Studie des National Institute of Child Health and Human Development Early Child Care Research Networks (vgl. NICHD ECCRN, 2002) können für das erste Schuljahr zwei Faktoren von Unterrichtsqualität bestätigt werden: emotionale Unterstützung (Emotional Support: Overcontrol, Intrusiveness, Detachment, Positive Emotional Climate, Negative Emotional Climate, Management, Sensitivity / Responsivity) und instruktionale Unterstützung (Instructional Support: Literacy Instruction, Evaluative Feedback, Instructional Conversation, Encourages Responsibility). Im Gegensatz zu den Studien von Pianta und Kollegen wurden Aspekte des ‚Classroom Managements‘ in dieser Studie zum ‚Emotional Support‘ gezählt: „because the description of that construct emphasized effective classroom management and skilled responses to children’s social and emotional behaviors“ (NICHD ECCRN, 2002, S. 378). Auf Basis dieser zwei Faktoren identifizierten Stuhlman und Pianta (2009) für das erste Schuljahr vier Profile von ‚Educational Quality (Types of Classrooms)‘. Die Autoren kamen zu dem Schluss, dass sich erste Klassen hinsichtlich des ‚Emotional Supports‘ und des ‚Instructional Supports‘ unterscheiden, wobei zusätzliche Kombinationen möglich sind. Übereinstimmend mit der Forschung charakterisieren sie einen typischen ‚early-educational classroom‘ durch „fairly positive social environment, and somewhat low levels of productivity and engagement in academics activities“ (Pianta & La Paro, 2003, S. 27).

Fasst man die zuvor dargestellten Konzeptualisierungen von Unterrichtsqualität in der Grundschule zusammen, so wird deutlich, dass für die Grundschule bislang keine einheitliche Operationalisierung von Unterrichtsqualität existiert. Jedoch kann in Anlehnung an das Konzept der drei Basisdimensionen die Bedeutung der Klassenführung und des Unterrichtsklimas auch für die Grundschule aufgezeigt werden. So wurden in allen hier aufgeführten Studien Merkmale der Klassenführung (oder des Klassenmanagements / des Managements oder der Classroom Organization) genannt (mit Ausnahme der KILIA-Studie). Auch Aspekte des Unterrichtsklimas sind in einem Großteil der Studien erhoben worden (jedoch mit unterschiedlichen Begrifflichkeiten: soziales Klima, schülerorientiertes sozial-emotionales Klima etc.).<sup>26</sup> Es kann demnach davon ausgegangen werden, dass auch in der Grundschule die Klassenführung und das Unterrichtsklima Grunddimensionen der Unterrichtsqualität darstellen. Große Ähnlichkeiten mit dem Konzept der drei Basisdimensionen aus der Sekundarstufe weisen die Konzeptualisierungen von Unterrichtsqualität aus der Studie von Hamre und Pianta (2007; Pianta et al., 2008) und dem FiS-Projekt (vgl. Eckerth et al., 2012; Hanke, 2007; Hein et al., 2010) auf. Kapitel 1.2.1 informierte bereits ausführlich über das Stufen-

26 Auf den Klimabegriff und dessen unterschiedliche Verwendung wird in Kapitel 4.1 noch einmal ausführlich eingegangen.

modell der Unterrichtsqualität von Pietsch et al., das unter anderem auf Basis von Unterrichtsbeobachtungen in der Grundschule entwickelt wurde (vgl. Pietsch, 2010).

*Erhebungsmethoden und -instrumente im Rahmen der Unterrichtsqualitätsforschung im Grundschulbereich:*

In den bisher dargestellten Studien aus dem Grundschulbereich wurden die Merkmale der Unterrichtsqualität mithilfe von Live-Beobachtungen oder durch Unterrichtstagebücher erfasst. Dies geht aus Tabelle 2 hervor. So wurden beispielsweise in der KILIA-Studie die Prädiktoren der Unterrichtsqualität (Freiheitsspielräume, Kooperationsmöglichkeiten, konstruktivistischer Unterricht) mithilfe eines Unterrichtstagebuches aus der Perspektive der Lehrperson erhoben. Im Gegensatz dazu schätzten sowohl in der SCHOLASTIK- und ECCE-Studie als auch im Kontext der Schulevaluation Hamburg trainierte Beobachter die Merkmale der Unterrichtsqualität hoch inferent ein.

Zur Analyse der Prozessqualität des mathematischen und schriftsprachlichen Anfangsunterrichts wurde im Rahmen des FiS-Projekts das Beobachtungssystem PQA (= Beobachtungsverfahren zur Erfassung der Prozessqualität des Anfangsunterrichts) entwickelt und eingesetzt (vgl. Eckert et al., 2012). Im Kontext des Teilprojekts 2 der BiKS-Forschergruppe (2009) kam ein umfangreiches hoch inferentes Beobachtungssystem zum Einsatz.<sup>27</sup>

Die Forschergruppe um R. C. Pianta entwickelte zur Erfassung der Unterrichtsqualität (mit Fokus auf Lehrer-Schüler-Interaktionen) das Beobachtungssystem ‚Classroom Assessment Scoring System‘ (CLASS, Pianta et al., 2008; vgl. auch Hamre & Pianta, 2007). Die Entwicklung des CLASS basiert auf Forschungsergebnissen, die bestätigen, dass die Interaktionen zwischen Schülern und Lehrpersonen als primärer Mechanismus der schulischen Entwicklung und des Lernens gelten. Das System wurde in den USA vielfach angewendet und ist auf das amerikanische Schulsystem ausgerichtet (Kindergartenklasse bis 6. Schuljahr). Die 3-Faktorenstruktur des CLASS (mit den Faktoren ‚Emotional Support‘, ‚Classroom Organization‘ und ‚Instructional Support‘) konnte anhand einer Stichprobe von mehr als 3.000 Klassen (von der Preschool bis zur fünften Klasse) bestätigt werden (vgl. Hamre, Mashburn, Pianta & Downer, 2006; zit. n. Pianta et al., 2008; Hamre et al., 2007). Im Rahmen des NICHD ECCRN (2002) wurden die Merkmale des ‚Emotional Supports‘ und des ‚Instructional Supports‘ durch Unterrichtsbeobachtungen mithilfe des COS-1 (Classroom Observation System for First Grade<sup>28</sup>) erhoben.

27 Dieses Beobachtungssystem zur Erfassung verschiedener Aspekte der Unterrichtsqualität steht mir zum Zeitpunkt der Fertigstellung der Dissertation nur als nicht publizierte Version zur Verfügung.

28 Es existieren vom COS auch Versionen für das dritte und fünfte Schuljahr (vgl. Hamre, Pianta, Mashburn & Downer, 2007). Der COS wurde speziell für die Nutzung im Rahmen der NICHD SECCYD entwickelt (SECCYD = Study of Early Child Care and Youth Development, vgl. Hamre & Pianta, 2007).

Tabelle 2: Erhebungsmethoden und -instrumente der aktuellen Unterrichtsqualitätsforschung im Grundschulbereich

Studie & Autoren	Erhebungsmethode
<i>Deutschsprachige Studien</i>	
Schulinspektion Hamburg: Pietsch (2008; 2010)	Unterrichtsbeobachtung
KILIA-Studie: Kammermeyer & Martschinke (2009)	Unterrichtstagebuch (Perspektive der Lehrperson)
ECCE-Study: Roßbach (2002a; b)	Unterrichtsbeobachtung
SCHOLASTIK-Studie: Helm- ke & Weinert (1997b)	Unterrichtsbeobachtung (Ausnahme: Klarheit wurde aus Schülersicht erhoben)
FIS-Projekt: Hanke (2007)	Unterrichtsbeobachtungen mithilfe des PQA (= Prozess- qualität im Anfangsunterricht)
Teilprojekt 2 der BIKS-Forscher- gruppe (2009)	Unterrichtsbeobachtung
<i>Internationale Studien</i>	
Hamre & Pianta (2007) / Pianta et al. (2008)	Unterrichtsbeobachtung mithilfe des CLASS-Beobach- tungssystems (Classroom Assessment Scoring System; Fokus auf Lehrer-Schüler-Interaktionen)
Stipek (1997, zit. n. Perry, Dono- hue & Weinstein, 2007)	Unterrichtsbeobachtung mit Early Childhood Classroom Observation Measure (ECCOM)
National Institute of Child Health and Human Development Early Child Care Research Net- work (NICHD ECCRN, 2002)	Unterrichtsbeobachtung mit COS-1 (Classroom Observation System for First Grade)

In der vorliegenden Arbeit werden die Unterrichtsqualitätsmerkmale Klassenführung und Unterrichtsklima jedoch nicht durch Live-Beobachtungen, sondern durch den Einsatz von Unterrichtsvideos erfasst. Diese videobasierte weist gegenüber der Live-Beobachtung relevante Vorteile auf.

*Vorteile der videobasierten Unterrichtsbeobachtungen gegenüber Live-Beobachtung durch einzelne Beobachter:*

In der Lehr-Lern-Forschung gilt besonders die Beobachtung von Unterrichts- und Lehr-Lern-Prozessen als wichtige Datenquelle und als „the most direct way to measure instructional quality“ (Clare, Valdés, Pascal & Steinberg, 2001, S. 2). Die externe Beobachtung von Unterricht (mit und ohne den Einsatz von Unterrichtsvideos) hat den Vorteil, dass Daten über den tatsächlich stattfindenden Unterricht generiert werden, die frei von Wahrnehmungen und Interpretationen der beteiligten Schüler und Lehrpersonen sind (vgl. Clausen, 2002; Seidel, 2003b). Vor allem Merkmale, die über Unterrichtsbeobachtungen generiert werden, weisen – im Vergleich zu mittels Fragebogen erfassten

Merkmale – bei der Erforschung von unterrichtsrelevanten Qualitätsmerkmalen größere Effektstärken in Bezug auf die Leistungen der Schüler auf (vgl. Seidel & Shavelson, 2007).

Verglichen mit der Live-Beobachtung, in der sich die Beobachter aufgrund der hohen Komplexität der Unterrichtsgeschehnisse und ihrer begrenzten Aufmerksamkeitsspanne nur auf einzelne wenige und zudem relativ leicht beobachtbare Aspekte beschränken können – diese müssen im Vorfeld der Beobachtung festgelegt werden –, können mithilfe von Unterrichtsvideos die komplexen Unterrichtsprozesse mehrmals und immer wieder aus unterschiedlichen Perspektiven und unter verschiedenen Fragestellungen beobachtet und analysiert werden (vgl. Jacobs, Kawanaka & Stigler, 1999; Petko, Waldis, Pauli & Reusser, 2003). Es muss demnach im Vorfeld keine Festlegung auf ein bestimmtes Auswertungsraster erfolgen. Besonders bei der Analyse von komplexen Segmenten des Unterrichtsgeschehens erfassen Videos im Vergleich zu anderen methodischen Zugängen aussagekräftigere Indikatoren beispielsweise für die Unterrichtsqualität (vgl. Ulewicz & Beatty, 2001). Die Möglichkeit der wiederholten Betrachtung und Re-Analyse des aufgezeichneten Unterrichts – aus unterschiedlichen Perspektiven und mit verschiedenen Fragestellungen – weist auf den „spezifischen Mehrwert von Unterrichtsvideos“ (Pauli & Reusser, 2006, S. 787) gegenüber der Live-Beobachtung hin. Nach Pauli und Reusser (2006) „ermöglichen Videoanalysen mit ihren Möglichkeiten kumulativer, zeitverzögerter Analysen ein perspektivenreiches und auf vielfältige Weise theoriegestütztes Vorgehen“ (S. 787f).

Allerdings sind Videostudien im Vergleich zur Live-Beobachtung mit sehr großem zeitlichen und kostenintensiven Aufwand verbunden, besonders wenn, wie im vorliegenden Fall, mehrere Unterrichtsstunden in verschiedenen Fächern in großer Zahl aufgenommen und analysiert werden. Letztendlich wird häufig argumentiert, dass Videoaufzeichnungen, die einmalig stattfinden, nur einen Ausschnitt der Realität darstellen. Nicht nur, dass das Verhalten von Lehrperson und Schülern in einem bestimmten Fach über die Zeit eine gewisse Variabilität aufweist (vgl. Bem & Allen, 1974), auch zwischen verschiedenen Fächern können Unterschiede vermutet werden (vgl. Borich & Klinzing, 1986). Neuere Studien zeigen jedoch, dass vor allem für die Einschätzung der Klassenführung und des Unterrichtsklimas eine aufgezeichnete Stunde hinreichend ist (vgl. Praetorius, Pauli, Reusser, Rakoczy & Klieme, 2013; vgl. Kapitel 17.3).

Im Bereich der Grundschulforschung kann jedoch erst in den letzten Jahren eine zunehmende Verwendung von videobasierten Erhebungsmethoden zur Erfassung von Unterrichtsmerkmalen festgestellt werden (vgl. Kapitel 2.2). Vor allem im Bereich der Unterrichtsqualitätsforschung wurde bislang hauptsächlich auf die Methode der Unterrichtsbeobachtungen durch im Vorfeld geschulte Beobachter zurückgegriffen, wie aus dem vorherigen Abschnitt bereits hervorgeht. Im Anschluss wird nun ein Überblick über diejenigen Studien aus dem Grundschulbereich gegeben, die mithilfe der Video-

analyse Merkmale des Unterrichts bzw. der Unterrichtsqualität erfasst haben. Dabei soll unter anderem dargestellt werden, inwieweit Studien existieren, die die drei Basisdimensionen guten Unterrichts – wie sie in der Sekundarstufe häufig erfasst werden – auch für den Grundschulbereich erhoben haben.

## 2.2 Videobasierte Studien zur Unterrichtsqualität in der Grundschule

Während der bisherige Kenntnisstand zur Analyse von Lehr-Lern-Prozessen im videografierten Unterricht hauptsächlich auf Erfahrungen aus der Sekundarstufe beruht, kann in den letzten Jahren eine Zunahme an Videostudien festgestellt werden, in denen vermehrt der Grundschulunterricht in den Fokus rückt (vgl. Tabelle 3). Wie aus Tabelle 3 hervorgeht, ist die Mehrzahl dieser Studien jedoch im Bereich des naturwissenschaftlichen Sachunterrichts angesiedelt (z. B. Blumberg, Möller & Hardy, 2004; Ewerhardy, 2010; Ewerhardy, Kleickmann & Möller, 2009; Kleickmann, Vehmeyer & Möller, 2010; Möller, Jonek, Hardy & Stern, 2002; Vehmeyer, 2009; Vehmeyer et al., 2007; Fauth, Decristan, Rieser, Klieme & Büttner, 2014).

Beispielsweise wurde im Rahmen des DFG-Schwerpunktprogramms BIQUA<sup>29</sup> die Videobeobachtung eingesetzt, um die Wirksamkeit von Fortbildungsmaßnahmen für Lehrpersonen im Bereich des naturwissenschaftlichen Sachunterrichts zu untersuchen. Im Kontext dieses Schwerpunktprogramms untersuchten Kleickmann et al. (2010), welche Bedeutung fachspezifische Vorstellungen vom Lehren und Lernen für den Einsatz von Scaffolding-Maßnahmen (operationalisiert über die Merkmale Hervorheben, Sequenzieren und kognitive Konflikte anregen) durch Lehrpersonen im Unterricht haben. Um den Einsatz von Scaffolding-Maßnahmen zu erfassen, wurde ein hoch inferentes Videorating durchgeführt. Zur Erfassung des Unterrichtshandelns der Lehrpersonen wurde ein weiteres hoch inferentes Ratingsystem entwickelt (vgl. Vehmeyer et al., 2007), welches die folgenden Aspekte beinhaltet: ‚Conceptual Change‘, Präkonzepte, eigene Ideen (eigene Formulierungen bzw. eigene Ideen zulassen), Ideen diskutieren, Motivation, anwendungsbezogenes Lernen, Laissez-faire-Haltung der Lehrperson und transmissives Lehrerverhalten.

Ähnlich wie im Schwerpunktprogramm BIQUA wurde auch im Rahmen der PLUS-Studie<sup>30</sup> der naturwissenschaftliche Sachunterricht mithilfe

29 BIQUA steht als Abkürzung für die **Bildungsqualität** von Schule.

30 PLUS steht als Abkürzung für **Professionswissen** von Lehrkräften, verständnisorientierter naturwissenschaftlicher Unterricht und Zielerreichung im Übergang von der Primar- zur Sekundarstufe. Das Projekt PLUS ist eingebettet in die DFG-Forschergruppe „naturwissenschaftlicher Unterricht“ (nwu) und wurde als Kooperationsprojekt zwischen der Ar-

von Unterrichtsvideos analysiert. So untersuchten Ewerhardy et al. (2009) die Verständnisorientierung von Lehrpersonen als Aspekt der Unterrichtsgestaltung durch Verwendung eines hoch inferenten Ratingsystems. Die Verständnisorientierung wurde über die vier Konstrukte Anregung von ‚Conceptual Change‘, Einbetten in natürliche, bedeutungsvolle Kontexte, ‚Scaffolding‘ sowie das Aushandeln von Bedeutungen und der Relevanz von Konzepten operationalisiert (vgl. auch Ewerhardy, 2010).

Während die oben aufgeführten Studien im Sachunterricht nur einzelne Kriterien guten Unterrichts anhand von Unterrichtsvideos untersuchten (‚Scaffolding‘ oder Verständnisorientierung), orientierte sich die Interventionsstudie IGEL<sup>31</sup> bei der Auswertung ihrer Videodaten aus dem naturwissenschaftlichen Sachunterricht an dem Konzept der drei Basisdimensionen guten Unterrichts. In Anlehnung an Studien aus dem Sekundarschulbereich wurden die drei Basisdimensionen ‚Classroom Management‘, unterstützendes Klima und kognitive Aktivierung nicht nur aus der Beobachterperspektive (sowohl Unterrichtsvideos als auch Live-Beobachtungen) erfasst, sondern zusätzlich aus Lehrer- und Schülersicht mithilfe von Fragebögen. Dies erlaubt unter anderem die Beantwortung der Frage, inwieweit bereits Schüleraussagen im dritten Schuljahr valide sind. Das IGEL-Projekt zielt zudem auf die Frage ab, inwieweit unterschiedliche Formen adaptiver Unterrichtsgestaltung (kognitive Strukturierung, ‚Peer Learning‘ und formatives Assessment) im Sachunterricht des dritten Schuljahres Auswirkungen auf die individuelle Förderung von Grundschulkindern haben.

Für die videobasierte Einschätzung der drei Unterrichtsqualitätsmerkmale muss jedoch einschränkend angemerkt werden, dass pro Basisdimension jeweils nur ein bis zwei Items von den Ratern hoch inferent eingeschätzt wurden. Das unterstützende Klima beinhaltet lediglich die Anerkennung durch die Lehrperson. Das ‚Classroom Management‘ wurde über das Ausmaß der Disziplinprobleme und die kognitive Aktivierung über die Exploration der Denkweisen und über die zum Denken herausfordernde Unterrichtsgestaltung erfasst (vgl. Fauth et al., 2014). Diese wenigen Items bilden jedoch nicht die Gesamtheit der drei Konstrukte ab. So beinhaltet das ‚Classroom Management‘ neben dem Ausmaß der Disziplinprobleme weitere reaktive und proaktive Elemente (vgl. Kapitel 3.1), die im Rahmen der IGEL-Studie nicht berücksichtigt wurden. Hierzu zählen unter anderem der Umgang mit Unterrichtsstörungen oder Disziplinproblemen sowie die Existenz eines klaren Regelsystems und von Ritualen oder die Vorbereitung des Klassenraums (vgl. Kapitel 3.2). Neben der Anerkennung durch die Lehrperson zeichnet sich ein

---

beitsgruppe um Prof. Dr. H. E. Fischer in Essen und der Arbeitsgruppe um Prof<sup>in</sup> Dr. K. Möller in Münster durchgeführt.

31 IGEL steht als Abkürzung für Individuelle Förderung und adaptive Lern-**G**elegenheiten in der Grundschule. (vgl. <http://www.idea-frankfurt.de/wissen/projekte/projekt-igel>, Zugriff: 05.12.2012)

unterstützendes Klima zusätzlich auch beispielsweise durch den positiven Umgang mit Fehlern oder die gegenseitige Anerkennung der Schüler aus (vgl. Kapitel 4.2).

Als eine Studie, die unabhängig vom Fach Aspekte des ‚Classroom Managements‘ (Regeln, Prozeduren, Rituale, Raumgestaltung, Rhythmisierung, Überblick behalten) in der Schuleingangsstufe in der Schweiz (Kindergarten und Unterstufe) untersucht, muss die von Wannack, Arnaldi und Schütz (2009) genannt werden. Mithilfe von videobasierter Unterrichtsbeobachtung gingen die Autorinnen in Anlehnung an Evertson, Emmer und Worsham (2006) der Frage nach, welche Elemente des ‚Classroom Managements‘ Lehrpersonen im Kindergarten und in den ersten Jahren der Schuleingangsstufe einsetzen, um die Spiel- und Lernaktivitäten der Schüler zu gestalten (vgl. Wannack, 2012). Die Stichprobe setzt sich aus zwölf Lehrpersonen zusammen, die gebeten wurden, sowohl eine offene als auch eine geführte Unterrichtssequenz zu zeigen. Das Videomaterial umfasst insgesamt 19 Stunden (vgl. Wannack, Herger & Barblan, 2011). Mithilfe zusätzlicher Interviews wurde im Rahmen dieser Studie ein vorläufiges Modell erarbeitet, „das in detaillierter Art und Weise das Classroom Management und damit grundlegende Dimensionen der Unterrichtsgestaltung [...] beschreibt“ (Wannack, 2012, S. 72).

Tabelle 3: Überblick über videobasierte Studien in der Grundschule (Fach, Merkmale der Unterrichtsqualität und Erhebungsmethode)

Studie bzw. Autoren	Fach	Merkmale der Unterrichtsqualität	Erhebungsmethode (niedrig vs. hoch inferent)
BIQUA: Kleickmann et al. (2010)	Sachunterricht	Scaffolding-Maßnahmen (Hervorheben, Sequenzieren, kognitive Konflikte anregen)	hoch inferent
BIQUA: Vehmeyer et al. (2007)	Sachunterricht	Unterrichtshandeln: ‚Conceptual Change‘, Präkonzepte, eigene Ideen (eigene Formulierungen bzw. eigene Ideen zulassen), Ideen diskutieren, Motivation, anwendungsbezogenes Lernen, Laissez-faire, transmissives Lehrerverhalten	hoch inferent
PLUS-Studie: Ewerhardy et al. (2009, vgl. auch Ewerhardy, 2010)	Sachunterricht	Verständnisorientierung (Anregung von ‚Conceptual Change‘, Einbetten in natürliche, bedeutungsvolle Kontexte, ‚Scaffolding‘, Aushandeln von Bedeutungen, Relevanz der Konzepte)	hoch inferent
IGEL-Studie: (Fauth et al., 2014)	Sachunterricht	Classroom Management, unterstützendes Klima und kognitive Aktivierung	hoch inferent

– Fortsetzung nächste Seite –

Studie bzw. Autoren	Fach	Merkmale der Unterrichtsqualität	Erhebungsmethode (niedrig vs. hoch inferent)
Connor et al. (2010)	Lese- und Schreibunterricht	Classroom Management (Zeitnutzung)	niedrig inferent
BiTe-Studie: Ohle et al. (2012)	Deutsch	kognitive Aktivierung, Lernunterstützung, Motivierung, Strukturierung und (meta)kognitive Strategieunterstützung, Klassenmanagement	k. A.
VERA-Studie: Helmke et al. (2008)	Deutsch	Zeitnutzung	niedrig inferent
VERA-Studie: Kleinbus (2010)	Deutsch	Kompetenzorientierung, Adaptivität, Klarheit und Strukturiertheit, Aktivierung im Sinne aktiver Teilhabe und Motivierung	hoch inferent
VERA-Studie: Praetorius et al. (2012)	Deutsch	Classroom Management, 'Personal Learning Support'	hoch inferent
Wannack, Arnaldi & Schütz (2009; vgl. Wannack, Herger & Barblan, 2011)	Fachunspezifisch	Classroom Management (Regeln, Prozeduren, Rituale, Raumgestaltung, Rhythmisierung, Überblick behalten) sowie didaktisch-methodische Aspekte und Spiel- und Lernbegleitung	hoch inferente (Sequenzkodierung)
ALPHA-Studie: Kucharz (2012)	Fachunspezifisch	Instruktionseffizienz, Schülerorientierung, kognitive Aktivierung, Klarheit und Strukturiertheit	hoch inferent
Lernberatung im offenen Unterricht: Kucharz et al. (2009)	Fachunspezifisch	Instruktionseffizienz, Schülerorientierung, kognitive Aktivierung, Klarheit und Strukturiertheit	hoch inferent

Eine weitere Studie, die unabhängig vom Fach Unterrichtsvideos zur Analyse von Grundschulunterricht einsetzt, soll an dieser Stelle zusätzlich erwähnt werden. Das einjährige Teilprojekt „Lernberatung im offenen Unterricht“ (vgl. Kucharz, Schnebel & Helming, 2011) der Forschergruppe WOLLF (= Wirksamkeit Offener Lehr-Lern-Formen) zielte darauf ab, Erkenntnisse über Quantität und Qualität von Lernunterstützung im offenen Unterricht der Grundschule mittels Videos zu erlangen. Die Studie konzentrierte sich dabei auf das Angebot der Lehrpersonen während Einzelinteraktionen und verglich anhand einer Stichprobe von sechs Grundschulklassen unter anderem die Verteilung der Inhalte der Lehrer-Schüler-Interaktionen in den schülerzentrierten mit denen in lehrerzentrierten Phasen. Im Rahmen eines hoch inferenten Ratings wurde erfasst, ob es der Lehrperson gelingt, eine Lernumgebung zu schaffen, in der selbstgesteuerte Lernprozesse der Schüler möglich sind. Der Fokus lag hierbei auf einer Einschätzung der Instruktions-

effizienz (z. B. Umgang mit der Unterrichtszeit, Qualität der Organisation), der Schülerorientierung (z. B. Umgang mit Fehlern, Motivierungsfähigkeit), der kognitiven Aktivierung (z. B. genügend zur Verfügung stehende Unterrichtszeit) sowie der Klarheit und Strukturiertheit (z. B. sprachliche Qualität, Gesprächsführung)<sup>32</sup>. Es ergaben sich vergleichsweise große Unterschiede auf der Ebene der Qualitätseinschätzungen zwischen den sechs Klassen, während die Einschätzungen für die jeweilige Grundschulklasse stabil ausfielen (vgl. Kucharz, Küster, Müller & Schnebel, 2009).

*Der Einsatz von Videografie in sprachlichen Fächern:*

Im Rahmen der videobasierten Forschung sind die sprachlichen Fächer bislang eher unterrepräsentiert, sowohl im Sekundar- als auch im Grundschulbereich. In den letzten Jahren wurden beispielsweise nur vereinzelt Videostudien im Fremdsprachenunterricht für die Sekundarstufe realisiert (z. B. Göbel, 2007; Helmke, Helmke, Kleinbub, Nordheider, Schrader & Wagner, 2007; Helmke, Helmke, Schrader, Wagner, Klieme, Nold & Schröder, 2008). Im Fach Deutsch zielt der Einsatz von Unterrichtsvideos derzeit eher auf die linguistische Analyse von Unterrichtskommunikation ab (zusf. Kleinbub, 2010). So bildet beispielsweise den Kern der Videografie bei Bremerich-Vos (2006) der Grammatikunterricht in den Jahrgangsstufen fünf bis sieben mit dem Fokus auf Wortarten und Satzgliedern im einfachen und zusammengesetzten Satz. Becker-Mortzek (2008) sieht in der Videobeobachtung „eine geeignete Methode von Sprachaneignungsprozessen sowie der Modellierung und Messung sprachlicher Fähigkeiten und somit zur Konturierung der Gesprächskompetenz“ (zit. n. Kleinbub, 2010, S. 119). Auch in einer jüngeren Studie von Föh (2005) steht die Untersuchung mündlicher Kommunikation und Kooperation bei der gemeinsamen Arbeit am Computer im Deutschunterricht im Vordergrund der videobasierten Analysen.

Ähnlich wie für den Sekundarschulbereich existieren auch für den Grundschulbereich bislang kaum videobasierte Analysen zur Unterrichtsqualität im Deutschunterricht (vgl. Tabelle 3). Zum Beispiel widmen sich Stude und Ohlhus (2005) im Rahmen ihrer Videostudie lediglich der Analyse schulischer Schreibprozesse. Allerdings wurden als Datengrundlage nur Videoaufnahmen aus zwei dritten Klassen herangezogen. Diese und weitere Studien zeugen nach Kleinbub (2010) davon, „dass viele beobachtungs-basierte Analysen im Umfeld der Deutschdidaktik im Gegensatz zu denen in der pädagogisch-psychologischen Unterrichtsforschung eher in Einzel- denn in Großprojekten

32 Die gleichen Merkmale wurden auch im Rahmen der ALPHA-Studie verwendet (vgl. Kucharz, 2012). Die Längsschnittstudie „ALPHA – Unterrichtskompetenz im Berufseinstieg“ untersucht die Entwicklung von Unterrichtskompetenz im Vorbereitungsdienst bzw. im ersten Berufsjahr von Grundschullehrpersonen. Die Studie wird gefördert durch die Internationale Bodensee Hochschule (vgl. <http://www.bodenseehochschule.org/alpha-unterrichtskompetenz/> Zugriff: 05.12.2012).

organisiert sind“ (S. 120). Außerdem werden im Bereich der Deutschdidaktik die entstandenen Unterrichtsvideos- und Transkripte häufig mit qualitativen Methoden ausgewertet (vgl. Kleinbub, 2010). Vor allem mangelt es in diesem Zusammenhang an curricular vergleichbaren Videostudien.

Als eine Ausnahme für das Fach Deutsch gilt die Studie VERA<sup>33</sup> – Gute Unterrichtspraxis (vgl. Helmke et al., 2008; Kleinbub, 2010; Praetorius et al., 2012). Gegenstand der Forschungen im Rahmen der VERA-Studie sind unter anderem die Erscheinungsformen und Bedingungen der Zeitznutzung (vgl. Helmke et al., 2008) als ein zentrales Merkmal der Klassenführung und als Schlüsselmerkmal des Unterrichtserfolgs (vgl. Helmke, 2007). Kleinbub (2010) fokussiert in ihrer Dissertation, die im Rahmen der VERA-Studie entstanden ist, lediglich den Leseunterricht und dessen Qualität, wobei der Fokus auf den folgenden fünf Unterrichtsqualitätsmerkmalen lag: Kompetenzorientierung, Adaptivität, Klarheit und Strukturiertheit, Aktivierung im Sinne aktiver Teilhabe und Motivierung. Sogenannte „dominant generische Merkmale“ (Kleinbub, 2010, S. 88) wie die Klassenführung, die kognitive Aktivierung oder das lernförderliche Klima wurden jedoch nicht berücksichtigt, da der Fokus eher auf einer Konkretisierung deutschdidaktischer Merkmale der Unterrichtsqualität lag. Im Rahmen einer Re-Analyse nutzten Praetorius et al. (2012) die Videos der VERA-Studie, um Aspekte des ‚Classroom Managements‘ und des ‚Personal Learning Supports‘ für das Fach Deutsch (viertes Schuljahr) hoch inferent einzuschätzen.

Im Kontext der BiTe-Studie<sup>34</sup> werden derzeit in etwa 48 Grundschulen drei Unterrichtsstunden zur Bild-Text-Integration videografiert. Diese Stunden werden anschließend u. a. im Hinblick auf bild-text-spezifische Unterrichtsaktivitäten untersucht. Die Studie geht der Frage nach, welchen Einfluss die Lehrerkompetenzen, vermittelt über die Unterrichtsqualität auf die Kompetenzentwicklung der Schüler haben. Unterrichtsqualität wird hier operationalisiert über die Merkmale kognitive Aktivierung, Lernunterstützung, Motivierung, Strukturierung und (meta)kognitive Strategieunterstützung sowie Klassenmanagement (vgl. Ohle et al., 2012). Ergebnisse zur Ausprägung der Unterrichtsqualität und deren Einfluss auf die bild-text-spezifischen Schülerkompetenzen liegen allerdings noch nicht vor.

---

33 VERA steht als Abkürzung für **V**ergleichsarbeiten in der Grundschule. Details zum Design und eine ausführliche Projektbeschreibung der Studie finden sich bei Helmke, Helmke, Heyne, Hosenfeld, Schrader und Wagner (2010).

34 BiTe steht als Abkürzung für **B**ild-**T**ext-Integration (Titel der Studie: „Entwicklung und Überprüfung von Kompetenzmodellen zur integrativen Verarbeitung von Texten und Bildern“). BiTe ist ein Kooperationsprojekt der Universität Koblenz-Landau (Prof. Dr. Wolfgang Schnotz), des Max-Planck-Instituts für Bildungsforschung, Berlin (Prof. Dr. Jürgen Baumert), der Goethe-Universität Frankfurt (Prof. Dr. Holger Horz) und der Technischen Universität Dortmund (Prof. Dr. Nele McElvany; <http://www.ifs-dortmund.de/1499.html> Zugriff: 05.12.2012). Die Lehrpersonen erhielten für die Stunden Informationen zu Bild-Text-Integration, Unterrichtszielen sowie Unterrichtsmaterial zum Thema „Südamerika“.

Connor et al. (2010) konnten im Rahmen ihrer Studie zeigen, dass Lehrpersonen, die im Vorfeld an einer Intervention teilgenommen haben, ein besseres ‚Classroom Management‘ zeigen als Lehrpersonen ohne entsprechende Intervention. Das ‚Classroom Management‘ wurde in dieser Studie über eine effiziente Zeitnutzung operationalisiert. Lehrpersonen mit gutem ‚Classroom Management‘ verschwenden wenig Zeit für Unterbrechungen, Disziplinstörungen und Übergänge und realisieren zusätzlich einen größeren Anteil an Gruppenarbeitsphasen im Sinne eines eher individualisierten Unterrichts. Zur Erfassung des ‚Classroom Managements‘ wurden niedrig inferente Kodierungen während eines ganzen Schuljahres (drei Messzeitpunkte) durchgeführt. Der Fokus lag hier auf dem Lese- und Schreibunterricht im ersten Schuljahr.

*Der Einsatz von Videografie im Fach Mathematik:*

Im Fach Mathematik wurden m. W. n. bislang keine videobasierten Studien im Anfangsunterricht durchgeführt, die unter curricular vergleichbaren Bedingungen Merkmale guten Unterrichts erfassen (vgl. Tabelle 3). Im Rahmen der Studie von Fetzer (2007) wurden zwar über drei Jahre (1. bis 3. Schuljahr) hinweg Episoden aus dem Mathematikunterricht einer Klasse videografiert, jedoch erfolgte die Auswertung der Videos lediglich aus argumentationstheoretischer Perspektive. Es sollte die Frage beantwortet werden, wodurch sich das Argumentieren von Grundschulern im Mathematikunterricht charakterisieren lässt.

*Vorteil der Standardisierung des Unterrichtsinhalts in Videostudien:*

Die zuvor dargestellten Studien aus dem Bereich des naturwissenschaftlichen Sachunterrichts in der Grundschule wurden unter standardisierten Richtlinien aufgezeichnet (vgl. Blumberg et al., 2004; Ewerhardy et al., 2009; Kleickmann et al., 2010; Vehmeyer et al., 2007; Fauth et al., 2014). Dazu zählt u. a., dass im Sinne einer quasiexperimentellen Kontrolle der Unterrichtsinhalt während der aufgezeichneten Stunden festgelegt wurde.<sup>35</sup> Im Rahmen des DFG-Schwerpunktprogramms BIQUA wurden die teilnehmenden Lehrpersonen zusätzlich gebeten, dass es sich bei der aufgezeichneten Unterrichtseinheit um eine Einführungsstunde handeln sollte. Um auch die materielle Ausstattung konstant zu halten, wurden den Lehrpersonen zusätzlich Materialboxen und eine Lehrerhandreichung für die Durchführung der Unterrichtseinheit zur Verfügung gestellt (vgl. Kleickmann et al., 2010). Im Gegensatz dazu konnten die Lehrpersonen, die an der PLUS-Studie teilgenommen haben, ihren aufgezeichneten Unterricht zum Thema Aggregatzustände

35 Sowohl im IGEL-Projekt (vgl. Fauth et al., 2014) als auch im DFG-Schwerpunktprogramm BIQUA (vgl. Kleickmann et al., 2010; Vehmeyer, Kleickmann & Möller, 2007) wurde in den Stunden das Thema „Schwimmen und Sinken“ behandelt. In der Studie von Ewerhardy und Kollegen (2009, PLUS-Studie) lernten die Schüler etwas über „Aggregatzustände und ihre Übergänge am Beispiel Wasser“.

und ihre Übergänge<sup>36</sup> am Beispiel Wasser selbstständig planen und durchführen (vgl. Ewerhardy et al., 2009). Auch im Projekt IGEL wurde den teilnehmenden Lehrpersonen das Unterrichtsmaterial in Form der Klassenkiste zum Schwimmen und Sinken zur Verfügung gestellt (vgl. Fauth et al., 2014).

Die Standardisierung der Unterrichtsinhalte, des Arbeitsmaterials oder von methodisch-didaktischen Maßnahmen ermöglicht eine Generalisierung von Aussagen über die teilnehmenden Klassen hinweg und stellt eine gewisse Vergleichbarkeit zwischen den Klassen her (vgl. Klieme et al., 2006). Die Frage, inwieweit eine solche Standardisierung Auswirkungen auf die Gestaltung alltäglichen Unterrichts hat, bleibt allerdings unbeantwortet.

Für die Fächer Deutsch und Mathematik können auf Basis der bisherigen videobasierten Forschung keine generalisierbaren Aussagen über die Ausgestaltung der Unterrichtsqualität für die Grundschule getätigt werden. Hier mangelt es an entsprechenden Studien, in denen Unterrichtsvideos unter curricular vergleichbaren Bedingungen aufgezeichnet wurden. An diesem Punkt setzt die PERLE-Studie an. Sowohl für das Fach Deutsch (1. Schuljahr) als auch für das Fach Mathematik (2. Schuljahr) wurden die Unterrichtsinhalte standardisiert. Im Fach Deutsch (1. Schuljahr) wurde den teilnehmenden Lehrpersonen zudem einheitliches Arbeitsmaterial in Form eines Bilderbuchs zur Verfügung gestellt (vgl. Kapitel 9).

Da für den Anfangsunterricht wenige videobasierte Studien existieren, auf deren Basis generell Aussagen über die Ausprägung der Unterrichtsqualitätsmerkmale Klassenführung und Unterrichtsklima gemacht werden können, werden im Anschluss zusätzlich Studien herangezogen, die Merkmale der Unterrichtsqualität durch den Einsatz geschulter Beobachter während des Unterrichts untersucht haben. Die videobasierte Beobachtung kann als Form der Beobachtung angesehen werden (vgl. Bortz & Döring, 2006), womit Rückschlüsse auf die Ausprägung der Unterrichtsqualitätsmerkmale in den Videos grundsätzlich möglich sind.

### **2.3 Befunde zur Ausprägung der Unterrichtsqualität im Anfangsunterricht**

Über die tatsächliche Ausgestaltung der Unterrichtsqualität im Anfangsunterricht liegen für Deutschland nur sehr wenige Befunde vor (z. B. Eckerth et al., 2012; Helmke & Renkl, 1993; Helmke & Weinert, 1997b; Kammermeyer & Martschinke, 2009; Roßbach, 2002a). Auf die Ergebnisse dieser Stu-

---

36 Mit dem Begriff Übergang ist hier der Übergang vom flüssigen in den gasförmigen Zustand (Verdunstung) oder vom gasförmigen in den flüssigen Zustand (Kondensation) gemeint.

dien wird anschließend eingegangen. Zusätzlich lohnt sich erneut ein Blick auf US-amerikanische Forschung. So kann für amerikanische Grundschulklassen bestätigt werden, dass sich ein Großteil der Klassen vor allem durch hohe Werte auf den Aspekten ‚Teacher Sensitivity‘ und ‚Positive Emotional Support‘ auszeichnet (vgl. Hamre & Pianta, 2007; NICHD ECCRN, 2002). Eher selten zeigten die Lehrpersonen in diesen Studien distanzierteres Verhalten bzw. es herrschte generell ein eher positives Klima. So schätzten Beobachter im Rahmen der NICHD ECCRN Studie (vgl. NICHD ECCRN, 2002) unter anderem Aspekte der Klassenführung und des Unterrichtsklimas für das erste Schuljahr durchweg sehr positiv ein. Mithilfe des COS-1 (Classroom Observation System for First Grade, vgl. Kapitel 2.1), dem eine siebenstufige Skala zugrunde liegt, wurde beispielsweise die ‚Sensitivity / Responsivity‘ der Lehrperson als ein Aspekt der Lehrer-Schüler-Beziehung eingeschätzt und erreichte im Mittel einen Wert von 5.33 (vgl. auch Hamre & Pianta, 2007). Auch für das Klima ( $MW = 5.36$ ) und das ‚Classroom Management‘ ( $MW = 5.03$ ) erhielten die Klassen ähnlich hohe Werte. Perry et al. (2007) bestätigen für das erste Schuljahr ähnlich hohe Ausprägungen des ‚Instructional Climate‘, ‚Social Climate‘ und ‚Management‘. Im Rahmen der Studie schätzten geschulte Beobachter mithilfe des ECCOM (Early Childhood Classroom Observation Measure, vgl. Stipek, 1997, zit. n. Perry et al., 2007) die drei Aspekte auf einer fünfstufigen Skala ein, wobei ‚Management‘ im Mittel am höchsten eingeschätzt wurde ( $MW = 3.18$ ), gefolgt vom ‚Instructional Climate‘ ( $MW = 3.12$ ) und dem ‚Social Climate‘ ( $MW = 3.04$ ). Große Unterschiede in den Ausprägungen zwischen den Klassen konnten die Autorinnen für das erste Schuljahr nicht feststellen.

Für deutsche Grundschulklassen bestätigen unter anderem Untersuchungen im Rahmen des FiS-Projekts diese eher positive Ausprägung der Merkmale der Unterrichtsqualität (vgl. Eckerth et al., 2012). Mithilfe von Unterrichtsbeobachtungen kommen die Autorinnen zu dem Ergebnis, dass die Merkmale Klassenführung und individuelle Unterstützung im Vergleich zu den beiden Merkmalen Strukturierung und kognitive Aktivierung durchschnittlich höher ausgeprägt sind, wobei die Werte vom ersten Schuljahr bis zum Ende des zweiten Schuljahres nochmal signifikant ansteigen. Es konnten keine Unterschiede in der Ausprägung der Merkmale Klassenführung und individuelle Unterstützung im schriftsprachlichen und mathematischen Anfangsunterricht bestätigt werden. Im mathematischen Anfangsunterricht wird die kognitive Aktivierung im Vergleich zum schriftsprachlichen Anfangsunterricht jedoch häufig höher eingeschätzt. Bezüglich der Ausprägung der Merkmale ergaben sich deutliche Unterschiede zwischen den zwei Lerngruppen (jahrgangshomogen versus jahrgangsübergreifend).

Für das zweite Schuljahr präsentiert sich die Befundlage uneinheitlich. Während die externen Beobachter in der Studie von Roßbach (2002a; b) die Klassenführung auf Basis einer fünfstufigen Skala deutlich positiver beurteilen als das schülerorientierte sozial-emotionale Klima, berichten Eckerth et

al. (2012) ähnlich hohe Ausprägungen der Klassenführung und der individuellen Unterstützung für das zweite Schuljahr, sowohl für den mathematischen als auch den schriftsprachlichen Bereich. Helmke und Renkl (1993) ließen unter anderem die Klassenführung und das Sozialklima von geschulten Ratern hoch inferent einschätzen und zeigen für beide Unterrichtsvariablen ähnlich hohe Ausprägungen auf einer fünfstufigen Skala.

In der Studie von Roßbach (2002a; b) wurden externe Unterrichtsbeurteilungen unter anderem auf Grundlage des Ratingbogens aus dem SCHOLASTIK-Projekt durchgeführt. Bei einer Faktoranalyse ergaben sich fünf Faktoren der Unterrichtsqualität (Klassenmanagement, schülerorientiertes sozial-emotionales Klima, direktes Unterrichten, Anwendungsbezug und Unterstützung bei Stillarbeit). Sowohl das Klassenmanagement als auch die Unterstützung bei Stillarbeit erreichten Werte über der theoretischen Mitte von 3,0, während das direkte Unterrichten, der Anwendungsbezug und das schülerorientierte sozial-emotionale Klima darunter lagen. Vor allem der Wert von 2,3 (Skala 1–5) für das schülerorientierte sozial-emotionale Klima war aus Sicht des Autors überraschend angesichts der übereinstimmenden Meinung, dass das Klima in der Grundschule als sehr bedeutsam erachtet wird. Für das dritte und vierte Schuljahr berichten Helmke und Schrader (1997) im Rahmen des SCHOLASTIK-Projekts unter Anwendung desselben Beobachtungsverfahrens zudem deutlich höhere Werte.

Bei näherer Betrachtung lässt sich jedoch in Bezug auf die Operationalisierung der Skala Klassenmanagement bei Roßbach (2002a; b) feststellen, dass diese neben dem Übergangmanagement und dem Vorhandensein eines klaren Regelsystems Aspekte des Unterrichtsklimas umfasst. Dazu gehört, inwieweit die Lehrperson sich um ein Klima wechselseitigen Respekts und sozialer Unterstützung des Lernens bemüht und schwächere Schüler nicht entmutigt werden. Die Skala schülerorientiertes sozial-emotionales Klima beinhaltet hingegen vor allem Aspekte der Wertschätzung. Demnach sollte bei der Ausprägung des Unterrichtsklimas berücksichtigt werden, ob Aspekte der emotionalen Unterstützung, die Fürsorglichkeit der Lehrperson oder Aspekte der Anerkennung und Wertschätzung beurteilt werden. In Anlehnung an Roßbach (2002a; b) können, je nachdem auf welchem Aspekt der Fokus liegt, Unterschiede in der Höhe der Ausprägung angenommen werden.

Im Vergleich zu Sekundarschulen erreichen Grundschulen allgemein bessere Bewertungen hinsichtlich der Merkmale der Unterrichtsqualität (Pietsch, 2010). Vor allem auf einzelnen Items des Klassenmanagements und des Klassenklimas sowie der Schülerorientierung und Unterstützung bestätigten sich statistisch signifikante Unterschiede, die 20 % einer Standardabweichung oder mehr betragen. Zusätzlich zeigten sich Unterschiede auf einzelnen Items, die den Merkmalen Motivieren, intelligentes Üben und aktives Lernen zugeordnet sind.

### 3 Klassenführung als Basisdimension der Unterrichtsqualität

Wie aus den vorherigen Kapiteln deutlich wurde, gilt eine effektive Klassenführung als ein fachübergreifend gültiges Prozessmerkmal des Unterrichts (vgl. Helmke et al., 2010) und als eine „Basisdimension qualitativ vollen Unterrichts“ (Kunter et al., 2005, S. 504; vgl. Kapitel 1.2). Das bedeutet, dass die Effektivität der Klassenführung als relativ unabhängig vom Fach betrachtet werden kann, gerade in der Grundschule (vgl. Martschinke, Kammermeyer, Frank & Mahrhofer, 2002), da hier das Fächerprinzip noch nicht so praktiziert wird, wie in der Sekundarschule. Nach Weinert (1996) kommt der Klassenführung eine „Schlüsselfunktion im Unterricht“ (S. 124) zu. Im Folgenden wird kurz der Begriff der Klassenführung geklärt.<sup>37</sup> Im Anschluss daran befasst sich das Kapitel 3.2 mit einer ausführlichen Darstellung der Merkmale der Klassenführung, die im Rahmen der vorliegenden Arbeit von Bedeutung sind.

#### 3.1 Klassenführung – eine Begriffsklärung

Eine effektive Klassenführung (oder Classroom Management<sup>38</sup>) betrifft unter anderem die erforderlichen Rahmenbedingungen eines lernwirksamen Unter-

37 Ophardt und Thiel (2008) differenzieren vier unterschiedliche Forschungszugänge zum Klassenmanagement. Neben dem Klassenmanagement als Basisdimension des Unterrichts (1) unterscheiden sie theoretisch zwischen (2) dem Klassenmanagement als individuelle Verhaltensmodifikation (behavioristischer Zugang); (3) dem Klassenmanagement als Steuerung von Aktivitätsstrukturen (ökologischer Zugang) und (4) dem Klassenmanagement als gezielte Gestaltung sozialer Verhaltenserwartungen (handlungstheoretisch basierter Zugang). Die Herstellung von Ordnung im Unterricht wird aus der behavioristischen Perspektive vor allem „als Förderung erwünschten und Einschränkung unerwünschten Verhaltens durch den Einsatz spezifischer Stimuli definiert“ (S. 265), wobei sich verschiedene Operationen der Verhaltensmodifikation unterscheiden lassen (z. B. Belohnung versus Bestrafung). Aus der Perspektive des ökologischen Zugangs wird die Herstellung von Ordnung als proaktive, konstruktive Steuerung von Aktivitäten betrachtet. Für die Etablierung von Ordnung sind hier beispielsweise reibungslose Übergänge, der Einsatz von Signalen oder der Aspekt des ‚Monitorings‘ bedeutsam. Bei dem handlungstheoretisch basierten Zugang rückt die Lehrperson als Gestalter sozialer Verhaltenserwartungen in den Vordergrund, indem von ihr u.a. Regeln aufgestellt und implementiert werden. Die Gestaltung bezieht sich zudem auf Prozeduren bzw. Routinen, die vor allem für die Durchführung von Unterrichtsstunden bedeutsam sind.

38 Im deutschsprachigen Raum wird der Begriff ‚Classroom Management‘ häufig mit Klassenführung übersetzt. Nach Hennemann und Hillenbrand (2010) wird diese Übersetzung allerdings kontrovers diskutiert. So wird beispielsweise von Helmke (2009) die verengte Fokussierung der Klassenführung auf den Umgang der Lehrperson mit Unterrichtsstün-

richts (vgl. Helmke et al., 2010). Die Herstellung solcher Rahmenbedingungen geschieht durch die frühzeitige Etablierung eines Regelsystems und die konsequente Anwendung dieser Regeln sowie durch den Aufbau von Handlungsroutinen oder die Einführung von Ritualen. Im Kern geht es darum, unerwünschtes Schülerverhalten im Sinne von Unterrichts- und Disziplinstörungen frühzeitig und effektiv zu unterbinden und so die aktive Lernzeit zu sichern. Eine effektive Klassenführung kann allgemein beschrieben werden als die

Fähigkeit einer Lehrperson, im Unterricht einen zügigen, gradlinigen und gut organisierten Stundenablauf zu gewährleisten, Disziplinprobleme schnell und effektiv abzuhandeln sowie die Klassenzimmerprozesse im Gesamten zu überblicken (die sprichwörtlichen „Augen im Hinterkopf“), um bei Schwierigkeiten und Problemen rasch und im Idealfall vorausschauend handeln zu können (Waldis, Grob, Pauli & Reusser, 2010b, S. 177).

Waldis et al. (2010b) verfolgen mit dieser Definition einen integrativen Forschungsansatz. Demnach zeichnet sich eine effektive Klassenführung nicht nur durch reaktive, sondern auch durch vorausschauende und präventive Handlungsweisen der Lehrpersonen im Umgang mit Unterrichtsstörungen aus. In diesem Sinne umfasst die Klassenführung präventive, proaktive und reaktive Elemente (vgl. Helmke, 2009). Forschungsbefunde zur Wirksamkeit von Lehrerreaktionen auf Unterrichtsstörungen bestätigen in diesem Zusammenhang, dass präventive Maßnahmen wirksamer sind als reaktive (vgl. Brophy & Good, 1986).

Als proaktive Kriterien gelten nach Hennemann und Hillenbrand (2010) unter anderem die Vorbereitung des Klassenraums, die Festlegung von Konsequenzen, aber auch das Schaffen eines positiven (Lern-)Klimas. In der Grundschule gelten vor allem die Festlegung von Regeln zum Schuljahresbeginn und deren gewissenhafte Einhaltung in den ersten Wochen bis hin zur Einführung von Routinen als wichtige Maßnahmen, um Störungen vorzubeugen. Zur Intervention von Unterrichtsstörungen sollte die Lehrperson zusätzlich über reaktive Strategien verfügen, z. B. Strategien zum Unterbinden unangemessenen Schülerverhaltens oder zum Reagieren auf potenzielle Probleme im Unterricht (vgl. Tabelle 4).

Helmke (2009) unterscheidet neben diesem integrativen drei weitere Ansätze, die an dieser Stelle jedoch nicht weiter ausgeführt werden: Klassenführung (1) als Klassenlehrer, (2) als Inbegriff erfolgreichen Unterrichtens und Führens sowie (3) als Reaktion auf Störungen. Martin und Sass (2010) definieren das ‚Classroom Management‘ im Rahmen ihrer Studie als ein „multifaceted construct that includes two independent constructs: behavior manage-

---

gen kritisiert, die nicht gleichzusetzen ist mit der breiten Definition des ‚Classroom Managements‘. In der vorliegenden Arbeit wird zwar der Begriff Klassenführung verwendet, dieser beinhaltet jedoch weitaus mehr als nur den Umgang der Lehrperson mit Disziplinproblemen (vgl. Kapitel 3.2).

ment and instructional management“ (S. 1126). ‚Behavior Management‘ beinhaltet präventive Maßnahmen im Umgang mit Fehlverhalten aufseiten der Schüler. Zudem umfasst dieser Aspekt die Etablierung von Regeln und Belohnungsstrukturen sowie die Schaffung von Lernmöglichkeiten für die Schüler. ‚Instructional Management‘ fokussiert die instruktionalen Ziele und Methoden und beinhaltet u. a. Aspekte wie ‚Monitoring‘ oder die Strukturierung von täglichen Routinen.

Tabelle 4: Proaktive und reaktive Kriterien effektiven ‚Classroom Managements‘ nach Hennemann und Hillenbrand (2010, S. 259)

Proaktive Kriterien	Reaktive Kriterien
Vorbereitung des Klassenraums	Unangemessenes Schülerverhalten unterbinden
Planung und Unterrichtung von Regeln und unterrichtlichen Verhaltensweisen	Strategien für potenzielle Probleme
Festlegung von Konsequenzen	
Schaffen eines positiven (Lern-)Klimas im Klassenraum	
Beaufsichtigung der Schüler	
Unterricht angemessen vorbereiten	
Festlegung von Schülerverantwortlichkeit	
Unterrichtliche Klarheit	
Kooperative Lernformen	

Eine weitere Einteilung der Merkmale einer effektiven Klassenführung erfolgt durch Gold und Holodynski (2011). Die Autoren unterscheiden zwischen der Klassenführung als Organisation der individuellen Lerntätigkeit im Klassenverband, der Einführung und Einhaltung von Regeln sowie prozessorientierten Merkmalen einer erfolgreichen Klassenführung. Bei der Organisation der individuellen Lerntätigkeit im Klassenverband bzw. bei der Planung einer Lerneinheit sind von der Lehrperson drei grundlegende Anforderungen zu lösen (vgl. Bromme, 1997): Zeitmanagement, Schaffung von Lerngelegenheiten und Rückmeldungen sowie die Organisation der Schüler-Lehrer-Interaktion in Form von Unterrichtsaktivitäten (vgl. Gold & Holodynski, 2011). Unter den prozessorientierten Merkmalen einer erfolgreichen Klassenführung listen die Autoren die Merkmale von Kounin (1976; 2006) auf. Dieser analysierte im Rahmen einer Videoanalyse von 49 Grundschulklassen, als einer der ersten Forscher auf diesem Gebiet, Aspekte der Klassenführung in systematischer Art und Weise. Ausgangspunkt seiner Studie war die Suche nach erfolgreichen Strategien der Lehrpersonen, die darauf abzielen, einerseits Störungen zu unterbinden und andererseits die Schüler zu aktivieren. Fol-

gende Techniken der Klassenführung nennt Kounin (1976; 2006): Allgegenwärtigkeit und Überlappung, Steuerung von Unterrichtsabläufen (Reibungslosigkeit und Schwung), Aufrechterhaltung des Gruppenfokus (Gruppenmobilisierung, Rechenschaftsprinzip, Beschäftigungsradius) und programmierte Überdrussvermeidung. Die Kriterien von Kounin (1976; 2006) sind bis heute von herausragender Bedeutung für die empirische Bildungsforschung und gelten als wegweisend (vgl. Seidel, 2009). Unter anderem konnte er zeigen, dass sich effektive und weniger effektive Klassenmanager vor allem in präventiven Strategien zur Vermeidung von Unterrichtsstörungen unterscheiden und weniger in ihren Reaktionen auf Störungen (Kounin, 1970). In Tabelle 5 sind die Merkmale und Beispielindikatoren einer effektiven Klassenführung nach Kounin überblicksartig dargestellt. Welche Merkmale einer effektiven Klassenführung im Rahmen der vorliegenden Arbeit übernommen bzw. adaptiert wurden, wird in Kapitel 3.2 beschrieben.

Tabelle 5: Überblick über die Merkmale einer effektiven Klassenführung nach Kounin (2006)

Merkmale	Beispiel Indikator
Allgegenwärtigkeit	Die Lehrperson vermittelt den Eindruck einer Allgegenwärtigkeit, d. h. sie hat alle Aktivitäten im Blick und sozusagen auch auf dem Rücken Augen und Ohren.
Überlappung	Die Lehrperson kann ihre Aufmerksamkeit simultan auf mehrere Dinge gleichzeitig richten, z. B. einen Schüler ermahnen und den Unterricht weiterführen.
Reibungslosigkeit und Schwung	Die Lehrperson sorgt besonders an Übergangsstellen für einen flüssigen Unterrichtsverlauf.
Gruppenaktivierung	Die Lehrperson konzentriert sich auf die ganze Klasse und schafft es, die einzelnen Schüler bzw. Kleingruppen bei der Arbeit zu halten.
Abwechslung und Herausforderung	Die Lehrperson gestaltet die Arbeitsphasen so, dass sie von den Schülern als abwechslungsreich und herausfordernd wahrgenommen werden.

Neben Kounin (1976; 2006) wurde die Diskussion um Klassenführung / ‚Classroom Management‘ in Amerika maßgeblich durch Carolyn M. Evertson geprägt. Auch Evertson, Emmer und Worsham (2006) unterteilen ihre Merkmale des effektiven ‚Classroom Managements‘ in proaktive und reaktive Kriterien, wobei proaktive Kriterien eindeutig überwiegen (vgl. Henneemann & Hillenbrand, 2010). Die Autoren benennen elf Merkmale eines effektiven ‚Classroom Managements‘, die sie im Rahmen von Trainingsprogrammen den Lehrpersonen vermittelten (vgl. Evertson & Weinstein, 2006; Evertson et al., 2006):

- (1) Klassenraum vorbereiten
- (2) Regeln und Verfahrensweisen planen

- (3) Konsequenzen festlegen
- (4) Sofortige und konsistente Unterbindung
- (5) Regeln und Prozeduren unterrichten
- (6) Zu Beginn Aktivitäten zum Zusammengehörigkeitsgefühl entwickeln
- (7) Strategien für potenzielle Probleme rechtzeitig planen
- (8) Überwachen des Schülerverhaltens
- (9) Vorbereiten des Unterrichts
- (10) Verantwortlichkeit des Schülers
- (11) Unterrichtliche Klarheit

Die vorliegende Arbeit folgt in Anlehnung an die vorherrschende Sichtweise dem integrativen Ansatz. Demnach wird die Klassenführung nicht eindimensional beispielsweise nur in Bezug auf den Umgang mit Störungen erfasst, weil sie damit nur auf ein Disziplinmanagement reduziert wäre (vgl. Henneemann & Hillenbrand, 2010), sondern es werden sowohl proaktive als auch reaktive Kriterien berücksichtigt. Im nächsten Abschnitt werden diese inhaltlich beschrieben.

## 3.2 Merkmale der Klassenführung in der Grundschule

An dieser Stelle werden detailliert die Merkmale einer effektiven Klassenführung beschrieben, für die angenommen wird, dass diese auch in der Grundschule berücksichtigt werden sollten. In den Blickpunkt rückt zu Beginn vor allem die Nutzung der zur Verfügung stehenden Zeit als Rahmenbedingung des Unterrichtens und als eine von drei grundlegenden Anforderungen bei der Planung einer Lerneinheit (vgl. Bromme, 1997; Gold & Holodynski, 2011).

### 3.2.1 *Zeitmanagement / effiziente Zeitnutzung*

Das Merkmal der optimalen / effektiven Zeitnutzung gilt als wesentlicher Aspekt einer wirksamen Klassenführung (vgl. Helmke, 2007; Seidel, 2009; Wang, Haertel & Walberg, 1993b; Weinert, 1996) und als zentraler Prädiktor des Lernerfolgs (vgl. Anderson, Evertson & Brophy, 1979; Brophy, 2000; Einsiedler, 1997; Helmke, 1992; Lipowsky, Rakoczy, Pauli, Reusser & Klieme, 2007; Meyer, 2004; Walberg & Paik, 2000; Walberg, 1986; Wang et al., 1993b; Weinert, Schrader & Helmke, 1990a). In Unterrichtsstunden, die sich durch einen hohen Anteil effektiver Arbeitszeit auszeichnen, zeigen Schüler außerdem weniger unangemessenes Verhalten (vgl. Rutter, Höhn & von Hentig, 1980). Dabei wird vermutet, dass vor allem Disziplinprobleme das gegen-

standsbezogene Arbeiten und Lernen erschweren oder gar verhindern (vgl. Neuenschwander, 2005).

Je besser es der Lehrperson also gelingt, die zur Verfügung stehende Zeit für gegenstandsbezogenes Arbeiten und Lernen zu nutzen und sogenannte ‚Non-Instructional Activities‘<sup>39</sup> bzw. Zeitdiebe aus dem Unterricht auszulagern, desto mehr aktive Lernzeit steht zur Verfügung. Zu den Zeitdieben gehören beispielsweise ein unpünktlicher Beginn, zu lange Administrativa, schleppende Übergänge zwischen einzelnen Unterrichtsaktivitäten, Schwierigkeiten mit Medien und Technik oder unnötige Wartezeiten und Störverhalten von Schülern (vgl. Gold & Holodynski, 2011). In Anlehnung an den Begriff des ‚Classroom Managements‘ kann die Verteilung der zur Verfügung stehenden Unterrichtszeit auf verschiedene Unterrichtsformen, verschiedene fachliche Inhalte und auf unterschiedliche Schüler auch als ‚Time Management‘ bezeichnet werden (vgl. Arlin, 1979). Gerade in dieser Kompetenz sind die Unterschiede zwischen verschiedenen Lehrpersonen sehr groß (vgl. Weinert, 1996).

Auch für die Grundschule konnten Studien bestätigen, dass die Lernzeit, die für die Behandlung von relevanten Unterrichtsgegenständen verwendet wird, positive Auswirkungen auf den Lernerfolg der Schüler hat (vgl. Anderson et al., 1979; Cameron, Connor, Morrison & Jewkes, 2008; Campbell, Kyriakides, Muijs & Robinson, 2002; Helmke et al., 2008; Hofmann, 2007; Klicpera & Gasteiger-Klicpera, 1993; Pianta, Belsky, Vandergrift, Houts & Morrison, 2008; Roos & Schöler, 2009; Rowan, Correnti & Miller, 2002; Weinert & Helmke, 1997). Anderson et al. (1979) konnten für den Lesunterricht im ersten Schuljahr zeigen, dass eine längere Arbeitszeit der Schüler in den Lesegruppen die Leseleistungen positiv beeinflusst. Dies ist angesichts der Tatsache, dass einzelne Schüler bis zu zehn Minuten mehr Lesunterricht als andere Schüler hatten, nicht verwunderlich. Cameron et al. (2008) kommen auf Basis von Mehrebenenanalysen zu dem gleichen Ergebnis. Die Autoren bestätigen, dass sich Schüler, die zu Beginn des Schuljahres weniger Wörter lesen konnten, in Klassen mit einem deutlich höheren Anteil an Lesezeit im Unterricht besser entwickeln. Auch Rowan et al. (2002) konnten in ihrer Grundschulstudie zeigen, dass eine effektiv genutzte Lernzeit positiv mit dem späteren Lernerfolg der Schüler korrespondiert. Das Ergebnis konnte sowohl im Lesen als auch bedingt in Mathematik bestätigt werden. In der Studie von Klicpera und Gasteiger-Klicpera (1993) zeichneten sich am Ende des ersten Schuljahres Klassen mit einem hohen Leistungsstand im Rechtschreiben durch einen höheren Anteil an Zeit aus, in dem Rechtschreibung geübt wurde. Im zweiten Schuljahr unterschieden sich Klassen mit einem hohen und gerin-

---

39 Als „Non-Instructional Activities“ (vgl. Evertson et al., 2006; Rosenshine, 1980) werden unterrichtsfremde Vorgänge oder nicht-gegenstandsbezogene Aktivitäten bezeichnet (z. B. Milchgeld einsammeln, Ausflüge planen). Diese sollten weitestgehend aus dem Unterricht ausgelagert werden.

geren Leistungsstand hinsichtlich der Länge der Beschäftigungsdauer nicht mehr signifikant voneinander. Für das vierte Schuljahr bestätigen Befunde aus der EVES-Studie<sup>40</sup> (vgl. Hofmann, 2007; Roos & Schöler, 2009) positive Auswirkungen einer effektiven Zeitznutzung auf die Leistungsentwicklung im Rechtschreiben. Schüler, in deren Klassen viel Zeit durch die Beschäftigung mit außerschulischen Themen und durch zahlreiche Unterrichtsstörungen verloren ging, gehörten am Ende des vierten Schuljahres verhältnismäßig oft zu der schlechteren Leistungsgruppe.

Bereits Carroll (1963; vgl. Kapitel 1.1) ging davon aus, dass die Wahrscheinlichkeit des Lernerfolgs unmittelbar von der aufgewendeten Lernzeit abhängt. In der Systematisierung von Brophy (2000) nimmt das Merkmal ‚Opportunity to Learn‘, als Teil des ‚Classroom Managements‘, den zweiten Platz ein und beinhaltet sowohl den Aspekt des Zeitmanagements der Lehrperson, die Nutzung der zur Verfügung stehenden Lernzeit als auch den Umgang mit Unterrichtsstörungen. In der Liste von Wang et al. (1993b), die unterschiedliche Faktoren von Lernerfolg und deren Effektstärken beinhaltet, ist das Kriterium hoher Anteil echter Lernzeit in das Merkmal Klassenführung integriert. Auch Helmke (2009) integriert den Faktor Lernzeit in das Merkmal Klassenführung und verdeutlicht in seinem theoretischen Rahmenmodell (Helmke, 2007; vgl. Kapitel 1.2.2, Abbildung 5) die direkte Wirkung der Zeitznutzung auf die aktive Lernzeit<sup>41</sup> der Schüler. In dem Modell von Carroll (1963; vgl. Kapitel 1.1) wird die aktive Lernzeit als Schülersaufmerksamkeit konzipiert. Als Synonyma werden in der angloamerikanischen Forschung die Begriffe ‚Time on Task / Time off Task‘ verwendet (z. B. Emmer, Evertson & Anderson, 1980; Caldwell, Huitt & Graber, 1982). Verschiedene Studien bestätigen für die Grundschule durchschnittliche Aufmerksamkeitsraten von 75 %, d. h. Schüler sind im Durchschnitt drei Viertel der Unterrichtszeit aufmerksam (vgl. Helmke & Renkl, 1992; Peterson & Fennema, 1985).

Neben einer effektiven Zeitznutzung (3.2.1) sind weitere Aspekte der Klassenführung für ein hohes Schülersaufmerksamkeitsniveau bzw. ‚Time on Task‘ verantwortlich: (3.2.2) die Organisation der Übergangsphasen von einer Unterrichtsaktivität zur anderen (Übergangsmanagement), (3.2.3) die Einführung verbindlicher Verhaltensregeln für den Unterricht sowie für das Sozialverhalten und (3.2.4) die Schaffung einer störungsarmen Lernumgebung durch Prävention von Unterrichtsstörungen bzw. Fehlverhalten der Schüler. Dies kann beispielsweise durch die Fähigkeit der Lehrperson, allgegenwärtig zu sein, oder durch die Mobilisierung und das Einbinden der gesamten Klasse in das Unterrichtsgeschehen erreicht werden (vgl. Kounin, 1979; 2006). Auf

40 EVES steht als Abkürzung für Evaluation eines Vorschultrainings zur Prävention von Schriftspracherwerbsproblemen sowie Verlauf und Entwicklung des Schriftspracherwerbs in der Grundschule (vgl. Roos & Schöler, 2009).

41 Unter aktiver Lernzeit wird diejenige Zeit verstanden, „in der sich Schüler mit den zu lernenden Inhalten aktiv, engagiert und konstruktiv auseinandersetzen“ (Weinert, 1996, S. 124).

alle drei Aspekte wird anschließend detailliert eingegangen. Die Ausführungen beginnen mit (3.2.2) der Organisation der Übergangsphasen, die im Zusammenhang mit der Zeitznutzung folglich viel Zeit in Anspruch nehmen kann und eine Quelle für Ablenkungen aufseiten der Schüler darstellt. Daraufhin wird auf die Aspekte 3.2.3 und 3.2.4 Bezug genommen. Im letzten Punkt wird auf die Bedeutung der Vorbereitung des Klassenraums als Merkmal der Klassenführung eingegangen (3.2.5).

### 3.2.2 *Übergangsmangement (Managing Transition)*

Da sich der Unterricht in der Grundschule durch eine große Anzahl von verschiedenen Unterrichtsmethoden und Phasen auszeichnet (vgl. Götz et al., 2005) und die Organisation dieser Vielzahl von Phasen eine große Herausforderung für die Lehrperson darstellt, impliziert ein gutes Zeitmanagement besonders in der Grundschule ein gutes Übergangsmangement (vgl. Carter & Doyle, 2006). Häufige Methoden- und Sozialformwechsel sind vor allem in der Grundschule besonders wichtig, da diese mit dem Bedürfnis der Schüler nach motorischer Bewegung einhergehen (vgl. Zimmer, 1994; 1997). Junge Schüler haben im Gegensatz zu älteren noch kurze Konzentrationszeiten. Thal und Ebert (2004) geben als Richtzeiten für die Aufmerksamkeitsphasen von 5- bis 10-Jährigen 15 bis 20 Minuten an. Deswegen sind solche Sozialformwechsel, aber auch kurze Bewegungspausen, die in den meisten Fällen nichts mit dem Unterrichtsstoff zu tun haben, wichtig zur Auflockerung des Unterrichts. Häufige Methoden- und Sozialformwechsel führen ihrerseits auch zu methodischen Herausforderungen bei der Aufzeichnung von Unterricht durch Videografie (vgl. Kapitel 6.1).

Studien in der Grundschule bestätigen, dass sich effektive ‚Classroom Manager‘ unter anderem durch ein produktives Übergangsmangement auszeichnen (vgl. Cameron, Connor & Morrison, 2005; Connor et al., 2010; Wharton-McDonald, Pressley & Hampston, 1998). Es kann davon ausgegangen werden, dass ein geringer zeitlicher Anteil der Übergangsphasen in der Grundschule mit einem größeren Anteil an aktiver Lernzeit einhergeht. Bereits Kounin (1976; 2006) greift den Aspekt des Übergangsmagements in dem Merkmal Reibungslosigkeit und Schwung auf (vgl. auch Kounin, 1970, Transition Smoothness) und begreift ihn als eine wesentliche Komponente zur Steuerung von Unterrichtsabläufen. Pianta et al. (2008) schätzen mithilfe des CLASS-Beobachtungssystems (vgl. Kapitel 2.1) die Qualität der Übergänge ein, jedoch in Verbindung mit dem Merkmal Produktivität. Eine hohe Einschätzung des Merkmals Produktivität impliziert, dass die Übergänge routiniert verlaufen und wenig Zeit verschwendet wird.

Studien bestätigen, dass im Vergleich zur verwendeten Übergangszeit in der Sekundarstufe<sup>42</sup> die Phasen der Übergänge in der Grundschule mehr Zeit in Anspruch nehmen (vgl. Anderson et al., 1979; Burns, 1984; Carter & Doyle, 2006; Doyle, 1986; Hamre & Pianta, 2007; Helmke et al., 2008; NICHD ECCRN, 2002; Rosenshine, 1980), wobei sich deutliche Unterschiede zwischen einzelnen Grundschulklassen feststellen lassen. In der VERA-Studie liegt die real benötigte Zeit für Übergänge beispielsweise im Mittel zwar nur bei 5.4 %, schwankt jedoch in den Klassen zwischen minimal 0.8 % und maximal 15.9 % (vgl. Helmke et al., 2008). Für das erste Schuljahr (Leseunterricht) entfallen bei Anderson et al. (1979) zwischen 7 % und 14 % der Unterrichtszeit auf Übergänge, je nachdem, ob es sich um eine geschulte oder ungeschulte Lehrperson handelt.<sup>43</sup> Über 80 % dieser Übergänge wurden mit einem Signal eingeläutet (ebd.) und folgen demnach einem gewissen Ritual. Zu ähnlichen Ergebnissen für das erste Schuljahr gelangt auch die NICHD-ECCRN-Studie (2002). Die Autoren berichten, dass 17 % der zur Verfügung stehenden Zeit in ‚Non-Instructional Time‘ investiert wird. Diese Zeit beinhaltet jedoch neben der reinen Übergangszeit auch diejenige Zeit, die für soziale Probleme der Schüler benötigt wird. Auch Studien im zweiten Schuljahr können vergleichbare Anteile der Übergangszeit bestätigen (vgl. Burns, 1984; Rosenshine, 1980). Beispielsweise entfällt in der Studie von Rosenshine (1980) der Großteil der Zeit für ‚Non-Instructional Activities‘ auf die Übergänge.

Damit Übergänge zwischen den einzelnen Unterrichtsaktivitäten reibungslos verlaufen, sollten sie nach Evertson et al. (2006) strukturiert sein bzw. bestimmten Regeln und Ritualen folgen. In diesem Zusammenhang spielt das frühzeitige Einführen solcher Regeln eine wichtige Rolle für das Übergangsmanagement. So konnten Cameron et al. (2005) in ihrer Studie zeigen, dass Lehrpersonen, die zu Beginn des ersten Schuljahres mehr Zeit in die Organisation solcher Übergangsphasen investierten, im Verlauf der folgenden Schuljahre weniger Zeit durch Übergänge verschwendeten.

42 Im Rahmen der DESI-Studie gaben beispielsweise 14 % der befragten Deutschlehrer an, dass sie sich „pro Unterrichtsstunde mehr als 11 Minuten mit unterrichtsfremden Aspekten“ (Klieme et al., 2008, S. 323), beispielsweise mit Übergängen, aber auch mit fachfremden Inhalten, beschäftigen. Da der zeitliche Anteil der Übergänge in DESI zusammen mit anderen unterrichtsfremden Aspekten erhoben wurde, liegen die zeitlichen Anteile der Übergänge unter den hier angegebenen 11 Minuten. Für den im Rahmen der IPN-Video-studie aufgezeichneten Physikunterricht wurden lediglich zwei Minuten der Unterrichtszeit für Übergänge benötigt (Seidel et al., 2007).

43 In der Studie von Anderson et al. (1979) wurden zwei Gruppen von Lehrpersonen in einem experimentellen Design miteinander verglichen. Einer Gruppe wurde im Vorfeld der Unterrichtsbeobachtung eine Reihe von Richtlinien für eine effektive Klassenführung im Leseunterricht präsentiert (Treatment-Gruppe). Lehrpersonen aus der Treatment-Gruppe benötigten signifikant weniger Zeit für Übergänge als Lehrpersonen aus der Kontroll-Gruppe. Die hier angegebenen prozentualen Werte beinhalten die Ergebnisse beider Gruppen (Treatment- und Kontroll-Gruppe).

Reibungslose und strukturierte Übergänge sind zudem dadurch gekennzeichnet, dass die Lehrperson (1) die vorherige Arbeit der Schüler zum Abschluss bringt, (2) dafür sorgt, dass die Schüler ihr Material verstauen und (3) das soziale Verhalten der Schüler während der Phase beobachtet und gegebenenfalls interveniert (vgl. Buck, 1999; Carter & Doyle, 2006). Des Weiteren sollte die Lehrperson während einer Übergangsphase darauf achten, dass alle Schüler anschließend wissen, welche Aufgabe ihnen gestellt wurde. Sobald die Lehrperson die Aufgabenstellung wiederholen muss und somit einzelne Schüler warten müssen, geht wichtige Unterrichtszeit verloren (vgl. Fudge, Skinner, Williams, Cowden, Clark & Bliss, 2008).

Während der Übergangsphasen spielt der Einsatz von Lob eine wichtige Rolle (vgl. Brophy, 1981). Lob ist vor allem nach dem Beenden einer Aufgabe ein wichtiges Signal, um zu der nächsten Tätigkeit überzugehen und kann von der Lehrperson genutzt werden, um die Schüler für die neue Aktivität zu motivieren und den Übergang zwischen den Schüleraktivitäten zu beschleunigen.

In der Studie von Arlin (1979) sind ‚Smooth Transitions‘ (reibungslose Übergänge) durch eine geringe Anzahl von notwendigen Einwüfen durch die Lehrperson gekennzeichnet. Arlin (1979) unterscheidet in seiner Studie zwischen strukturierten und nicht strukturierten Übergängen. Strukturierte Übergänge zeichnen sich durch festgelegte Prozeduren und Regeln aus. Beispielsweise gilt ein Übergang als strukturiert, wenn die Lehrperson den Übergang frühzeitig mit den Worten ankündigt: „Ok, in fünf Minuten machen wir weiter mit Mathematik. Ihr solltet jetzt anfangen, eure Tische freizuräumen und eure Mathematikhefte bereitlegen“. Übergänge, die kurzfristig eingeleitet werden, ohne dass Schüler die Möglichkeit der Vorbereitung bekommen, werden hingegen als unstrukturiert bezeichnet. Der reibungslose Ablauf von Übergängen geht einher mit weniger Off-Task-Verhalten der Schüler. So konnte Arlin (1979) zusätzlich zeigen, dass im Vergleich zu anderen Unterrichtsphasen vor allem in den Phasen des Übergangs durchschnittlich mehr Unterrichtsstörungen stattfinden, was durch ein vermehrtes Off-Task-Verhalten der Schüler erklärbar ist.

### 3.2.3 Regelklarheit bzw. -verwendung

Für eine effektive Nutzung der Unterrichtszeit sowie die Strukturierung der Übergangsphasen sind eingeschliffene Handlungsrouninen, Rituale und rechtzeitig eingeführte Regeln unerlässlich. Regeln, Routinen und Prozeduren überlappen sich in ihrer Bedeutung teilweise mit Ritualen (vgl. Helmke, 2009). Aus diesem Grund werden die Begriffe kurz gegenübergestellt und inhaltlich voneinander abgegrenzt, bevor anschließend auf die Bedeutung eines effektiven Regelsystems in der Grundschule eingegangen wird.

Meyer (2004) versteht unter *Ritualen* regelmäßig wiederkehrende Tätigkeiten, die eine von der Lehrperson und den Schülern „geteilte symbolische

Bedeutung“ (S. 37) erhalten haben. Rituale bilden einen festen Bestandteil des Unterrichtsalltags. Sie begünstigen die Entwicklung von Sicherheit durch Routinen und können damit einen Beitrag zur Lernförderung leisten. „[...] Rituale haben den großen Vorteil, dass sie sehr präzise benannt und durchgesetzt werden können. Sie begünstigen sowohl die Selbständigkeit und Selbstverantwortung auf Schülerseite als auch und nicht zuletzt die Entlastung der Lehrkräfte im Unterricht“ (Klippert, 2008, S. 109).

Im Gegensatz zu Ritualen, bei denen es eher um Symbolhandlungen geht, versteht man unter *Regeln* Vorschriften (z. B. die Vorschrift, sich im Unterricht zu melden) und Verhaltenserwartungen, die zum Ausdruck bringen, welches Verhalten angemessen bzw. unangemessen ist. Wannack (2012) nennt fünf verschiedene Regelinhalte, die im Zusammenhang mit einem effektiven ‚Classroom Management‘ im Anfangsunterricht relevant sind: „Regeln zur Bewahrung der Integrität des einzelnen Schülers, zur Kommunikation in der Klasse und unter den Kindern, zur Sorgfalt im Umgang mit Material, zur Ordnung persönlicher Gegenstände sowie zur Mobilität im und außerhalb des Klassenraums“ (S. 70).

*Handlungsroutinen* bzw. *Routinen* beziehen sich hingegen auf spezifische Verhaltensweisen, die nach einem bestimmten Muster ablaufen (vgl. Havers, 1998). Solche wiederkehrenden Situationen werden synonym auch als Prozeduren bezeichnet (z. B. Wannack, 2012). Beispielsweise kann die Lehrperson die Übergänge zwischen den einzelnen Unterrichtsaktivitäten so gestalten, dass sie immer ähnlich verlaufen. Diese Handlungsroutinen können durch Signale (z. B. Läuten einer Glocke), Gesten oder Symbole (vgl. Helmke, 2009) unterstützt werden. Der Einsatz von Signalen steht zwar in keinem Zusammenhang mit dem Lernerfolg der Schüler z. B. im Lesen (1. Schuljahr, vgl. Anderson et al., 1979), kann jedoch von entscheidender Bedeutung sein, wenn es um die Schaffung einer ruhigen Arbeitsatmosphäre geht.

Die Etablierung und Einhaltung verhaltenswirksamer Regeln gilt als entscheidender Faktor erfolgreichen Lehrens und Lernens im Unterricht (Helmke, 2006). Nach Helmke (2006) dienen Regeln nicht nur der Prävention von unangemessenem Verhalten bzw. Störungen, sondern fördern gleichzeitig die Aufmerksamkeit, Anstrengung und Motivation der Schüler. Für das frühe Grundschulalter bestätigt Marzano (2003) durch eine Metaanalyse, dass Regeln und Prozeduren im Mittel einen hohen negativen Effekt auf die Anzahl von Unterbrechungen haben. Das bedeutet, dass in Klassen, die über ein effektives Regelwerk und Prozeduren verfügen, im Mittel deutlich weniger Unterrichtsstörungen ermittelt werden können, als in Klassen, in denen dieses effektive Regelwerk nicht existiert.

Im Sinne eines proaktiven Kriteriums der effektiven Klassenführung (vgl. Kapitel 3.1) ist es sinnvoll, die Regeln in der Klasse frühzeitig festzulegen (vgl. Carter & Doyle, 2006). Die Forschung bestätigt, dass die frühzeitige Einführung von Regeln den Grundstein für eine effektive Klassenführung darstellt. Emmer, Evertson und Anderson (1980) verglichen beispielsweise er-

folgreiche ‚Classroom Manager‘ mit weniger erfolgreichen und kamen für das dritte Schuljahr unter anderem zu dem Ergebnis, dass beide Gruppen von Lehrpersonen zwar Regeln und Prozeduren zu Beginn des Schuljahres im Unterricht etablierten, die erfolgreichereren ‚Classroom Manager‘ jedoch vor allem die ersten Tage in einer neuen Klasse intensiv nutzten, um die Regeln einzuführen und bei gegebenem Anlass ständig daran zu erinnern.

Den Ergebnissen der Studie folgend, sollten Lehrpersonen demnach vor allem nach der Übernahme einer Klasse Zeit aufwenden, um solche Regeln und Handlungsroutinen zu vereinbaren und zu festigen. Im Vergleich zur Sekundarstufe „geht es im Primarbereich darum, nicht nur Regeln und Routinen, sondern auch die prinzipiellen Inszenierungsmuster schulischen Unterrichts erstmals zu etablieren“ (Ophardt, 2008, S. 246). In den ersten Monaten der Schulzeit ist es Aufgabe der Lehrperson, vor allem die technischen Regeln und Verfahrensweisen wie beispielsweise das Ablegen der Hefte etc. einzuführen (vgl. Boostrom, 1991). Helmke (2003) weist darauf hin, dass Regeln nicht nur aufgestellt werden sollten, sondern in Abhängigkeit von der Altersgruppe immer wieder erläutert und auch erklärt werden müssen. Neu eingeführte Regeln sollten von der Lehrperson direkt vorgemacht (vgl. Nolting, 2008) und mit den Schülern eingeübt werden (vgl. Holodynski, 1998). Auch nach Gold und Holodynski (2011) sollten die Regeln für alle sichtbar aufgehängt und kontinuierlich eingeübt werden. Neben dem frühzeitigen Vereinbaren und Etablieren eines solchen Regelsystems führen die folgenden Aspekte zum Erfolg (Gold & Holodynski, 2011, S. 142):

- „Regeln sollten unmissverständlich formuliert sein.
- Sie sollten als beobachtbare Verhaltensweisen formuliert sein.
- Regeln sollten positiv formuliert sein und nicht die Form von Verbotslisten haben.
- Höchstens fünf Regeln gleichzeitig einführen, um den Überblick zu behalten. Weitere Regeln erst nach Etablierung der ersten einführen.
- Regeln im Konsens zwischen Schülern und Lehrperson auszuhandeln, zeigt die beste Wirkung.
- Ein von Schülern und Lehrpersonen unterzeichneter Vertrag erhöht die Verbindlichkeit. Wenig Nutzen bringen Regeln, die den Schülern diktiert und somit aufgezwängt werden.
- [...]“

Das Fehlen eines gemeinsamen Regelsystems für Lehrperson und Schüler kann Auswirkungen auf formale Unterrichtsabläufe oder auf das Verhalten der Schüler haben und somit den Unterricht stören (vgl. Boostrom, 1991). Aus diesem Grund wird ein geringer Grad an Regelhaftigkeit bzw. das Fehlen von regulativen Prinzipien und sinnhaften Ablaufstrukturen auch als „Chaos“ (Schwarzer, 1983, S. 132) bezeichnet.

Gleichermaßen bedeutsam ist es, dass Lehrpersonen neben der Etablierung eines Regelsystems die Einhaltung der Regeln bzw. Handlungsrou­tinen überwachen und Regelverstöße gerecht sanktionieren (vgl. Evertson et al., 2006; Holodynski, 1998; Gold & Holodynski, 2011; Wannack, 2012). Die Einhaltung von Regeln zielt vor allem auf das Aufrechterhalten angemessenen Verhaltens ab. Die konsequente Unterbindung unangemessenen Verhaltens sowie Sanktionen können von der Lehrperson unter anderem durch Verweise auf die abgesprochenen Regeln begründet werden. Wie die Lehrperson effektiv mit auftretenden Störungen im Unterricht umgehen kann, außer dass sie ein Regelsystem etabliert, wird nachfolgend erläutert.

### 3.2.4 *Störungsfreiheit / effektiver Umgang mit auftretenden Unterrichtsstörungen*

Als weiterer wichtiger Aspekt einer effektiven Klassenführung und als zentrale Aufgabe der Lehrperson gilt der wirksame Umgang mit auftretenden Disziplin- und Unterrichtsstörungen<sup>44</sup> und somit die Schaffung einer relativ störungsfreien Lernumgebung. Ein störungsarmer und disziplinierter Unterricht ist eng verbunden mit einem positiven Kompetenzerleben aufseiten der Schüler (DESI Englisch: Helmke et al., 2008; DESI Deutsch: Klieme et al., 2008; Rakoczy, 2007), auch in der Grundschule (vgl. Anderson, Evertson & Brophy, 1979; Klicpera & Gasteiger-Klicpera, 1993). So bestätigen Klicpera und Gasteiger-Klicpera (1993), dass am Ende des ersten Schuljahres in Klassen mit einem hohen Leistungsstand im Lesen und Schreiben die Schüler den Aufforderungen der Lehrperson eher nachkommen und es pro Unterrichtsstunde seltener zu Disziplinierungen durch die Lehrperson kommt als in Klassen mit einem geringeren Leistungsstand. Ein hoher prozentualer Anteil an ‚Corrections of Misbehavior‘ (Korrektur von Fehlverhalten), gemessen an der Gesamtzahl von Lehrer-Schüler-Interaktionen im ersten Schuljahr, hat in der Studie von Anderson et al. (1979) schlechtere Leseleistungen der Schüler zur Folge.

Unterrichtsstörungen können von der Lehrperson auf vielfältige Art und Weise vermieden werden. Um ein hohes Aufmerksamkeitsniveau der Schüler zu erreichen und somit Ablenkungen zu minimieren, werden vor allem

---

44 Neuenschwander (2005) fasst verschiedene Typologien und Klassifikationsschemata von Disziplinproblemen bzw. Unterrichtsstörungen zusammen. Nach Neuenschwander (2005) sind Unterrichtsstörungen „Blockaden und Widerstände, die ein effektives Funktionieren des Unterrichtsystems erschweren oder gar verhindern“ (S. 326). Das Urteil über das Ausmaß einer Unterrichtsstörung kann je nach befragter Person (Schüler, Lehrperson) unterschiedlich interpretiert werden, da sich nicht alle Personen von einem bestimmten Verhalten gleichermaßen gestört fühlen (vgl. Clausen, 2002). In Abgrenzung zu Unterrichtsstörungen definiert Neuenschwander (2006) Disziplinstörungen als „Verletzungen von sozialen Normen und Regeln im Unterricht“ (S. 190).

in amerikanischen Publikationen die folgenden Aspekte der Klassenführung als relevant eingestuft (vgl. Doyle, 1986; Evertson, 1989) und beispielsweise auf deutsche Grundschulen übertragen (vgl. Helmke & Renkl, 1993): Zusätzlich zu der frühen Einführung eines verbindlichen Regelsystems und zur guten Organisation von Übergängen zwischen den einzelnen Unterrichtsaktivitäten (s. o.) sollte das Schülerverhalten während des Unterrichts kontinuierlich überwacht werden. Dies wird als wesentliche Voraussetzung dafür angesehen, Störungen frühzeitig erkennen und unterbinden zu können. Kounin (1976; 2006) spricht in diesem Zusammenhang von der Fähigkeit der Lehrperson, allgegenwärtig zu sein (1). Außerdem zeichnet sich eine effektive Klassenführung durch ein hohes Maß an Gruppenmobilisierung aus. Die Aktivierung der gesamten Klasse minimiert die Ablenkung einzelner Schüler. Kounin (1976; 2006) bezeichnet diese Fähigkeit in seinen Arbeiten als Gruppenfokus (2). Im Detail stellen sich diese beiden Aspekte wie folgt dar:

### 3.2.4.1 Allgegenwärtigkeit der Lehrperson

Die Schaffung einer störungsarmen Lernumgebung geht einher mit einer kontinuierlichen „störungspräventiven Überwachung der Schülertätigkeiten“ (Gruehn, 2000, S. 129), welche eine Voraussetzung dafür ist, Störungen frühzeitig zu erkennen und gegebenenfalls sofort unterbinden zu können. Unter Rückgriff auf Kounin (1976; 2006) wird in diesem Zusammenhang der Begriff Allgegenwärtigkeit oder ‚Whititness‘ (vgl. Kounin, 1970; 1976; 2006) verwendet. Gleichbedeutend wird der Begriff ‚Monitoring‘ gebraucht (z. B. Evertson, Emmer, Sanford & Clements, 1983; Klieme et al., 2001; Kunter, 2005; Kunter, Baumert & Köller, 2007). Nach Kounin (1976; 2006) muss die Lehrperson allgegenwärtig sein bzw. buchstäblich „Augen im Hinterkopf“ (Kounin, 2006, S. 90) haben. Eine allgegenwärtige Lehrperson richtet ihren Fokus häufig auf die gesamte Klasse, auch wenn sie mit einzelnen Schülern oder Schülergruppen interagiert. Kounin (1976; 2006) bezeichnet diese Fähigkeit als Überlappung, d. h. die Lehrperson kann gleichzeitig mehreren Geschehnissen in der Klasse Aufmerksamkeit zuwenden.

Eine hohe Allgegenwärtigkeit der Lehrperson ist förderlich für die Mitarbeit und hemmend in Bezug auf das Fehlverhalten der Schüler (vgl. Kounin, 1976; 2006). Die Ergebnisse der Metaanalyse von Hattie (2009) belegen, dass die Allgegenwärtigkeit (Whititness) als zentrales Attribut der Lehrperson Auswirkungen auf die Leistungen der Schüler hat. In Klassen, die durch eine geringe Allgegenwärtigkeit der Lehrperson charakterisiert sind, wird tendenziell aufseiten der Schüler mehr Off-Task-Verhalten begünstigt bzw. die zur Verfügung stehende Unterrichtszeit durch häufigere Störungen reduziert und somit das On-Task-Verhalten minimiert (vgl. Emmer et al., 1980; Mayr, 2006; Rheinberg & Hoss, 1979).

Im Rahmen der TIMSS-Videostudie wird anstelle des Begriffs Allgegenwärtigkeit der Begriff ‚Monitoring‘ verwendet. ‚Monitoring‘ beinhaltet unter

anderem, inwieweit die Lehrperson darauf achtet, dass alle Schüler aufpassen (vgl. Klieme et al., 2001). Auch Kunter et al. (2007) benutzen den Begriff ‚Monitoring‘. Sie konnten zeigen, dass die durch die Schüler wahrgenommene Lehrerallgegenwärtigkeit sowie die Regelklarheit positiv mit dem Interesse der Schüler im Fach Mathematik zusammenhängen. Des Weiteren konnten positive korrelative Zusammenhänge mit dem Autonomie- und Kompetenzerleben der Schüler ermittelt werden. Monitoring der Lehrperson wurde in dieser Studie unter anderem durch folgende Items erfasst: „Unser Lehrer stellt sicher, dass wir ausreichend im Unterricht mitschreiben“, „Unser Lehrer weiß immer, was im Unterricht vorgeht“, „Unser Lehrer merkt sofort, wenn wir unkonzentriert sind“ oder „Unser Lehrer achtet darauf, dass wir aufmerksam sind“ (Kunter et al., 2007, S. 507, dt. Übersetzung d. Verf.).

Auch im ersten Schuljahr sind effektive ‚Classroom Manager‘ unter anderem durch ein hohes ‚Monitoring‘ und somit durch ein schnelles Eingreifen bei auftretenden Unterrichtsstörungen charakterisiert (vgl. Emmer et al., 1980). Vor allem Zurechtweisungen einzelner Schüler oder Schülergruppen (vgl. Kounin, 1976; 2006) bzw. Augenkontakt (vgl. Emmer et al., 1980), als non-verbale Zurechtweisung, gelten als wichtige Indikatoren dafür, dass die Lehrperson durch ihr Verhalten den Schülern mitteilt, dass sie über die Geschehnisse im Unterricht im Bilde ist und somit das Verhalten der Schüler kontrolliert. Aus diesem Grund wird die Allgegenwärtigkeit auch als Aspekt der Verhaltenskontrolle (vgl. Mayr, 2006) bzw. Kontrolle (vgl. Helmke & Schrader, 1997) angesehen.

### 3.2.4.2 Gruppenfokus / Gruppenaktivierung der Lehrperson

Neben einer ständigen störungspräventiven Überwachung der Schülertätigkeiten kann die Lehrperson Unterrichtsstörungen auch vorbeugen, indem sie versucht, nicht nur einzelne Schüler, sondern gleichzeitig möglichst alle Schüler in das Unterrichtsgeschehen einzubinden und zu aktivieren. Diese sogenannte Aufrechterhaltung des Gruppenfokus (vgl. Kounin, 1976; 2006) geht einher mit einer demokratischen Beteiligung aller Schüler im Unterricht.

Die gleichmäßige Beteiligung der Schüler am Unterricht hängt mit einer aktiveren Nutzung der Lernangebote und einer effektiven Klassenführung zusammen (vgl. Lipowsky, Pauli & Rakoczy, 2008; Kounin, 1976; 2006; Rheinberg & Hoss, 1979). Lipowsky et al. (2008) konnten beispielsweise mit Daten aus der Pythagorasstudie für den Klassenunterricht bestätigen, dass Klassen mit einer gleichmäßigen Verteilung der Schülerbeiträge auf alle Schüler von den externen Beobachtern sowohl in der Klassenführung als auch im Unterrichtsklima positiver wahrgenommen werden. Für die Schülerarbeitsphasen zeigten sich hingegen keinerlei Zusammenhänge. In der Studie von Rheinberg und Hoss (1979), die Kounins Kategorisierung anhand von 22 deutschen Sekundarschulklassen überprüften, ergab sich ein enger Zusammenhang zwischen der Kounin-Kategorie Gruppenaktivierung und dem be-

obachteten Schülerverhalten. Klassen, in denen Schüler angaben, dass sie von der Lehrperson zu jeder Zeit unvorhersehbar aufgerufen werden, hatten eine niedrigere Störungs- und eine höhere Mitarbeitsrate. Für Grundschulklassen zeigt sich ein ähnliches Bild. Kounin (1976; 2006) stellte für amerikanische Grundschulklassen fest, dass die Mobilisierung und Aktivierung der ganzen Klasse mit einer besseren Mitarbeit der Schüler und einer effektiveren Klassenführung korrespondiert. Erklärt werden kann dieser Befund damit, dass Unterrichtsgespräche in Klassen mit einer effektiven Klassenführung geordneter ablaufen als in Klassen mit einer weniger effektiven Klassenführung. In Klassen mit einer weniger effektiven Klassenführung wurden die Beiträge der Schüler nicht ausgewogen verteilt und nicht alle Schüler bekamen Gelegenheit zu einer inhaltlichen Auseinandersetzung mit dem Unterrichtsgegenstand (vgl. Lipowsky et al., 2008).

Die Aktivierung der gesamten Klasse (Klassenaktivierung bzw. Gruppenmobilisierung) kann von der Lehrperson erreicht werden, indem die Art der Ansprache durch die Lehrperson, z. B. nach einer Aufgabenstellung, so erfolgt, dass alle Schüler damit rechnen müssen, aufgerufen zu werden. Die Mobilisierung der Klasse / Gruppe wäre hingegen weniger kollektiv, wenn die Lehrperson vor einer Aufgabe den Namen des Schülers oder die Reihenfolge des Drankommens (z. B. nach Alphabet oder Sitzplan) vorgibt.

### 3.2.5 *Vorbereitung des Klassenraums*

Abgesehen von den bereits genannten Aspekten der Klassenführung kann angenommen werden, dass auch die räumliche Gestaltung des Klassenzimmers, gerade in den ersten Schuljahren, einen wesentlichen Faktor für den Lernerfolg darstellt (vgl. Knauf, 2001) und ebenfalls eine wichtige Bedeutung im Zusammenhang mit einer effektiven Klassenführung erfährt (vgl. Evertson et al., 2006; Toman, 2007; Wannack, 2012). Lehrpersonen können den Klassenraum so gestalten bzw. vorbereiten, dass dieser effektiv für ein erfolgreiches Lernen eingerichtet ist und somit Störungen, z. B. durch herumstehende Gegenstände oder durch schwer zugängliche Unterrichtsmaterialien vermieden werden. Einerseits muss der Klassenraum so vorbereitet sein, dass Staus und Störungen vermieden werden können, andererseits sollte der Raum aus Sicht der Lehrperson gut überschaubar sein, d. h. Allgegenwärtigkeit ermöglichen. Die Materialien für die Schüler sollten vorbereitet (vgl. Emmer et al., 1980) und schnell zugänglich sein (vgl. Evertson et al., 2006; Toman, 2007). Der Aspekt der Materialzugänglichkeit, wie beispielsweise, dass für jedes Kind eine eigene Mappe und / oder ein Fach vorhanden ist, spielt für die Reduzierung von Unterrichtsstörungen und für den Aufbau von Handlungsrouninen eine wesentliche Rolle (vgl. Meyer, 2004). Die Schüler wissen dadurch, wo sie ihre benötigten Arbeitsmaterialien finden bzw. an welchem Ort sie nicht benötigtes Material schnell unterbringen können. Die

Lehrperson sorgt somit dafür, dass Übergänge zwischen den Unterrichtsphasen einerseits ritualisiert werden und andererseits zwischen den einzelnen Sozialformen wenig Zeit verloren geht. Da besonders in der Grundschule häufig Sozialformwechsel stattfinden (vgl. Götz et al., 2005), zählt dieser Aspekt der Vorbereitung des Klassenraums zum Klassenmanagement (vgl. Evertson et al., 2006). Zusätzlich „versteht sich [die Klassenraumgestaltung] von Anfang an auch als Lebensraum zum Wohlfühlen“ (Toman, 2007, S. 167). Demnach ist die Klassenraumgestaltung nicht nur eine Grundlage für das ‚Classroom Management‘, sondern hat auch für das Unterrichtsklima eine zentrale Bedeutung (vgl. Evertson et al., 2006).

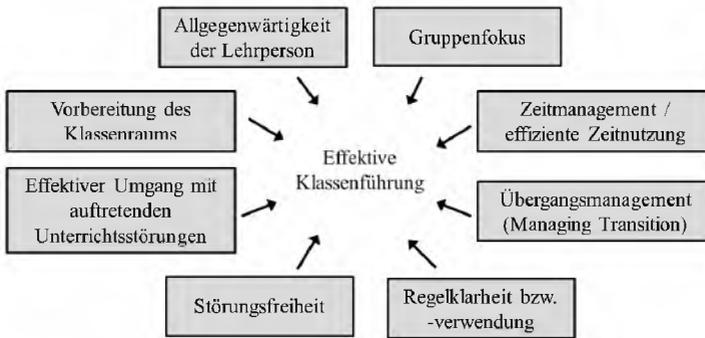


Abbildung 7: Untersuchte Merkmale der Klassenführung im Überblick

*Zusammenfassend* zeichnet sich auch in der Grundschule eine effektive Klassenführung dadurch aus, dass die Lehrperson die Strukturen des Unterrichts so gestaltet, dass der Alltag der Schüler geregelt und rhythmisiert abläuft, dass ausreichend Lernzeit zur Verfügung gestellt sowie Störungen vorgebeugt wird und der Klassenraum sowie alle benötigten Arbeitsmaterialien für ein erfolgreiches Lernen vorbereitet sind. Dafür sollte die Lehrperson allgegenwärtig sein und alle Schüler am Unterrichtsgeschehen beteiligen (Gruppenfokus). In Unterrichtsklima und Selbstkonzept sind abschließend alle Merkmale der Klassenführung überblicksartig dargestellt. Die Operationalisierung der insgesamt acht beschriebenen Merkmale der Klassenführung erfolgt in Kapitel 12.1.

## 4 Unterrichtsklima als Basisdimension der Unterrichtsqualität

Wie bereits im Kapitel 2 dargestellt, hängt die Qualität des Anfangsunterrichts nicht nur von einer effektiven Klassenführung ab, sondern unter ande-

rem auch von einem positiven Unterrichtsklima. Im Gegensatz zur Klassenführung ist das Unterrichtsklima jedoch nicht klar zu definieren. Anderson (1982) spricht nicht ohne Grund ein großes Problem der Unterrichtsklimaforschung an: „Are we all hunting the same beast?“ (S. 376) und weist damit auf die zahlreichen Definitionen des Begriffs Klima hin. Auch Eder (2001) setzt sich mit der Diskussion um die unterschiedlichen Verwendungsweisen des Begriffs ausführlich auseinander. Je nachdem, von welchem Standpunkt aus und unter welchem Forschungsinteresse man Klima betrachtet, werden andere Merkmale als bedeutsam erachtet (zusf. Eder, 1998; Eder, 2001; Gruehn, 2000). Dies liegt nicht zuletzt an der fehlenden Theoriebildung, die im Laufe der Jahre zu einer Fülle von Merkmalen und Operationalisierungen des Klimas geführt hat (zusf. Gruehn, 2000). Beispielsweise fassen Klieme et al. (2006) unter dem Begriff unterstützendes Lernklima eine positive Lehrer-Schüler-Beziehung, ein warmes unterstützendes Verhalten der Lehrperson oder den produktiven Umgang mit Schülerfehlern sowie positive und sachlich-konstruktive Rückmeldungen zusammen. Ein ‚Supportive Classroom Climate‘ (unterstützendes Klima) beinhaltet bei Brophy (2000) unter anderem, dass „the teacher displays concern and affection for students, is attentive to their needs and emotions, and socializes them to display these same characteristics in their interactions with one another“ (S. 8). Außerdem geht Brophy (2000) davon aus, dass in einem unterstützenden Klima Fehler als natürlicher Teil des Lernprozesses angesehen werden sollten.

In diesem Kapitel wird kurz auf die gegenwärtige Diskussion und derzeitige Definitionen von Klima eingegangen. Eine ausführliche Zusammenfassung zu Ansätzen und Befunden im Bereich Schul- und Klassenklimaforschung erfolgte bereits in der Arbeit von Gruehn (2000). Im Anschluss werden hier lediglich die theoretischen Aspekte sowie zentrale Forschungsbefunde zum Unterrichtsklima aufgezeigt, die für die vorliegende Arbeit relevant sind. In Anlehnung an das Kapitel 3.1 erfolgt zunächst die Definition des Klimabegriffs (vgl. Kapitel 4.1). In Kapitel 4.2 werden anschließend die für die Arbeit und die Entwicklung des hoch inferenten Ratingsystems relevanten Merkmale des Unterrichtsklimas referiert.

## 4.1 Der Klimabegriff – eine langjährige Debatte

Wie sich bereits andeutete, fällt es bis heute nicht leicht, eine eindeutige Definition des Klimabegriffs zu geben. In der pädagogischen Psychologie herrscht die dominierende Ansicht, dass Klima als eine subjektiv wahrgenommene Lernumwelt zu verstehen ist (vgl. Arbinger & v. Saldern, 1984; Dreesmann, 1980; Dreesmann, Eder, Fend, Pekrun, v. Saldern & Wolf, 1992; Eder, 1996; Eder, 2001; Lange, Kuffner & Schwarzer, 1983; v. Saldern, 1987). Arbinger

und von Saldern (1984) definieren Klima beispielsweise als „die (subjektive) Wahrnehmung, die (subjektive) Beurteilung, das (subjektive) Erleben schul- und unterrichtsbezogener Aspekte durch den Schüler“ (S. 88).

Klima als eine subjektiv wahrgenommene Lernumwelt kann auf unterschiedlichen Erfassungsebenen bzw. Bedeutungsebenen verortet werden. Hier wird zwischen Schul-, Klassen- oder Unterrichtsklima (zuf. Eder, 2001; Eder, 2002) bzw. zwischen Schul-, Klassen- und Schülerebene (zuf. Satow, 1999a) differenziert. Die Ebene des Unterrichtsklimas ist dabei die „konkreteste, da sie die Wahrnehmungen der Schüler einer Klasse hinsichtlich eines definierten Lehrers in einem bestimmten Fach betrifft“ (Clausen, 2002, S. 28). Auch für von Saldern (2000) ist das Unterrichtsklima „auf eine bestimmte didaktische Situation, genannt Unterricht“ (S. 160) bezogen. Nach Eder (2002) enthält der Begriff Unterrichtsklima sowohl eine klassenspezifische als auch eine lehrer- und fachspezifische Komponente. Die Definition von Dreesmann (1982) betont zusätzlich die Mehrdimensionalität und Stabilität des Unterrichtsklimas: „Unterrichtsklima ist eine relativ überdauernde Qualität der Umwelt eines Unterrichts, die sich auf einen bestimmten Satz von Merkmalen bezieht, der von den Schülern erlebt werden kann und ihr Verhalten potentiell beeinflusst“ (S. 43).

Unter dem Klassenklima wird hingegen „die sozial geteilte subjektive Repräsentation wichtiger Merkmale der Schulklasse als Lernumwelt“ (Eder, 2002, S. 215) verstanden, was wiederum erklärt, warum das Klassenklima häufig auch als Sozialklima beschrieben wird (z. B. Arbinger & v. Saldern, 1984; v. Saldern, 2000; v. Saldern & Littig, 1987). Bei dem Begriff Schulklima handelt es sich „um Wahrnehmungen der Schulumwelt durch die Schulmitglieder (genauer: um Wahrnehmungen von Umweltmerkmalen, die nicht episodisch-zufälliger, sondern typischer bzw. dauerhafter Art sind)“ (Dreesmann et al., 1992, S. 656).

Einen weiteren Vorschlag zur Strukturierung bzw. Typologisierung des Klimabegriffs macht Eder (1996; vgl. auch Eder, 1998; Eder, 2001). Seiner Ansicht nach sollten zum Verständnis des Klimabegriffs fünf Facetten berücksichtigt werden: der Inhalt, der Organisationsbezug, der Subjektbezug, die Beschreibungsbasis / Aggregierungsebene und die Datenquelle. Die Unterrichtsklima und Selbstkonzept gibt einen Überblick über die Vielzahl an daraus resultierenden Möglichkeiten, den Klimabegriff zu erfassen.

*Inhalt:* In dieser Definition des Klimabegriffs beziehen sich die Klimabeschreibungen auf unterschiedliche inhaltliche Bereiche der Organisation Schule, z. B. auf die sozialen Beziehungen der Schüler einer Klasse (Sozialklima) oder auf den speziellen Unterricht einer Lehrperson (Unterrichtsklima). *Organisationsbezug:* Bei dieser Definition wird auf institutionelle und organisatorische Einheiten, wie die Klasse (Klassenklima) oder Schule (Schulklima), Bezug genommen. *Subjektbezug:* Je nachdem, ob sich die Klimabeschreibungen eher auf das subjektive Erleben der Lernumwelt beziehen (psychologisches Klima) oder Aussagen über Merkmale der Organisation

als konsensuell abgesichert gelten (kollektives Klima), wird die Facette Subjektbezug angesprochen. *Aggregierungsebene*: Je nach Forschungsinteresse kann das Klima auf Individualebene (Individualklima) oder beispielsweise auf Ebene der Klasse als aggregiertes Klima (aggregiertes Klima) verwendet werden. *Datenquelle*: Zur Erfassung des Klimas können unterschiedliche Personengruppen befragt werden, was wiederum Auswirkungen auf die Begriffsverwendung hat (Schüler-, Lehrer-, Elternklima). Neben der Erfassung durch Schüler, Lehrpersonen oder Eltern können objektive und geschulte Beobachter das Klima in einer Klasse einschätzen. Die Datenquelle bildet somit die Einschätzung der Beobachter, die beispielsweise mittels Live-Beobachtung oder durch die Analyse von Unterrichtsvideos objektive Einschätzungen bezüglich der Merkmale des Unterrichtsklimas abgeben. Ergänzend zu Eder (1996; vgl. auch Eder, 1998; Eder, 2001) könnte somit die Datenquelle Beobachter in der Kategorisierung aufgenommen werden. Für die vorliegende Arbeit gilt diese Datenquelle als zentral, da das Unterrichtsklima weder aus der Perspektive der Schüler, Lehrpersonen oder Eltern erfasst wird, sondern mithilfe des Einsatzes von Unterrichtsvideos aus der Beobachterperspektive.

Tabelle 6: Übersicht über die Facetten des Klimabegriffs nach Eder (1998, S. 40)

Inhalt	Organisationsbezug	Subjektbezug	Aggregierungsebene	Datenquelle
Sozialklima	Klassenklima	Psychologisches Klima	Individualklima	Schülerklima
Unterrichtsklima	Schulklima	Kollektives Klima	Aggregiertes Klima	Lehrerklima
Werteklima	Lehrkörperklima			Elternklima

Im nachfolgenden Kapitel werden die Begriffe ‚Supportive Classroom Climate‘ bzw. unterstützendes Unterrichtsklima und Schülerorientierung gegenübergestellt, da sich in der bisherigen Forschung zum Unterrichtsklima zeigt, dass zwischen den beiden Begriffen große konzeptuelle Überschneidungen existieren.

#### 4.1.1 Unterstützendes Unterrichtsklima versus Schülerorientierung

Wenn man die Literatur zum Thema Unterrichtsqualität betrachtet, werden die Begriffe unterstützendes Unterrichtsklima (oder Supportive Classroom Climate) und Schülerorientierung häufig synonym verwendet (vgl. Clausen, Reusser & Klieme, 2003; Helmke, 2009; Klieme et al., 2001; Klieme et al., 2006; Kunter et al., 2005; Rakoczy, 2008), obwohl sich beispielsweise Helmke (2009) für eine klare Trennung beider Konstrukte ausspricht. Zwar hängt seiner Meinung nach das Qualitätsmerkmal lernförderliches Klima eng mit der

Schülerorientierung zusammen, jedoch unterscheiden sich beide Merkmale voneinander. Bei der Schülerorientierung geht es nach Helmke (2009) primär darum, dass die Lehrperson ihren Schülern nicht nur in fachlichen Fragen, sondern auch in nicht-fachlichen, persönlichen Belangen als Ansprechpartner zur Verfügung steht, unabhängig von Lernen und Leistung. Dies hat vor allem etwas „mit dem affektiven Aspekt der Lehrer-Schüler-Beziehung und dem Wohlbefinden von Schülerinnen und Schülern zu tun“ (Helmke, 2009, S. 230). Das lässt die Annahme zu, dass ein hoher Grad an Schülerorientierung einhergeht mit Wohlbefinden, größerem Selbstvertrauen und einer größeren Lernmotivation. Auch Klieme et al. (2001) subsumieren im Rahmen der TIMSS-Studie unter dem Begriff Schülerorientierung hauptsächlich Aspekte der Lehrer-Schüler-Interaktion (vgl. Kapitel 1.2). Im Rahmen der schweizerischen Videostudie im Fach Mathematik werden unter dem Begriff folgende Aspekte subsumiert: positives Sozialklima, Individualisierung, individuelle Bezugsnormorientierung, diagnostische Kompetenz der Lehrperson im sozialen Bereich, Fehlerkultur, Interaktionstempo und individuelle Lernunterstützung (vgl. Waldis et al., 2010b). Im Vergleich zum unterstützenden Unterrichtsklima thematisiert die Schülerorientierung somit hauptsächlich Aspekte der Beziehungsqualität (vgl. Helmke & Schrader, 1997; Tausch & Tausch, 1998). Bereits Fend (1977) nutzt zur Beschreibung der schulischen Umwelt neben den Merkmalen Inhalts- und Interaktionsaspekt den Beziehungsaspekt. Dieser beinhaltet bei Fend (1977) unter anderem die Bemühungen der Lehrperson um eine persönliche Beziehung zu den Schülern, die einerseits durch Vertrauen und andererseits durch Hilfsbereitschaft gekennzeichnet ist. Rakoczy (2008) übernimmt in ihrer Arbeit die Bezeichnung von Klieme, Schümer und Knoll (2001). Für sie zeichnet sich schülerorientierter Unterricht dadurch aus,

daß die Lehrkraft gute diagnostische Kompetenz im Sozialbereich zeigt und sich um die Probleme der Schülerinnen und Schüler kümmert. Solche Lehrkräfte üben wenig Druck aus und verwenden häufig einen individuellen Referenzrahmen zur Bewertung von Leistungen. Der Unterricht bietet den Schülerinnen und Schülern Freiräume, indem er individualisierende Maßnahmen einsetzt und den zu lernenden Stoff in multiplen authentischen Kontexten behandelt (Rakoczy, 2008, S. 61f).

Ein unterstützendes / lernförderliches Unterrichtsklima kann indes eher durch den konstruktiven Umgang mit Fehlern und ein geringes Maß an Wettbewerb und Konkurrenz charakterisiert werden. Außerdem ist ein unterstützendes Klima durch eine entspannte und lockere Lernatmosphäre gekennzeichnet (vgl. Helmke, 2009). Im Gegensatz zur Schülerorientierung beschreibt das unterstützende Unterrichtsklima somit eher Aspekte der Lernumwelt (vgl. Kapitel 4.1).

Eine klare Trennung beider Merkmale erfolgt beispielsweise in der Studie von Krammer (2009). Die Autorin unterscheidet Lehrer-Schüler-Interaktionen nach deren Bedeutung einerseits als Beziehungsgestaltung und ande-

rerseits als Lernunterstützung. (1) Lehrer-Schüler-Interaktionen als Beziehungsgestaltung beinhalten in Anlehnung an Tausch und Tausch (1998) die emotionalen Dimensionen Wertschätzung versus Geringschätzung sowie emotionale Wärme und Zuneigung versus emotionale Kälte / Abneigung. Die Forschergruppe um Tausch und Tausch (1998) konnte insbesondere der emotionalen Wärme positive Effekte auf das Wohlbefinden der Schüler zuschreiben (vgl. Kapitel 4.2). (2) Lehrer-Schüler-Interaktionen als Lernunterstützung implizieren das Anregen, Begleiten und Unterstützen der Schüler in Lernprozessen, womit zusätzlich Aspekte der kognitiven Aktivierung angesprochen werden.

Obwohl Helmke (2009) eine klare Trennung der Konstrukte Schülerorientierung und unterstützendes Unterrichtsklima für sinnvoll hält, existieren Ansätze, die einzelne Aspekte der Schülerorientierung als Teil des lernförderlichen Klimas ansehen (vgl. Bremerich-Vos, 2006; Meyer, 2004; Rakoczy, 2008). So enthält das lernförderliche Klima in der Klassifizierung von Meyer (2004) neben Aspekten der Schülerorientierung (Fürsorge, Verantwortungsübernahme sowie gegenseitiger Respekt) die Einhaltung von Regeln sowie die Gerechtigkeit der Lehrperson gegenüber allen Schülern. Klieme et al. (2006) sowie Roßbach (2002a; b) umgehen die eben beschriebene Problematik der synonymen Verwendung der Begrifflichkeiten, indem sie Begriffe wie unterstützendes schülerorientiertes Sozialklima oder schülerorientiertes sozial-emotionales Klima verwenden.

Im Rahmen der vorliegenden Arbeit wird dem zuletzt genannten Ansatz gefolgt. Da dieser Ansatz impliziert, dass die Aspekte der Schülerorientierung als Teil des lernförderlichen Klimas anzusehen sind, werden in der vorliegenden Arbeit sowohl Aspekte der Schülerorientierung (z. B. Fürsorge oder gegenseitige Wertschätzung) als auch des unterstützenden Unterrichtsklimas (z. B. Fehlerklima oder sachlich-konstruktive Rückmeldungen) als Komponenten des Unterrichtsklimas betrachtet (vgl. Kapitel 4.2).

#### *4.1.2 Verwendung des Klimabegriffs in der vorliegenden Arbeit*

Im Rahmen der Arbeit wird der konkrete Unterricht in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) mithilfe von Unterrichtsvideos analysiert und somit die Ebene des Unterrichtsklimas betrachtet. Nach Eder (1996; 1998; 2001) ist der Begriff Unterrichtsklima eher dadurch gekennzeichnet, dass er bereits inhaltlich beschreibt, auf welchen Aspekt er sich innerhalb der Organisation Schule bezieht. In Anlehnung an Bessoth (1989), der zwischen einem allgemeinen, vom Fachlehrer unabhängigen Klassenklima und dem fachlehrerbezogenen Unterrichtsklima unterscheidet, wird in der vorliegenden Arbeit unter dem Begriff Unterrichtsklima primär ein lehrerabhängiges „soziales Klima“ (Bessoth, 1989, S. 4) verstanden. Auch von Saldern (2000) definiert das Unterrichtsklima als eine spezifische Form des Sozi-

alklimas. Demzufolge gelten in einer Klasse interpersonale Beziehungen der Schüler zu der Lehrperson sowie die Schüler-Schüler-Beziehungen als zentrale Merkmale des Klimas (vgl. Zumhasch, 2006).

Kritisiert werden kann, dass eine derartige Festlegung auf den Begriff Unterrichtsklima aus rein formalen Gesichtspunkten erfolgt, da im Rahmen der Studie der konkrete Unterricht bei einer Lehrperson bzw. einem Lehrerteam beobachtet wird. Die Frage nach den Inhalten, die mit dem Begriff verbunden sind, ist damit noch nicht beantwortet. Welche Inhalte und Merkmale im Rahmen der vorliegenden Arbeit erfasst und untersucht werden sollen, wird ausführlich in Kapitel 4.2 dargestellt. In Anlehnung an die Definition des Unterrichtsklimas von Dreesmann (1982, vgl. Kapitel 4.1) wird dabei auf die Mehrdimensionalität des Konstrukts Bezug genommen. Die Mehrdimensionalität als ein Aspekt des Klimas bedeutet, dass das Klima theoretisch sowie empirisch als Überbegriff vieler Einzelmerkmale gebildet wird. Auf die unterschiedlichen Ansätze und Einteilungen geht Gruehn (2000) ausführlich ein. Resultierend aus der Gegenüberstellung zahlreicher Klimainstrumente konnte Gruehn (2000) in Anlehnung an Arbeiten von Eder (1996) feststellen, dass ein Großteil der eingesetzten Instrumente die einzelnen Aspekte des Klimas dahingehend differenzieren, ob sie sich auf Merkmale der Lehrer-Schüler-Beziehung, der Schüler-Schüler-Beziehung oder des Unterrichts beziehen. Mithilfe des Fragebogens zur Erfassung des Sozialklimas in Schulklassen (Landauer Skalen zum Sozialklima, LASSO, v. Saldern & Littig, 1985; v. Saldern, 1991) konnte diese Dreiteilung empirisch auch belegt werden (vgl. v. Saldern & Littig, 1985). In den LASSO zählt beispielsweise der wahrgenommene Leistungsdruck im Unterricht zu den Merkmalen des Unterrichts, während die Hilfsbereitschaft der Mitschüler als Merkmal der Schüler-Schüler-Beziehung und die Fürsorglichkeit der Lehrperson als Merkmal der Lehrer-Schüler-Beziehung betrachtet werden (vgl. v. Saldern, 1991). Im Anschluss werden nun die Merkmale des Unterrichtsklimas vorgestellt, wie sie für die Grundschule und im Rahmen der vorliegenden Arbeit operationalisiert wurden.

## **4.2 Merkmale des Unterrichtsklimas in der Grundschule**

Aufgrund der uneinheitlichen Definition des Klimabegriffs und der fehlenden Theorieanbindung variiert die Anzahl der erfassten Merkmale bzw. deren Operationalisierung zwischen den einzelnen Klimainstrumenten (zusf. Gruehn, 2000; vgl. Kapitel 4.1). Moos (1979) gliedert Unterrichtsklima beispielsweise in drei Bereiche: (1) soziale Beziehungen, (2) persönliche Entwicklung und (3) Bestand und Entwicklung der Institution. Arbinger und von Saldern (1984) gehen hingegen von fünf Bereichen aus, die das Klima konstituieren (Kontext der Schule / Klasse; physikalische und architektoni-

sche Merkmale der Schule; organisatorische Faktoren; Merkmale und Verhalten der Lehrperson sowie Merkmale der Schüler). In der eher pragmatisch-deskriptiven Einteilung der Landauer Skalen zum Sozialklima (LASSO, v. Saldern & Littig, 1985) werden die einzelnen Merkmale des Klimas dahingehend unterschieden, ob sie sich auf Merkmale der Schüler-Schüler-Beziehungen, der Lehrer-Schüler-Beziehungen oder auf Merkmale des Unterrichts beziehen. Auch für Eder (2011) gelten vor allem diese drei Bereiche als bedeutsam, „damit überhaupt eine sinnvolle Messung des Klimas möglich ist“ (S. 115). In Anlehnung an Gruehn (2000) und Eder (2011) erfolgt in Tabelle 7 die Einordnung der Merkmale des Unterrichtsklimas, die für die vorliegende Arbeit relevant sind, in diese Dreiteilung.

Tabelle 7: Einteilung der erfassten Einzelmerkmale des Unterrichtsklimas in drei Komponenten in Anlehnung an LASSO (vgl. v. Saldern & Littig, 1985)

Lehrer-Schüler-Beziehungen	Schüler-Schüler-Beziehungen	Merkmale d. Unterrichts
Anerkennung der Schüler durch die Lehrperson	Gegenseitige Anerkennung der Schüler	Humorvolle Lernatmosphäre
Fürsorglichkeit, Herzlichkeit und Wärme	Positives Fehlerklima zwischen den Schülern	Wettbewerbs- und Leistungsdruck (-)
Einsatz von Lob (Häufigkeit)		
Ermutigung durch die Lehrperson (Qualität von Lob)		
Sachlich-konstruktive Rückmeldung		
Positiver Umgang der Lehrperson mit Fehlern		

Da davon ausgegangen werden kann, dass die Gesamtheit der Unterrichtshandlungen der Lehrperson einen wesentlichen Beitrag zum Unterrichtsklima innerhalb einer Klasse leistet (vgl. Jerusalem & Schwarzer, 1991), wird in der vorliegenden Arbeit der Großteil der Merkmale des Klimas aus der Perspektive der Lehrer-Schüler-Beziehung erhoben (vgl. Tabelle 7, linke Spalte). Speziell im Grundschulalter spielt die Qualität der Lehrer-Schüler-Beziehung eine bedeutende Rolle (vgl. Ang, 2005; Brophy, 1983; Davis, 2001; Kammermeyer & Martschinke, 2009; Pianta, Steinberg & Rollins, 1995). Verschiedene Studien dokumentieren positive Zusammenhänge zwischen Aspekten der Lehrer-Schüler-Beziehung und dem Verhalten sowie den Leistungen der Schüler (z. B. Buyse, Verschueren, Verachtert & Van Damme, 2009; Cornelius-White, 2007; Hughes, Cavell & Jackson, 1999; Marzano, 2003; Murrey & Greenberg, 2000; Pianta et al., 1995; Roorda, Koomen, Split & Oort, 2011), wobei die Zusammenhänge mit dem Verhalten höher sind als mit den Leistungen (vgl. Roorda et al., 2011). Übereinstimmend mit Hamre und Pianta (2001) argumentieren Roorda et al. (2011), bezüglich der Lehrer-Schü-

ler-Beziehung: „They are partly a measure of social adjustment and therefore more proximal to behavioral than academic outcomes“ (S. 515). So zeigen Grundschüler (= Upper Elementary) mit positiven Beziehungen zu ihrem Lehrer in der Studie von Marzano (2003) weniger störendes Verhalten. Gleichzeitig empfinden Schüler, die von einer positiven Lehrer-Schüler-Beziehung berichten, die Lernumwelt deutlich positiver (vgl. Burnett, 2002). Andere Studien bestätigen zudem, dass im Laufe der Grundschule die Qualität der Lehrer-Schüler-Beziehung sowie die Nähe der Schüler zur Lehrperson sinken, während die Konflikte zunehmen (z. B. Jerome, Hamre & Pianta, 2009; O'Connor, 2010). Aus diesem Grund wird häufig argumentiert, dass junge Schüler im Vergleich zu älteren eine größere emotionale Bindung zu ihrer Lehrperson haben, während im Sekundarschulalter die Rolle der Peers wichtiger wird (z. B. Roorda et al., 2011). Vermutlich spielt in diesem Zusammenhang auch die Tatsache eine wesentliche Rolle, dass in der Grundschule noch das Klassenlehrerprinzip gilt und die Schüler somit einen Großteil des Schultages mit ein und derselben Lehrperson verbringen. Aus diesem Grund kann angenommen werden, dass im Anfangsunterricht die Lehrer-Schüler-Beziehung eine wichtige Rolle spielt, unabhängig vom jeweiligen Fach. Davis (2001) fasst die besondere Rolle der Lehrer-Schüler-Beziehung in der Grundschule wie folgt zusammen:

For the elementary school students, the teacher wears many hats such as friend, protector, mentor, disciplinarian, and gatekeeper to academic success. A positive relationship with teacher can act as an additional resource for a student while a difficult relationship may act as obstacle to academic success (S. 431f).

Ausgehend von der besonderen Rolle der Lehrer-Schüler-Beziehung in den ersten Schuljahren werden anschließend die Merkmale des Unterrichtsklimas dargestellt, die die Beziehungen zwischen der Lehrperson und den Schülern untereinander beschreiben. Dabei liegt der Fokus zu Beginn auf der grundlegenden Wertschätzung bzw. Anerkennung. Diese impliziert unter anderem einen respektvollen Umgang miteinander. Alle im Folgenden definierten Merkmale der Lehrer-Schüler-Beziehung, aber auch der Schüler-Schüler-Beziehung, stehen im engen Zusammenhang mit dem Gefühl der sozialen Eingebundenheit von Schülern, sowie dem Autonomie- und Kompetenzerleben (vgl. Deci & Ryan, 1993). Für ein gutes Unterrichtsklima ist vor allem das Gefühl der sozialen Eingebundenheit von großer Bedeutung. Befunde weisen darauf hin, dass die positive Qualität der Beziehung zur Lehrperson eng mit einem selbstbestimmten und engagierten Lernen aufseiten der Schüler zusammenhängt. Zudem bestätigen Studien, dass die Wahrscheinlichkeit intrinsisch motivierter Schüleraktivitäten steigt, je wertschätzender und interessierter die Lehrperson ist (z. B. Ryan & Deci, 2000a).

#### 4.2.1 Gegenseitige Wertschätzung / Anerkennung

Als ein zentraler Gesichtspunkt der sozialen Eingebundenheit werden in zahlreichen Arbeiten oder Überblicksartikeln die grundlegende Wertschätzung und der respektvolle Umgang miteinander angesehen (vgl. Eder, 2011; Kramis, 1990; Mayr, 2006; Meyer, 2004; Rakoczy, 2008). Pianta et al. (2008) betrachten die Wertschätzung als Aspekt der emotionalen Unterstützung (vgl. Kapitel 2.1). Solche wertschätzenden Beziehungen zwischen Lehrpersonen und Schülern (sogenanntes *Caring*<sup>45</sup>) gelten als eine wichtige Bedingung für motiviertes Lernen (vgl. Cornelius-White, 2007; Rakoczy, 2008; Wentzel, 1997) und werden als Kernkompetenz von Lehrpersonen in allen Schulstufen betrachtet (Dubs, 2002). Eine hohe Wertschätzung der Lehrperson als ein beziehungsförderndes Mittel führt zu weniger Problemverhalten aufseiten der Schüler (Mayr, 2006).

Das Verständnis von wertschätzenden Beziehungen weist enge Bezüge zu dem Begriff Schülerorientierung auf (vgl. Kapitel 4.1.1). Im Unterricht sind solche wertschätzenden Beziehungen dadurch gekennzeichnet, dass sich die Lehrperson bemüht, die Perspektive der Schüler einzunehmen. Wertschätzung ist außerdem geprägt durch Empathie und durch eine kollegiale Lernatmosphäre. Mayr (2006) betrachtet die Wertschätzung als ein Merkmal der Beziehungsförderung, das weniger Unruhe, Regelverletzungen oder Problemverhalten (Abschalten und Aggression) aufseiten der Schüler zur Folge hat.

Nicht nur wertschätzende Beziehungen zwischen der Lehrperson und den Schülern, sondern auch zwischen den Schülern untereinander, sind in Bezug auf ein positives Unterrichtsklima von Bedeutung (vgl. Eder, 2011; v. Saldern & Littig, 1987). Schüler, die sich gegenseitig respektieren, lassen einander ausreden und nehmen die Vorschläge von Mitschülern ernst. Außerdem sollte die Schüler-Schüler-Beziehung durch Akzeptanz und Höflichkeit gekennzeichnet sein. Für von Saldern und Littig (1987) stellt ein gewisses Maß an Hilfsbereitschaft und Zufriedenheit mit den Mitschülern ein Merkmal dieser Schüler-Schüler-Beziehungen dar. Satow (1999a) fasst diesen Aspekt unter dem Begriff *supportive Schüler-Schüler-Beziehungen* zusammen und versteht darunter primär die gegenseitige Hilfsbereitschaft, Rücksichtnahme und soziale Verantwortung. Nach seiner Meinung bilden positive soziale Bezie-

---

45 Unter ‚Caring‘ versteht Dubs (2002, S. 4) „eine Kombination von Übernahme von Verantwortung für jemanden und von Aufbau einer positiven emotionalen Beziehung zu dieser Person [...]. Caring kann aber auch heißen, [...] sich persönlich verantwortlich zu fühlen für das Wohlbefinden von anderen [...]“. In diesem Zusammenhang spielt vor allem die Empathie der Lehrperson eine wesentliche Rolle (vgl. Dubs, 2002). Wentzel (1997) verweist in diesem Zusammenhang auf Noddings (1992) und beschreibt unterstützende und fürsorgliche Lehrpersonen folgendermaßen: „(a) model caring behavior to their students, (b) engage students in dialogues that lead to mutual understanding and perspective taking, and (c) expect as well as encourage students to do the best they can given their abilities“ (S. 412).

hungen zu Gleichaltrigen den Kern unterstützender Schüler-Schüler-Beziehungen.

Das Gegenteil eines respektvollen Umgangs ist die Demütigung von Schülern (vgl. Meyer, 2004). Demütigende Verhaltensweisen im Unterricht sind erkennbar durch herabsetzende oder verletzende Äußerungen, Gesten des Verächtlichmachens oder durch das Bloßstellen einzelner Schüler durch die Lehrperson oder Mitschüler und haben im Umkehrschluss negative Auswirkungen auf die Beziehungen zwischen Lehrperson und Schülern bzw. zwischen Schülern untereinander.

#### *4.2.2 Positiver Umgang mit Fehlern im Unterricht / positive Fehlerkultur*

Auch der positive Umgang mit Fehlern kann als Aspekt der sozialen Eingebundenheit betrachtet werden (vgl. Seidel & Prenzel, 2003b) und hängt mit dem Wohlbefinden der Schüler in einer Klasse zusammen. Eine Atmosphäre, die durch einen konstruktiven Umgang mit Fehlern und Akzeptanz gekennzeichnet ist, verstärkt beispielsweise das wertschätzende und respektvolle Miteinander (vgl. Dalehefte, 2001). Das beinhaltet sowohl den positiven Umgang der Lehrperson mit Fehlern als auch die Fehlerkultur zwischen den Schülern. Dabei spielen sowohl die Einstellungen der Lehrperson zu Fehlern sowie der Umgang der Lehrperson mit Fehlern als auch das Fehlerklima zwischen den Schülern, d. h. deren Verhältnis zu Fehlern im Lernprozess, eine wichtige Rolle. Während Fehler bzw. fehlerhafte Schüleräußerungen als notwendiger Bestandteil von Lernaktivitäten angesehen werden (vgl. Meyer, Seidel & Prenzel, 2006) und ihnen besonderes Potenzial für den Wissensaufbau bei Schülern bestätigt werden kann (vgl. Oser & Spychiger, 2005), können Fehler in Leistungssituationen als negativ und frustrierend wahrgenommen werden.

Um Schülerfehler für den Wissensaufbau produktiv nutzen zu können, muss im Unterricht ein Lernklima existieren, in dem das Auftreten von Fehlern erlaubt ist (vgl. Hascher & Oser, 1996; Seidel & Prenzel, 2003b). Schüler sollten von der Lehrperson vermittelt bekommen, dass Fehler im Unterricht generell vorkommen können und dürfen (vgl. Oser, Hascher & Spychinger, 1999) und nicht vermieden werden. In diesem Zusammenhang spricht man von einer positiven Fehlerkultur (zusf. Schoy-Lutz, 2005; Hascher & Oser, 1996; Seidel & Prenzel, 2003b). Hascher und Oser (1996) sprechen dann von einer positiven Fehlerkultur, „wenn das Lernumfeld so gestaltet ist, daß der Lernprozess den konstruktiven Umgang mit Fehlern einschließt“ (S. 179). Eine positive Fehlerkultur geht einher mit einer guten Lehrer-Schüler-Kommunikation in der Klasse (vgl. Schoy-Lutz, 2005) und gilt somit als Aspekt des Sozialklimas (vgl. Hascher & Oser, 1996). Wenn der Lernprozess von der Lehrperson eher dahingehend gesteuert wird, dass Fehler unbedingt vermie-

den werden sollen, spricht man von einer negativen Fehlerkultur im Unterricht (vgl. Hascher & Oser, 1996). Durch die Vermeidung von Fehlern wird ein wichtiges Lernpotenzial von der Lehrperson nicht genutzt.

Zusätzlich kann das Publizieren von Fehlern durch Äußerungen der Lehrperson als ein problematisches Verhaltensmuster im Umgang mit Fehler-situationen angesehen werden. Die öffentliche Bekanntgabe negativen Wissens kann Auswirkungen auf die Emotionen der Schüler haben (vgl. Oser et al., 1999). Dies geschieht vor allem dann, wenn es der Lehrperson nicht gelingt, auf das Fehlermachen einzelner Schüler so objektiv wie möglich zu reagieren. Sobald Schüler die Reaktion der Lehrperson als persönlich bzw. gegen sich selbst gerichtet erleben, können negative Emotionen entstehen. Das Ereignis wird als negativ eingestuft und die Angst vor Fehlern steigt (vgl. Oser et al., 1999). Des Weiteren leistet eine solche negative Fehlerkultur „[...] auch Begleiterscheinungen wie dem Verlust an Motivation und Lernfreude Vorschub, unter welchen das Lernklima leidet und welche die Grundlagen des selbstständigen Lernens untergraben“ (Reusser, 1999, S. 203). Auch das Aufrufen anderer Schüler zur Fehlerkorrektur kann als Rüge bzw. als negative Verstärkung aufgefasst werden, je nach Attribution dieses Handlungsablaufs durch den Schüler (vgl. Oser et al., 1999).

Die Beobachtung von Fehlerkultur mithilfe von Unterrichtsvideos erfreut sich in den letzten Jahren zunehmender Beliebtheit. Durch Fehleranalysen konnten Bos, Bonsen, Baumert, Prenzel, Selter und Walther (2008) für die TIMSS-Stichprobe beispielsweise die nonverbale Reaktion von Lehrpersonen auf Schülerfehler in förderliche und hinderliche Verhaltensmuster differenzieren. Oser et al. (1998) kommen im Zuge einer quantitativen Auswertung von zehn videografierten Mathematikstunden der Sekundarstufe zu dem Ergebnis, dass ein Großteil der Fehlersituationen (43 %) einem positiven Klima zugeordnet werden kann. Dieses Ergebnis bestätigte sich auch nach der Befragung der Schüler, die insgesamt die Fehlerkultur der Schule als positiv wahrgenommen haben.

Es existieren jedoch auch Befunde, die für den deutschen Sekundarschulunterricht bestätigen, dass eine öffentliche Thematisierung von Fehlern bzw. eine negative Reaktion auf Fehler eher selten im Unterricht und somit anhand von Unterrichtsvideos zu beobachten ist (vgl. Meyer et al., 2006; Heinze, 2004; Oser et al., 1998). In den von Heinze (2004) untersuchten Klassen traten beispielsweise im Mathematikunterricht im Mittel weniger als fünf Fehler pro Stunde auf, was eine systematische Analyse der Reaktion auf Fehlersituationen nur eingeschränkt ermöglichte. Meyer et al. (2006) bestätigen im Rahmen der IPN-Physikstudie, dass in einem Großteil der untersuchten Klassen Aspekte der Fehlervermeidung und die negative Reaktion auf Fehler nicht beobachtet werden konnten. Klassen, in denen besondere Fähigkeitsdemonstrationen bzw. verächtliche Reaktionen auf Fehler vorkamen, wiesen insgesamt niedrige Ausprägungen auf. Die Autoren schlussfolgern daraus, dass es gerade ein Kennzeichen des deutschen Unterrichts ist, dass fast ausschließlich

positive Beispiele oder Elemente des Lernstoffs behandelt werden und somit Fehler selten im Unterricht zu beobachten sind. Die Autoren sehen als ein zusätzliches Hindernis für die videobasierte Beobachtung das sogenannte Bermuda-Dreieck an. Dieses meint, dass falsche oder unerwünschte Schülerbeiträge von der Lehrperson übergangen werden. Aufgrund dieser Nicht-Reaktion kann während des Kodiervorgangs häufig gar nicht entschieden werden, ob es sich um einen Fehler handelt. Solche Situationen sind wiederum eng mit der Qualität der Fehlerkultur verknüpft.

Für die Grundschule konnte im Rahmen der WOLFF-Studie bestätigt werden, dass lediglich in fünf Prozent der Einzelgespräche zwischen Lehrperson und Schüler die Lehrpersonen im Sinne einer kognitiv aktivierenden Fehlerkultur interagieren, unabhängig davon, ob es sich um schüler- oder lehrerzentrierte Unterrichtsphasen handelt (vgl. Kucharz, Schnebel & Helming, 2011). Nichtsdestotrotz kann auch für die Grundschule angenommen werden, dass die Beurteilung der Leistungen der Schüler über die Anzahl der Fehler erfolgt und somit Auswirkungen auf das Unterrichtsklima haben kann, je nachdem wie mit Fehlern umgegangen wird. In der Untersuchung von Spychiger, Mahler, Hascher und Oser (1998; zit. n. Chott, 1999) aus der Schweiz brachten beispielsweise sowohl jüngere Schüler (4. Schuljahr) als auch ältere Schüler (9. Schuljahr) zum Ausdruck, dass der Umgang mit Fehlern in der Schule allgemein als verletzend wahrgenommen wird und mit negativen Erfahrungen verbunden ist.

Eine positive Fehlerkultur kann eine wichtige Rolle für die Entwicklung des Selbstkonzepts der Schüler spielen. Aufseiten der Lehrperson haben das Ignorieren von Fehlern oder das Aufrufen von Mitschülern zur Fehlerkorrektur zur Folge, dass Schüler den Eindruck bekommen, Fehler seien emotional negativ besetzt. Daraus „folgen Angst und Schamgefühl auf Seite der Schüler bezüglich des Fehlermachens“ (Heinze, 2004, S. 228), was vermutlich Auswirkungen auf das Selbstkonzept haben kann. Neben Effekten auf nicht kognitive Merkmale hängt eine positive Fehlerkultur zudem auch mit kognitiven Zielkriterien zusammen. In der Studie von Klieme et al. (2008) bestätigte sich beispielsweise ein Zusammenhang mit dem Hörverstehen im Fach Englisch (9. Schuljahr) und dem Lerninteresse, wobei der Effekt der positiven Fehlerkultur auf das Hörverstehen schwächer ausfiel als auf das Lerninteresse.

#### *4.2.3 Lehrerfürsorglichkeit als Merkmal einer positiven Lehrer-Schüler-Beziehung*

Besonders im Anfangsunterricht ist ein herzlicher und einfühlsamer Umgang der Lehrperson mit den Schülern wichtig, da Zuwendung, körperliche Nähe und Berührung von jungen Kindern als Anreiz und Belohnung erfahren werden und zu den sozialen Belohnungen zählen (vgl. Reimann, 2004). In der Literatur wird in diesem Zusammenhang häufig von der Lehrerfürsorglich-

keit<sup>46</sup> bzw. Fürsorge als einem wesentlichen Merkmal der Lehrer-Schüler-Beziehung und somit des Unterrichtsklimas gesprochen (z. B. Bülter & Meyer, 2004; Meyer, 2004; v. Saldern & Littig, 1985; 1987; Satow, 1999a). Der Begriff bezieht sich vor allem auf „die individuelle, persönliche Zuwendung der Lehrperson zu den Schülern, ihren Bedürfnissen und auch außerschulischen Problemen“ (Satow, 1999a, S. 78). Von Saldern und Littig (1987) verstehen unter der Fürsorglichkeit die Unterstützung und Kooperationsbereitschaft der Lehrperson und beziehen sich dabei weniger auf das Geschehen im Unterricht als vielmehr auf die persönlichen Interessen und Probleme der Schüler. Eine fürsorgliche Lehrperson nimmt die Anliegen ihrer Schüler ernst und ermöglicht ihnen im Unterricht einen gewissen Raum an Gestaltungsfreiheiten. Dies fördert nicht nur die aktive Beteiligung bzw. Lernbereitschaft der Schüler am Unterricht, sondern auch die Eigenverantwortlichkeit, da sich die Schüler als Verursacher ihrer Handlungen sehen. Die Lehrerfürsorglichkeit trägt über diesen Prozess zusätzlich zum Aufbau von Selbstwirksamkeitserwartungen bei (vgl. Satow, 1999a).

Pianta et al. (2008) haben in diesem Zusammenhang die sogenannte ‚Teacher Sensitivity‘ (Sensibilität der Lehrperson) als Aspekt der emotionalen Unterstützung (Emotional Support) erfasst, wobei sich für das dritte Schuljahr hohe Korrelationen mit einem positiven Klima, der Verhaltenskontrolle, der Berücksichtigung der Schülerperspektiven sowie mit der Qualität des Feedbacks bestätigen ließen. Auch das NICHD (NICHD ECCRN, 2002) ließ mithilfe eines Beobachtungssystems für die erste Klasse die ‚Sensitivity/Responsivity‘ der Lehrperson einschätzen, wobei die Werte von ungefähr 80 % der untersuchten ersten Klassen für dieses Merkmal oberhalb des Mittelwerts von vier (siebenstufige Skala) lagen (vgl. auch Hamre & Pianta, 2007). Rimm-Kaufman et al. (2002) zeigten zusätzlich, dass Schüler in Klassen mit einer sensiblen Lehrperson engagierter und selbstbewusster sind als Schüler in Klassen mit einer weniger sensiblen Lehrperson.

Im Rahmen der vorliegenden Arbeit wird die Lehrerfürsorglichkeit um die Aspekte Herzlichkeit und Wärme der Lehrperson ergänzt. Diese Erweiterung erfolgt in Anlehnung an bisherige Ergebnisse, die bestätigen, dass die Wärme der Lehrperson einen wesentlichen Zusammenhang mit Schülermerkmalen aufweist (zusf. Cornelius-White, 2007; Skinner & Belmont, 1993). Generell scheint sich die emotionale Wärme einer Lehrperson günstig auf die interpersonellen Beziehungen zwischen der Lehrperson und den Schülern sowie auf das Klassenklima allgemein auszuwirken. Tausch und Tausch (1998) konnten im Rahmen einer Schülerbefragung zeigen, dass diejenigen Schüler, die von Lehrern mit einem hohen Maß an emotionaler Wärme unterrichtet

---

46 Die Lehrerfürsorglichkeit kann auch als Aspekt der Schülerzentriertheit bezeichnet werden (vgl. Eder, 1996). Eder (1996) fasst beispielsweise unter der Schülerzentriertheit alle Verhaltensweisen der Lehrperson zusammen, in denen es um die Förderung sowie Einbeziehung der Schüler geht. In der vorliegenden Arbeit wird jedoch der Begriff Lehrerfürsorglichkeit verwendet.

tet wurden, weniger ängstlich sowie zufriedener sind. Beziehungen zu Leistungsbereitschaft und -verhalten der Schüler konnten allerdings nicht bestätigt werden. In einer älteren Studie von Christensen (1960) hing jedoch die Wärme / Zuwendung bei Lehrpersonen im vierten und fünften Schuljahr bedeutsam mit dem Leistungsgewinn der Schüler in Sprachen (Vocabulary) und Mathematik (Arithmetic Skills) zusammen. Im Rahmen der Studie von Pianta und Steinberg (1992) können Zusammenhänge der durch die Lehrperson eingeschätzten Wärme mit verschiedenen Aspekten des Schülerverhaltens belegt werden. Lehrpersonen, die die Beziehung zu ihren Schülern als wertschätzend, warm und herzlich einschätzen und ihren Schülern auch tröstend zur Seite stehen, nehmen auch weniger Verhaltensprobleme ihrer Schüler wahr bzw. berichten von positivem Arbeitsverhalten und sozialen Fähigkeiten. In der Studie von Skinner und Belmont (1993) empfanden sich junge Schüler (3.–5. Schuljahr) als glücklicher und enthusiastischer, wenn sie die Lehrperson als warm und herzlich wahrnahmen.

Die Ergebnisse lassen darauf schließen, dass in der Grundschule die Fürsorglichkeit und Sensibilität der Lehrperson als Aspekte eines positiven Klimas eine große Rolle spielen und dementsprechend hoch ausgeprägt sein sollten. Vor allem die implizierte emotionale Wärme der Lehrperson hat einen positiven Effekt auf das Wohlbefinden der Schüler. Die individuelle und persönliche Zuwendung der Lehrperson und die empfundene Wärme sind außerdem von Bedeutung für die Selbstkonzeptentwicklung der Schüler. Durch die Schaffung eines Vertrauensverhältnisses kann die Lehrperson den Schülern Sicherheit bei der Bewältigung schulischer Herausforderungen geben und somit die Selbstwirksamkeitserwartung günstig beeinflussen. Satow (1999b) konnte beispielsweise für die Sekundarstufe einen Zusammenhang zwischen Selbstwirksamkeitserwartungen der Schüler und der Lehrerfürsorglichkeit zeigen, sowohl auf Schüler- als auch auf Klassenebene<sup>47</sup>.

Wie bereits erwähnt verwendet Eder (1996) den Begriff Schülerzentriertheit anstatt Lehrerfürsorglichkeit. Unabhängig von der Begrifflichkeit konnte jedoch auch Eder (1996) positive Zusammenhänge mit dem Selbstkonzept der Schüler bestätigen. Die wahrgenommene Schülerzentriertheit der Lehrperson beeinflusst zusätzlich über die Zufriedenheit die Leistungen der Schüler (ebd.).

---

47 In der Studie von Satow (1999b) wird der Effekt der Lehrerfürsorglichkeit auf die Selbstwirksamkeitserwartung untersucht. Die Selbstwirksamkeitserwartung kann jedoch nicht mit dem Selbstkonzept gleichgesetzt werden. Das Selbstkonzept ist ein globales Konstrukt, das viele verschiedene Selbstwahrnehmungen umfasst, darunter auch die Selbstwirksamkeit. Diese ist im Vergleich zum Selbstkonzept eher zukunftsorientiert (vgl. Woolfolk, 2008). Es kann jedoch vermutet werden, dass die Lehrerfürsorglichkeit auch einen Effekt auf das Selbstkonzept hat.

#### 4.2.4 *Positiv unterstützende (Lob / Ermutigung) und sachlich-konstruktive Rückmeldungen*

Damit Schüler sich in einer angenehmen Atmosphäre als kompetent erleben, ist die Art der Rückmeldung z. B. in einer Fehlersituation ein entscheidender Faktor. Bisherige Videostudien aus der Sekundarstufe unterscheiden einfache Rückmeldungen und Rückmeldungen, die die soziale Bezugsnorm betreffen sowie positiv unterstützende und sachlich-konstruktive Rückmeldungen (vgl. Kobarg & Seidel, 2003; Rakoczy & Pauli, 2006). Durch den Einsatz positiv unterstützender Rückmeldungen erfahren die Schüler von der Lehrperson positive Unterstützung. In der Grundschule gelten der Einsatz von Lob sowie das Vermitteln von Zutrauen im Sinne einer Ermutigung als ein wichtiges Element der Rückmeldung. Eine solche Rückmeldung kann sowohl im öffentlichen Unterrichtsgespräch als auch in Schülerarbeitsphasen erfolgen, wobei Grundschüler privates Lob bevorzugen (zusf. Burnett, 2002). Lipowsky et al. (2008) zählen Lob und Ermutigung zu den affektiven Aspekten der unterrichtlichen Kommunikation und begreifen sie als Indikatoren eines unterstützenden Klimas in der Klasse. Diesem Ansatz wird auch in der vorliegenden Arbeit gefolgt.

Für Brophy (1981) enthält das Lob einer Lehrperson eine positive Konnotation, da lobende Äußerungen der Lehrperson „express positive teacher affects (surprise, delight, excitement) and/or place the student’s behavior in context by giving information about its value or its implications about the student’s status“ (S. 6). Demnach geht Lob „über die schiere Rückmeldung der Richtigkeit hinaus und enthält eine positive gefühlsmäßige Komponente der Freude“ (Hofer, 1985, S. 416). Brophy (1981) nennt neben der spontanen Äußerung von Erstaunen oder Bewunderung weitere sieben Funktionen von Lob: (1) Lob als ein Ausgleich für eine vorangegangene Kritik, (2) Lob als stellvertretende Bekräftigung, (3) Lob als positive Anleitung oder Vermeidung von Kritik, (4) Lob als Eisbrecher oder Friedensangebot, (5) Lob als „Student-elicited Stroking“, (6) Lob als Übergangsritual (vgl. Kapitel 3.2) und (7) Lob als Trost oder Aufmunterung.

Die besondere Relevanz des Lobs für die Lernentwicklung von jungen Schülern wird durch eine Untersuchung im Bereich der Hirnforschung untermauert. Duijvenvoorde, Zanolie, Rombouts, Raijmakers und Crone (2008) fanden heraus, dass bei Schülern im Alter zwischen acht und neun Jahren vor allem Lob im Sinne positiven Feedbacks wirkt und sie auf negatives Feedback hingegen kaum emotional reagieren. Bereits Brophy (1981) bescheinigt dem Lob ein großes Potenzial für die ersten Schuljahre. Bis zu einem Alter von sieben bis acht Jahren orientieren sich die Schüler sehr daran, der Lehrperson als einer Autoritätsperson zu gefallen. „Erfolg und darauf folgendes Lob werden als Zeichen sozialer Zuwendung interpretiert, Mißerfolge als Hinweis für soziale Ablehnung“ (Fend, 1997, S. 132). Für junge Schüler stellt Lob eine

Form der Orientierung dar und impliziert, dass ein Verhalten oder eine Leistung des Schülers der Lehrperson gefällt. Schüler, die versuchen, der Lehrperson zu gefallen, strengen sich mehr an und zeigen wiederum weniger störendes Verhalten (vgl. Brophy, 1981). Lob hat somit im Sinne einer stellvertretenden Bekräftigung<sup>48</sup> die Funktion, positives Schülerverhalten zu stärken. Im Vergleich zur Sekundarstufe, in der die Schüler vorwiegend bezüglich ihrer erbrachten Leistungen gelobt werden wollen, geht es in der Grundschule demnach auch darum, positives Verhalten der Schüler zu würdigen (vgl. Burnett, 2001).

Im Sinne der stellvertretenden Bekräftigung sollte der Einsatz von Lob nach O'Leary und O'Leary (1977; vgl. auch Brophy, 1981) möglichst spontan nach einem Schülerverhalten erfolgen, sowie ernst und aufrichtig gemeint sein. Außerdem sollte der einzelne Schüler bzw. die gesamte Klasse das Lob auf das eigene Verhalten oder die eigenen Leistung zurückführen können. In diesem Zusammenhang spricht man von einem spezifischen Lob. O'Leary und O'Leary (1977) beschreiben die Qualität des Lobes mithilfe der Begriffe ‚Contingency‘ (Unvorhersehbarkeit), ‚Specificity‘ (Spezifität) und ‚Sincerity / Variety / Credibility‘ (Ernsthaftigkeit / Abwechslung / Glaubhaftigkeit).

Lob kann nach Brophy (1981) von der Lehrperson zusätzlich eingesetzt werden, um eine freundliche Atmosphäre zu schaffen. Brophys Behauptung wird jedoch durch eine Studie aus Australien teilweise revidiert (vgl. Burnett, 2002). In einer Längsschnittstudie mit Grundschulern (3.–6. Schuljahr<sup>49</sup>) konnte Burnett mittels Pfadanalysen keinen Effekt der Frequenz generellen Lobens auf die durch die Schüler eingeschätzte Lernumwelt und die wahrgenommene Lehrer-Schüler-Beziehung bestätigen, wobei hingegen Lehrerfeedback bezüglich der Anstrengungsbereitschaft (Effort Feedback) und der Leistungen (Ability Feedback) positiv mit der wahrgenommenen Lernumwelt (Classroom Environment) der Schüler zusammenhängt. Die Studie macht deutlich, dass nicht generelles Lob hilfreich für den Aufbau positiver Lehrer-Schüler-Beziehungen ist, sondern die Anzahl spezifischen Feedbacks. Je mehr die Lehrperson die Anstrengungsbereitschaft der Schüler würdigt, desto häufiger berichten die Schüler von positiven Lehrer-Schüler-Beziehungen, die wiederum einen positiven Effekt auf die wahrgenommene Lernumwelt der Schüler haben. Die wahrgenommene Lernumwelt wird direkt beeinflusst durch die Häufigkeit des ‚Ability Feedbacks‘.

Die Befundlage bezüglich der Häufigkeit / Frequenz von Lob zeigt, dass Lehrerlob typischerweise selten, willkürlich, eher global als spezifisch und häufig durch individuelle Qualitäten der Schüler bedingt ist und weniger von

48 Stellvertretende Bekräftigung meint, dass durch das öffentliche Loben eines Schülers durch die Lehrperson, z. B. für das schnelle Aufräumen des Arbeitsplatzes, andere Schüler ihrerseits motiviert werden.

49 Im Vergleich zur Grundschule in Deutschland dauert die Grundschulzeit in Australien sechs anstatt vier Jahre.

ihrem Verhalten oder ihren Leistungen, womit Lob aus der Perspektive der Verstärkungstheorie nicht effektiv ist. Auch Grundschulstudien bestätigen den Befund, dass der Einsatz von Lob eher selten zu beobachten ist (zusf. Brophy, 1981). Im Leseunterricht des ersten Schuljahres rangiert die Häufigkeit der Nutzung von Lob zwischen 4 % und 16 % aller schulischen Interaktionen, korrelierte jedoch negativ mit den Leseleistungen der Schüler (vgl. Anderson et al., 1979). Lediglich für die Nutzung von spezifischen ‚Academic Praise‘ konnten Anderson et al. (1979) positive Zusammenhänge mit den Leseleistungen der Schüler bestätigen. Diese spezifische Art des Lobes scheint im Leseunterricht des ersten Schuljahres jedoch gar nicht bis kaum (0–8 %) vorzukommen. Auch die Ergebnisse von La Paro, Pianta und Stuhlman (2004) bestätigen, dass Lehrpersonen in der Grundschule eher kurze generelle Lobbe wie „Gut!“ gegenüber informierendem Feedback bevorzugen, d. h. erweitertes Lob, das das Lernen und Denken der Schüler unterstützt, kommt kaum vor. Diese und andere Studien legen den Schluss nahe, dass auf eine richtige Antwort die Lehrperson nicht mit generellem Lob reagieren, sondern dem Lob eine inhaltliche Erläuterung anschließen sollte (vgl. Brophy & Good, 1986; Hamre & Pianta, 2007).

Inhaltliche Erläuterungen können in Form von Ermutigungen durch die Lehrperson erfolgen. Lob wird als einfachste Art der Ermutigung verstanden, ist jedoch nicht gleichbedeutend damit. Während Lob, wie bereits dargestellt, viel mit Belohnung gemeinsam hat, kann Ermutigung einen Lernvorgang direkt einleiten (vgl. Dinkmeyer & Dreikurs, 1970). Ermutigung hängt eng mit anderen emotionalen Merkmalen wie beispielsweise der Wertschätzung zusammen (vgl. Tausch & Tausch, 1998). Damit Schüler ermutigt werden, muss die Lehrperson Einzelheiten einer Leistung oder des Verhaltens ansprechen. Das reine Lob wird von Tausch und Tausch (1998) hingegen eher als ungünstige Form der Ermutigung angesehen. Als günstige Formen der Ermutigung gelten nach Auffassung der Autoren Achtung und Wärme, positive Zuwendung sowie Verständnis einer Person für andere. Durch ermutigende Worte, wie „Ich bin sicher, dass du zu denen gehörst, die die Aufgabe rechnen können“, ermutigt die Lehrperson den Schüler nicht nur, sondern stärkt gleichzeitig das Selbstvertrauen und die Anstrengungsbereitschaft. In diesem Zusammenhang bedeutet Ermutigen, dass die Lehrperson den Schüler seine eigenen Fähigkeiten und seine Werte erkennen lässt (vgl. Dinkmeyer & Dreikurs, 1970). Die Ermutigung spielt auch dann eine große Rolle, wenn die Lehrperson versucht, die Entwicklung von Ideen zu fördern (vgl. Neuenschwander, 2005).

Aufgrund der eben dargestellten Unterschiede zwischen Lob und Ermutigung werden in der vorliegenden Arbeit beide Aspekte separat betrachtet. Einerseits soll die Häufigkeit von Lob hoch inferent erfasst werden, andererseits soll darauf aufbauend eine Einschätzung getroffen werden, inwieweit Lob eher generell oder spezifisch eingesetzt wird und inwieweit zusätzlich eine Ermutigung durch die Lehrperson folgt. Dalehefte (2001) konnte anhand

der Analyse von sechs Klassen wenig Rückmeldungen feststellen, die seitens der Lehrperson mit Emotionen verbunden sind (z. B. Lob, aber auch Tadel) bzw. positiv unterstützende Rückmeldungen waren kaum Bestandteil des öffentlichen Unterrichtsgesprächs (im Physikunterricht, 8. / 9. Schuljahr). Die Rückmeldungen der Lehrperson in ihrer Studie waren zumeist inhaltlich und nach Meinung der Autorin werden sie vermutlich eher als Zurechtweisung und Kontrolle von den Schülern erlebt, jedoch keineswegs als wertschätzend oder kompetenzunterstützend.

Rückmeldungen der Lehrperson können den Kindern nicht nur signalisieren, dass sie ernst genommen und wertgeschätzt werden (= positiv unterstützende Rückmeldungen), sondern auch, dass ihre Lernaktivitäten sowie -fortschritte gewürdigt werden. Neben positiv unterstützenden Rückmeldungen wie Lob oder Ermutigung können sachlich-konstruktive Rückmeldungen unterschieden werden (vgl. Kobarg & Seidel, 2003; Rakoczy & Pauli, 2006; Rakoczy, 2008). Solche sachlich-konstruktiven Rückmeldungen in Bezug auf die Lernaktivität der Schüler sind wichtige Bedingungen kompetenzunterstützender Maßnahmen. Konstruktive Rückmeldungen zeichnen sich nach Rakoczy und Pauli (2006) (vgl. auch Rakoczy, 2008) durch korrigierende und zukunftsgerichtete Hilfestellungen der Lehrperson aus. Dem Schüler werden demnach Hinweise darauf gegeben, wie man Mängel beheben könnte (vgl. Kobarg & Seidel, 2003). Die sachliche Formulierung einer solchen Hilfestellung kann sowohl inhaltlich als auch prozessbezogen sein. Sachlich-konstruktive Rückmeldungen sollten in einem wohlwollenden Ton erfolgen (vgl. Rakoczy & Pauli, 2006; Rakoczy, 2008). Für die Sekundarstufe konnte ein positiver Zusammenhang zwischen dem Maß an sachlich konstruktiven Rückmeldungen und der Motivation bestätigt werden (vgl. Seidel, 2003b). Das bedeutet, dass Schüler, die häufig sachlich-konstruktive Rückmeldungen erhalten, in einem hohem Maße introjiziert motiviert sind.

Für die frühen Grundschuljahre bestätigen Studien, dass prozessorientiertes Feedback eher selten vorkommt, obwohl Feedback durchgängig als unverzichtbar angesehen wird, wenn es um die kognitive Entwicklung der Schüler geht (z. B. Anderson et al., 1979; Hattie, 2009; La Paro et al., 2004). Pianta et al. (2008) zeigten beispielsweise im Rahmen der Validierung des CLAS-Systems für das dritte Schuljahr deutlich niedrige Ausprägungen für die Qualität des Feedbacks. Im Vergleich zum positiven Klima oder Aspekten der Verhaltenskontrolle erreichen die Klassen auf einer siebenstufigen Skala im Mittel lediglich einen Wert von 3.54. Auch im Kontext der Unterrichtsbeobachtungen mit dem COS-1 (vgl. NICHD ECCR, 2002) ergab sich für die Mehrheit der beobachteten ersten Klassen auf einer siebenstufigen Skala ein Wert zwischen 2 und 3, d. h., dass es in einem Großteil der Klassen an konstruktivem Feedback seitens der Lehrperson mangelt.

#### 4.2.5 *Humorvolle Lernatmosphäre als Aspekt des Unterrichtsklimas*

Nicht nur positive Lehrer-Schüler- bzw. Schüler-Schüler-Beziehungen spielen für das Schaffen von Geborgenheit und Wohlgefühl eine große Rolle. Der Gestaltung der Interaktionen wird zusätzlich eine wesentliche Rolle im Zusammenhang mit dem Unterrichtsklima zugeschrieben. Eine Unterrichtsatmosphäre, die auch mal durch Scherze und Lachen aufgelockert wird, kann das Unterrichtsklima bzw. das Sozialklima und damit die sozialen Interaktionen zwischen den Schülern und der Lehrperson positiv beeinflussen (vgl. Kassner, 2002). Durch den Einsatz von Humor können Lehrpersonen eine entspannte Unterrichtsatmosphäre schaffen, die das Lernklima in der Klasse begünstigt (vgl. Raaf, 2005). Aus diesem Grund wird auch der Einsatz von Humor zu den Merkmalen der Beziehungsförderung gezählt (vgl. Mayr, 2006).

Humor kann von der Lehrperson zusätzlich als Technik der Verhaltenskontrolle bzw. Führungsstrategie eingesetzt werden (vgl. Powell & Anderson, 1985). In diesem Zusammenhang können Lehrpersonen Humor als Mittel nutzen, um unerwünschtes Verhalten von Schülern, z. B. Störungen während des Unterrichts, zu unterbinden. Humor kann außerdem in konkreten Unterrichtssituationen stressvolle Situationen entschärfen bzw. Konflikte lösen (vgl. Veith, 2007), Aufmerksamkeit aktivieren (vgl. Mayr, 2006) und das Behalten von Unterrichtsinhalten fördern (vgl. Perrez, Huber & Geißler, 2001; Powell & Anderson, 1985; Raaf, 2005), d. h. Humor kann gezielt eingesetzt werden, um den jeweiligen Lernstoff zu veranschaulichen. So nennen Lehrpersonen in der Studie von Raaf (2005) Humor als Lernhilfe und Erinnerungsstütze für den Englischunterricht. Kassner (2002) konnte in seinen Untersuchungen zeigen, dass die positiven Emotionen des erlebten Humors im Unterricht bei den Schülern bzw. jungen Mitarbeitern und Auszubildenden bewirkten, dass damit verbundene Lernprozesse positiv besetzt werden und somit besser behalten werden und dass sich die Schüler gleichzeitig wohl, motiviert und ermutigt fühlen.

In der bisherigen Forschung spielt Humor als ein Persönlichkeitsmerkmal der Lehrperson kaum eine Rolle (vgl. Helmke, 2009), was dem Umstand geschuldet ist, dass es zwischen Lehrern und Schülern häufig zu diskrepanten Wahrnehmungen von Humor kommt (vgl. Kassner, 2002). Was Lehrpersonen für einen leicht verständlichen Humor halten, wird von den Schülern eventuell nicht verstanden. Dabei spielt besonders das Alter der Schüler eine entscheidende Rolle. Vor allem Grundschullehrer müssen bei dem Einsatz von Humor bedenken, „dass für viele Arten von Humor ein komplexer kognitiver Prozess zum Verstehen notwendig ist“ (Raaf, 2005, S. 82). Beispielsweise haben junge Schüler Probleme mit dem Verständnis von Ironie oder Satire (ebd.). Nach Veith (2007) ist Humor nur dann wirksam und verständlich, wenn die Form und der Inhalt des Humors dem Verständnishorizont, dem Wortschatz sowie dem Alter des Schülers entsprechen. Grundschullehrer müssen somit bei dem Einsatz von Humor im Unterricht bedenken, dass ih-

re Art von Humor eventuell wenig Anklang findet, denn Schüler lachen über Witze, „die ihren kognitiven Voraussetzungen angemessen sind [...]“ (Rißland, 2002, S. 84).

Bestimmte Formen des Humors können jedoch auch negative Effekte auslösen. Beispielsweise kann ein herabsetzender Humor die Lehrer-Schüler-Beziehung beeinträchtigen und somit Auswirkungen auf die Atmosphäre haben. Außerdem kann ein Zuviel an Humor durch die Lehrperson dazu führen, dass die Schüler den Lehrer nicht mehr ernst nehmen. Um einen positiven Einfluss auf die Lernprozesse von Schülern zu haben, muss Humor zusätzlich einen Bezug zum Thema haben und in optimaler Dosis dargeboten werden (vgl. Ziv, 1988).

Die Befundlage zum Einfluss von Humor auf die Leistung präsentiert sich uneinheitlich. Einerseits existieren Studien, die eine Leistungssteigerung durch Humor belegen, andererseits konnten einzelne Studien keine Effekte bestätigen (zusf. Powell & Anderson, 1985; Ziv, 1988). Über den Effekt von Humor auf das Selbstkonzept von Schülern sind m. W. n. bislang keine Studien bekannt. Es kann jedoch vermutet werden, dass Humor eine angemessene Ressource darstellen kann, um Misserfolgssituationen erfolgreicher zu verarbeiten und somit das Selbstkonzept der Schüler nicht allzu sehr geschwächt wird. Ein herabsetzender Humor, der auf Kosten einzelner Schüler eingesetzt wird, kann jedoch einen negativen Effekt auf die Motivation und das Selbstkonzept einzelner Schüler haben.

#### *4.2.6 Wettbewerbs- und Leistungsdruck als Merkmal des Unterrichtsklimas*

Ergänzend zu den bereits genannten Merkmalen kann das Unterrichtsklima durch ein gewisses Maß an Leistungs- und Konkurrenzdruck geprägt sein (vgl. Toman, 2007; Eder, 1998). Jerusalem und Schwarzer (1991) definieren den subjektiv wahrgenommenen Leistungsdruck als eine kognitive Einschätzung der schulischen Anforderungen durch den Schüler. Je nachdem wie groß der Leistungsdruck in einer Klasse ist, hat dieser positive oder negative Folgen für die Schüler und für das Unterrichtsklima in der Klasse. Von Saldern und Littig (1987) verstehen den Leistungsdruck als Aspekt des Unterrichts und fassen darunter eher Merkmale wie mangelnde Zeit für die Schüler zum selbstständigen Arbeiten, zu wenige Wiederholungen, zu viele Klassenarbeiten und die zu rasche Abfolge umfangreicher Lerninhalte. Nach Eder (1998) erfasst der Leistungsdruck in einer Klasse „das quantitative Ausmaß der Lern- und Leistungsanforderungen, mit denen die Schüler und Schülerinnen konfrontiert sind. Hoher Leistungsdruck bedeutet demzufolge, daß in einer Klasse hohe quantitative Ansprüche gestellt werden, die ein hohes Ausmaß an Lernzeit und Lernanstrengung erfordern“ (S. 20).

In der Diskussion um die Folgen von Leistungsdruck wird immer wieder auf die negativen Zusammenhänge mit Schülerleistungen hingewiesen (z. B. Eder, 1996; Fend, 2001). Versteht man Leistungsdruck im Sinne einer „objektiven Überforderung“ (Satow, 1999, S. 61), erhält der Begriff eher eine negative Konnotation und es resultieren negative Folgen für die Leistungen der Schüler (vgl. Eder, 1986; 1996). Sobald man jedoch versucht, Leistungsdruck als „Aspekt des Anforderungsniveaus“ (Satow, 1999a, S. 61) zu betrachten, wird ihm ein positiver Effekt auf die Leistungen zugeschrieben (vgl. Eder, 1986). Somit kann ein gewisses Maß an Leistungsdruck bzw. Wettbewerb in der Klasse nicht nur als Bedrohung oder Schädigung wahrgenommen werden, sondern auch als Herausforderung oder als Anreiz (vgl. Schwarzer, 1984; Satow, 1999a).

Ein hoher Leistungsdruck geht häufig einher mit einem erlebten Konkurrenzdruck in der Klasse. Nach Schwarzer (1983) gilt die Schaffung von Konkurrenzdruck bzw. Wettbewerb als bewährtes pädagogisches Mittel zur kurzfristigen Motivierung von Schülern, kann jedoch auch disziplinierend wirken (vgl. Eder, 1986). Demnach kann Wettbewerb verhaltenswirksam sein, indem mehr Zeit für schulisches Lernen investiert wird (vgl. Fend, 2001; Eder, 1998). Mangelt es in einer Klasse jedoch an bewältigbaren Herausforderungen, wird der Unterricht von den Schülern eher frustrierend eingeschätzt (vgl. Oswald, Pfeifer, Ritter-Berlach & Tanzer, 1989). Der erlebte Konkurrenzdruck bezieht sich auf Aspekte des sozialen Vergleichs (vgl. Festinger, 1954) und des Wettbewerbs (vgl. Jerusalem & Schwarzer, 1991), die wiederum einen Ausgangspunkt für die Bewertung des Selbst darstellen (vgl. Schwarzer, 1983, S. 131). Studien über alle Schulformen hinweg (Ausnahme Grundschule) bestätigen, dass ein hoher Leistungs- und Konkurrenzdruck dazu führt, dass Schüler in solchen Klassen ein geringeres Selbstbewusstsein haben (Eder, 1996; Fend, 1977; Fend, 2001). In leistungsbetonten Klassen, in denen das Klassenklima auf Konkurrenz ausgerichtet ist und zusätzlich kein geordneter Unterrichtsverlauf erkennbar ist, können nur einzelne, vermutlich leistungsstärkere Schüler ein positives Selbstkonzept aufbauen. Dreesmann (1979) konnte in seiner Studie außerdem zeigen, dass die Konkurrenz als Klimamerkmale in Beziehung mit dem Zurückführen schlechter Leistungen auf mangelnde Anstrengung steht. Weitere Folgen von zu hohem Leistungsdruck können Resignation und Angst sein (vgl. Eder, 1998; Fend, 2001; Oswald et al., 1989). Leistungsdruck hat außerdem Auswirkungen auf die Lernanstrengungen bzw. Leistungsbereitschaft (vgl. Fend, 2001) sowie die Mitarbeit der Schüler im Unterricht (vgl. Eder, 1996; Eder, 1998). Im Unterricht äußert sich Leistungsdruck unter anderem in Situationen, in denen die Lehrperson Leistungen einzelner Schüler öffentlich rückmeldet, indem sie Noten oder andere Formen der expliziten Beurteilung oder Bewertung von Leistungen (z. B. Vergabe von Punkten) thematisiert.

Abschließend fasst die Abbildung 8 nochmals alle für die vorliegende Arbeit relevanten Merkmale des Unterrichtsklimas zusammen. Auch in der vor-

liegenden Arbeit werden unter dem Unterrichtsklima eine Reihe von Merkmalen subsumiert werden, die unterschiedlichen Forschungstraditionen entstammen. Einerseits wird auf Forschungstraditionen zum Humor im Unterricht eingegangen, andererseits bezieht sich die Mehrzahl der hier einbezogenen Aspekte auf die Selbstbestimmungstheorie von Deci und Ryan (1993). Vor allem die Förderung des Kompetenzerlebens und Kontexte, in denen Schüler sich sozial eingebunden fühlen, werden im Rahmen der Arbeit betrachtet.

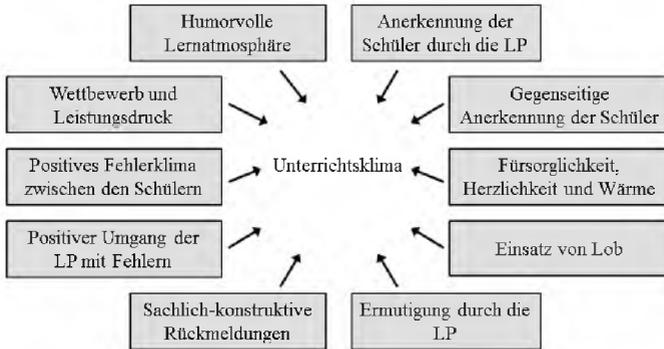


Abbildung 8: Untersuchte Merkmale des Unterrichtsklimas im Überblick (LP = Lehrperson)

## 5 Unterrichtsqualitätsmerkmale und multikriteriale Effekte

In der empirischen Unterrichtsforschung wird die Qualität von Unterricht vor allem unter dem Gesichtspunkt der lernförderlichen Wirkungen auf die Schüler betrachtet (zuseh. u. a. Brophy & Good, 1986; Einsiedler, 1997; Helmke & Weinert, 1997a). Während in den 50er und 60er Jahren hauptsächlich die Lehrperson und deren Persönlichkeitseigenschaften im Fokus zahlreicher Untersuchungen standen, setzte sich im Verlauf der darauffolgenden Jahrzehnte die Auffassung durch, dass Verhaltensmuster bzw. Unterrichtsstile guter Lehrpersonen im Zusammenhang mit Leistungszuwächsen bzw. Lernerfolg stehen (Prozess-Produkt-Paradigma, vgl. Helmke, 2009). Auf Basis von Korrelationen wurden vor allem Zusammenhänge zwischen Merkmalen des Lehrerhandelns und Unterrichtswirkungen (z. B. Leistungsentwicklung) betrachtet (zuseh. Brophy & Good, 1986).

In den 1990er Jahren erfuhr das Prozess-Produkt-Modell durch die zunehmende Orientierung an sogenannten Angebots-Nutzungs-Modellen (vgl. Fend, 1998) wesentliche Erweiterungen. Das Angebots-Nutzungs-Modell

fasst wichtige Aspekte des Unterrichts und dessen Wirkungen nicht nur in korrelativer Weise, wie das Prozess-Produkt-Paradigma, zusammen, sondern integriert die Wirkungsweise und Zielkriterien des Unterrichts in einem umfassenderen und differenzierteren Modell und kann somit als Ordnungsrahmen verstanden werden (vgl. Helmke, 2009). Die Qualität von Unterricht gilt in diesem Modell, in Anlehnung an Fend (1998), als Ko-Produktion von Lehrpersonen und Schülern und beinhaltet eine Fülle von Merkmalen. Neben Merkmalen der Lehrperson bzw. Lehrerpersönlichkeit berücksichtigt das Angebots-Nutzungs-Modell ergänzend zum Prozess-Produkt-Modell Merkmale des Unterrichts, des Klassenkontextes (kognitive und soziale Merkmale der Klasse, z. B. mittlere Klassenleistung, Klassenzusammensetzung), der individuellen Voraussetzungen sowie die Mediationsprozesse und die Lernaktivitäten auf Schülerseite. Eine lernförderliche Wirkung (= Ertrag) einzelner Unterrichtsmerkmale (= Angebot) resultiert aus dem Zusammenwirken dieser Vielzahl von unterrichtlichen und kontextuellen Merkmalen und deren Nutzung durch die Schüler.

Als eine entscheidende Determinante des Unterrichtsangebots gilt die Prozess- bzw. Unterrichtsqualität, die wiederum über fächerübergreifend sowie fachspezifisch gültige Merkmale mediiert wird. Fachübergreifend gültige Merkmale sind unter anderem eine effektive Klassenführung im Sinne eines effektiven Zeitmanagements und der Einführung und Etablierung von Regeln zur Vermeidung von Unterrichtsstörungen oder das Lernklima in der Klasse (= Basisdimensionen, vgl. Kapitel 1.2). Neben der Prozessqualität des Unterrichts spielt ebenso die Unterrichtsquantität im Sinne der Zeitnutzung eine große Rolle. Die Nutzung der zur Verfügung stehenden Zeit stellt nicht nur in der Grundschule eine wichtige Determinante des Lernerfolgs der Schüler dar (vgl. Helmke & Weinert, 1997b; Helmke et al., 2008).

Das Angebots-Nutzungs-Modell weist zusätzlich auf unterschiedliche Zielkriterien des Unterrichts hin, indem der Lernerfolg (= Wirkungen) nicht nur fachliche Effekte (Fachwissen oder Fertigkeiten) beinhaltet, sondern auch affektiv-motivationale Merkmale aufseiten der Schüler aufführt. Zwar wird in diesem Zusammenhang häufig von der „Konkurrenz von Zielkriterien“ (Helmke, 1999) oder „Zielkonflikten“ (Baumert, 1997) gesprochen, jedoch bestätigt eine wachsende Anzahl an aktuellen Studien, dass sowohl für die Sekundarstufe als auch für die Grundschule die Förderung verschiedener Zielbereiche im Unterricht grundsätzlich kompatibel ist (vgl. Baumert & Köller, 2000; Weinert, 1996; Gruehn, 1995; 1996; Helmke & Schrader, 1990; Kammermeyer & Martschinke, 2009; Kunter, 2005; Kunter & Voss, 2011; Schrader et al., 1997; Weinert & Helmke, 1996). Dies hat zur Folge, dass die Bedeutung verschiedener unterrichtlicher Qualitätsmerkmale nicht nur für die Leistungsentwicklung, sondern zunehmend auch für die Entwicklung affektiv-motivationaler Schülermerkmale untersucht wird.

Vor allem in der Grundschule ist die simultane Betrachtung kognitiver und nicht-kognitiver Schülermerkmale von besonderer Bedeutung, da es be-

reits hier um das Erreichen „gute[r] Leistungen in den Lernbereichen und [die] Förderung von Zielen im Persönlichkeits- und motivationalen Bereich“ (Einsiedler, 2003, S. 309) geht. Als nicht-kognitive Ziele gelten unter anderem die Selbstständigkeit aber auch affektiv-motivationale Ziele, wie das Interesse, die Lernfreude oder ein positives Selbstkonzept<sup>50</sup> (vgl. Gautschi, Moser, Reusser & Wiher, 2007). Auch nach Kammermeyer und Martschinke (2009) zeichnet sich Unterricht in der Grundschule dadurch aus, „dass er multikriteriale Ziele erreicht und als Kriterium für den Schulerfolg die Leistungsentwicklung und die Persönlichkeitsentwicklung gleichermaßen berücksichtigt“ (S. 36). Die Autorinnen sprechen in Anlehnung an Einsiedler (1988; vgl. auch Einsiedler, 2001) dem Erreichen von nicht-kognitiven Zielen, im Fall ihrer Studie der Entwicklung eines positiven Selbstkonzepts (als affektiv-motivationale Zielkriterium, vgl. Petillon, 1997), eine zweifache Bedeutung zu. Einerseits steht die Selbstkonzeptentwicklung in einer wechselseitigen Beziehung mit der Leistungsentwicklung,<sup>51</sup> andererseits haben Selbstkonzepte einen bedeutenden Eigenwert, da sie ein wesentliches Persönlichkeitsmerkmal darstellen (vgl. Einsiedler, 1988).

Besonders der Schulanfang und die neuen Anforderungen haben Auswirkungen auf die Persönlichkeit (vgl. Einsiedler, 1988). Die Schüler bekommen „im Unterricht erstmals regelmäßig Informationen über ihre Leistung und sind damit dem sozialen Vergleich in einer relativ altershomogenen Klasse ausgesetzt“ (Kammermeyer & Martschinke, 2003, S. 487), was wiederum Auswirkungen auf das noch sehr labile Selbstkonzept der Schüler hat. Schlechte bzw. gute Leistungen werden jedoch nicht von Beginn an als Ursache für entsprechende Selbstkonzepte herangezogen. Der Zusammenhang zwischen der Leistung und dem Selbstkonzept wird erst im Verlauf der ersten Schuljahre enger (vgl. Kammermeyer & Martschinke, 2003; Kastens, Gabriel & Lipowsky, 2013), sodass besonders in den ersten Schuljahren der Qualität des Unterrichts eine größere Bedeutung für die Selbstkonzeptentwicklung zukommen kann als in darauffolgenden Jahren. Um positive Selbstkonzepte bei den Schülern auszubilden und somit auch die Leistungsentwicklung der Schüler positiv zu beeinflussen, kann die Lehrperson neben positiven Leistungsrückmeldungen genügend aktive Lernzeit zur Verfügung stellen oder

---

50 Im Rahmen der vorliegenden Arbeit wird der Begriff Selbstkonzept verwendet. Für die Darstellung der Wirkungen des Unterrichtsklimas werden jedoch auch Studien herangezogen, die Effekte des Unterrichtsklimas auf das Fähigkeitsselbstkonzept, das akademische Selbstkonzept, die Selbstwirksamkeit, das Selbstwertgefühl etc. untersuchen.

51 Auf den Ursache-Wirkungs-Zusammenhang von Leistung und Selbstkonzept wird an dieser Stelle nicht genauer eingegangen. Für das erste Schuljahr konnte der Self-Enhancement-Ansatz bestätigt werden, d. h. das Selbstkonzept hat einen Effekt auf spätere Leistungen. Im zweiten Schuljahr kippt diese Wirkrichtung jedoch. Die Höhe des Pfades von der Leistung auf das Selbstkonzept fällt zu diesem Zeitpunkt höher aus als der vom Selbstkonzept auf die Leistung (vgl. Kammermeyer & Martschinke, 2006).

ein angenehmes Unterrichtsklima schaffen, in dem die Schüler wertgeschätzt und gelobt werden.

*Basisdimensionen der Unterrichtsqualität und deren vermutete Wirkmechanismen:*

Im Kontext der TIMSS-Videostudie wurden die postulierten Wirkungen der drei Basisdimensionen (vgl. Kapitel 1.2), unter Berücksichtigung der Vermittlungsprozesse (Wahrnehmung und Nutzung der Unterrichtsangebote durch die Schüler) zusammenfassend dargestellt (vgl. Klieme & Rakoczy, 2008; Klieme et al., 2006). Das theoretische Modell (vgl. Abbildung 9) korrespondiert mit dem Konzept der direkten Instruktion. Die direkte Instruktion wird häufig im Zusammenhang mit einer effektiven Klassenführung genannt und beschreibt nach Klieme et al. (2006) „einen eher lehrergesteuerten, auf effiziente Zeitnutzung fokussierten Unterricht, schließt in seiner ursprünglichen Formulierung aber auch die explizite Vermittlung kognitiver Strategien und korrigierendes Feedback – beispielsweise zu Hausaufgaben – mit ein“ (S. 128). Ein solcher Unterricht fördert vor allem Schüler mit ungünstigen Lernvoraussetzungen (ebd.). Da guter Unterricht nicht nur die Leistung der Schüler steigern, sondern vor allem in der Grundschule auch die Persönlichkeitsentwicklung unterstützen sollte (im Sinne der Multikriterialität von Unterricht), wird in dem Modell von Klieme et al. (2006) auch die Selbstbestimmungstheorie von Deci und Ryan (1993) aufgegriffen. Im Sinne der Selbstbestimmungstheorie profitieren Schüler in motivationaler Hinsicht von Unterricht, wenn sie von der Lehrperson in ihrem Streben nach Autonomie, Kompetenz und sozialer Eingebundenheit unterstützt werden. Vor allem wertschätzende Beziehungen zwischen der Lehrperson und den Schülern sowie positive Rückmeldungen (vgl. Kapitel 4.2), aber auch ein strukturierter Unterrichtsablauf gehören zu denjenigen Merkmalen, die das emotionale Erleben unterstützen (vgl. Kunter, 2005; Rakoczy, 2006).

Zusätzlich werden in dem Wirkungsmodell Hypothesen zur Wahrnehmung und Nutzung der Lernangebote (Vermittlungsprozesse) integriert. Das Modell postuliert der Klassenführung über die Wahrnehmung der aktiven Lernzeit (Time on Task) einen Effekt auf die Leistungen der Schüler im Sinne eines konzeptuellen Verständnisses, während ein unterstützendes Unterrichtsklima über das Erleben von Autonomie, Kompetenz und sozialer Eingebundenheit Auswirkungen auf die Motivation bzw. auf affektiv-motivationale Schülermerkmale hat. Jedoch kann auch ein strukturierter Unterrichtsablauf über diese Vermittlungsprozesse emotionales Erleben unterstützen (vgl. Rakoczy, Klieme, Drollinger-Vetter, Lipowsky, Pauli & Reusser, 2007).

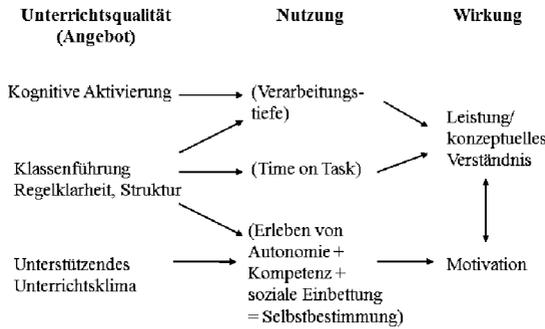


Abbildung 9: Basisdimensionen der Unterrichtsqualität und deren vermutete Wirkmechanismen (nach Klieme, Lipowsky, Rakoczy & Ratzka, 2006, S. 131)

*Personenzentrierter versus variablenzentrierter Ansatz:* Empirische Studien liefern Ergebnisse zur multikriterialen Zielerreichung häufig in Form von Clusteranalysen bzw. durch Betrachtung von Profilen erfolgreicher Klassen (z. B. Gruehn, 1995; Helmke & Schrader, 1990; Kammermeyer & Martschinke, 2009; Kunter, 2005; May, 2001; Schrader, Helmke & Dotzler, 1997; Stuhlman & Pianta, 2009; Weinert & Helmke, 1996). In diesem Zusammenhang spricht man von einem *personenzentrierten* Ansatz (vgl. Helmke, 2009; Weinert, Schrader & Helmke, 1990b). Studien, die diesem Ansatz folgen, identifizieren in einem ersten Schritt beispielsweise sogenannte Optimalklassen oder Best-Practice-Gruppen, um diese dann in einem weiteren Schritt in ihrer Ausprägung einzelner Unterrichtsqualitätsmerkmale zu untersuchen. Mithilfe dieses personenzentrierten Vorgehens konnte mehrfach bestätigt werden, dass sich erfolgreiche Klassen u. a. durch eine effektive Klassenführung und durch ein positives Sozialklima charakterisieren lassen (z. B. Weinert & Helmke, 1996; Schrader et al., 1997).

Im Gegensatz zum personenzentrierten Ansatz geht der *variablenzentrierte* Ansatz davon aus, „dass der Unterricht verschiedener Lehrpersonen, aber auch bei der gleichen Lehrkraft in verschiedenen Situationen (Fächer, Klassen, Zeitpunkte), hinsichtlich bestimmter Merkmale (Variablen) variieren kann [...]“ (Helmke, 2009, S. 27). Studien, denen dieser Ansatz zugrunde liegt, untersuchen beispielsweise, ob unterschiedliche Ausprägungen einzelner Variablen (z. B. Merkmale der Unterrichtsqualität) mit Unterschieden im Lernerfolg oder in der Entwicklung affektiv-motivationaler Schülermerkmale einhergehen. Aktuelle Studien wählen dabei zunehmend Verfahren, die u. a. die Mehrebenenstruktur der Daten berücksichtigen. – An diesem Punkt setzt die vorliegende Arbeit an. Um die Bedeutung der Unterrichtsqualitätsmerkmale Klassenführung und Unterrichtsklima in Bezug auf die Leistungen und Selbstkonzepte in den ersten beiden Schuljahren in den Fächern Deutsch und Mathematik zu untersuchen, werden in dieser Studie Mehrebenenregressions-

modelle mit *MPlus* (Version 5; Muthén & Muthén, 1998–2007) gerechnet. Im Unterschied zur Berechnung einfacher korrelativer Zusammenhänge greift *MPlus* auf die multiple Regressionsanalyse zurück und erlaubt die gleichzeitige Betrachtung mehrerer Prädiktorvariablen (vgl. Geiser, 2010; Kapitel 16) unter Kontrolle der Mehrebenenstruktur. Anhand ausgewählter Studien wird nachfolgend die Bedeutung der Klassenführung und des Unterrichtsklimas für multikriteriale Schülermerkmale in der Grundschule dargestellt.

## 5.1 Klassenführung und multikriteriale Effekte

### 5.1.1 *Klassenführung und Leistung*

Die positiven Wirkungen einer effektiven Klassenführung auf die Leistungsentwicklung der Schüler in den verschiedenen Fächern sind mittlerweile vielfach dokumentiert (zusf. Einsiedler, 1997; Evertson & Weinstein, 2006; Gruehn, 1995; Helmke, 1992; Helmke & Schrader, 1990; Helmke et al., 2002; Helmke, Schneider & Weinert, 1986; Helmke et al., 2008; Klieme et al., 2008; Kunter, 2005; Kunter et al., 2006; Lipowsky et al., 2009; Seidel & Shavelson, 2007; Wang et al., 1993b; Walberg & Paik, 2000). Vor allem die Wirksamkeit des Prozessmerkmals Klassenführung im Sinne einer effizienten Zeitnutzung und Störungsprävention ist immer wieder bestätigt worden (vgl. Brophy & Good, 1986; Creemers, 1994a; b; Hattie, 2009; Helmke, 1992; Helmke & Schrader, 1993; Klieme et al., 2001; Lipowsky et al., 2008; Lipowsky et al., 2009; Wang et al., 1990; Wang et al., 1993b). Bereits Carroll (1963) beschrieb in seinem Modell schulischen Lernens die besondere Bedeutung der Lernzeit als Aspekt der Klassenführung für den Lernerfolg eines Schülers (vgl. Kapitel 1.1). Lehrpersonen, deren Klassen einen überdurchschnittlichen Lernerfolg haben, nutzen die verfügbare Unterrichtszeit für die Vermittlung und Einübung des Lernstoffes und weniger für organisatorische Angelegenheiten (vgl. Helmke, 1992).

Für die Sekundarstufe konnte Helmke (1992) beispielsweise im Rahmen der Münchner Hauptschulstudie einen positiven Zusammenhang zwischen der Effektivität der Klassenführung und dem Leistungszuwachs in Mathematik zeigen. Lehrer von Klassen mit überdurchschnittlichem Leistungszuwachs in Mathematik praktizieren ein wirksames Klassenmanagement, sie haben ihre Klasse – umgangssprachlich – „im Griff“. In diesen Klassen kommen Disziplinstörungen selten vor oder werden sofort durch Eingreifen der Lehrperson unterbunden. Ein hoher Anteil an Disziplinproblemen innerhalb der Klasse hat wiederum negative Auswirkungen auf das Leistungsniveau (vgl. Klieme & Rakoczy, 2003). Auch die Befunde der COACTIV-Studie bestätigen, dass insbesondere die Klassenführung, aber auch das Potenzial zur kognitiven Aktivierung, die Mathematikleistungsentwicklung bedeutsam beeinflusst.

sen (vgl. Baumert et al., 2010). Die Resultate von Gruehn (1995) zeigen, dass Unterrichtsmerkmale wie Zeitverschwendung, ein hoher Leistungsdruck und ein hohes Interaktionstempo negative Auswirkungen auf die Entwicklung der Leistungen in Mathematik haben, während die Unterrichtsvariablen Disziplin, Strukturiertheit sowie die individuelle Bezugsnormorientierung positiv mit den Leistungen der Schüler zusammenhängen. Für das Fach Deutsch konnten Klieme et al. (2008) hingegen keinen Effekt der wahrgenommenen Klassenführung und Strukturiertheit (Schülersicht) auf die Entwicklung der Lesekompetenz und Sprachbewusstheit bis Ende des neunten Schuljahres bestätigen. Ein erwartungsgemäß positiver, jedoch schwacher Effekt zeigte sich allerdings für die wahrgenommene Disziplin im Unterricht auf den Leistungszuwachs im Bereich Sprachbewusstheit.

Betrachtet man Befunde aus der videobasierten Unterrichtsforschung in der Sekundarstufe, dann lassen sich teilweise keine oder nur sehr geringe Zusammenhänge der im Rahmen von Videostudien erfassten Klassenführung mit den Leistungen der Schüler bestätigen (z. B. Klieme, Schümer & Knoll, 2001; Lipowsky et al., 2008). Beispielsweise konnte im Rahmen der Pythagorasstudie ein positiver, wenn auch geringer Zusammenhang zwischen dem Classroom Management und dem Lernerfolg der Schüler im Fach Mathematik aufgezeigt werden (vgl. Lipowsky et al., 2009).

Ähnlich wie bereits für die Sekundarstufe kann auch für den Grundschulbereich der positive Einfluss einer effektiven Klassenführung für die Leistungen der Schüler vielfach bestätigt werden (z. B. Anderson et al., 1979; Cadima et al., 2010; Helmke et al., 2010; Helmke & Schrader, 1990; Helmke & Weinert, 1997b; Helmke et al., 2008; Klicpera & Gasteiger-Klicpera, 1993; Perry et al., 2007; Pianta et al., 2008). Allerdings ist die Befundlage für die verschiedenen Fächer bzw. Lernbereiche eher uneinheitlich (vgl. Tabelle 8).

Während die Ergebnisse aus der SCHOLASTIK-Studie (3./4. Schuljahr) für die Klassenführung (ermittelt durch hoch inferente Ratings geschulter Unterrichtsbeobachter) einen positiven Zusammenhang mit dem Leistungszuwachs im Fach Mathematik bestätigen ( $r = .36, p < .01$ ), zeigte sich für den Bereich Rechtschreiben kein signifikanter Zusammenhang (vgl. Helmke & Weinert, 1997b). Allerdings geht der Zusammenhang in die erwartete positive Richtung ( $r = .26, n.s.$ ). Zu ähnlichen Ergebnissen für den Rechtschreibunterricht gelangen auch Roos und Schöler (2009) im Rahmen der EVES-Studie. Für die beobachtete Effektivität der Klassenführung ergab sich kein signifikanter Zusammenhang mit den Rechtschreibleistungen am Ende des vierten Schuljahres (vgl. auch Hofmann, 2007). Zur Erfassung der Klassenführung und weiterer Unterrichtsmerkmale wurde der Unterricht der Lehrpersonen zu insgesamt zehn Messzeitpunkten beobachtet. Aufgrund der relativ kleinen Stichprobe von zehn Klassen sollten die Befunde jedoch mit Vorsicht betrachtet werden.

Inwieweit sich die Befunde für den Bereich Rechtschreiben aus den späteren Grundschuljahren auch für den Anfangsunterricht bestätigen lassen, untersuchten unter anderem Klicpera und Gasteiger-Klicpera (1993). Die Au-

toren kamen zu dem Ergebnis, dass in Klassen mit einem hohem Leistungsstand am Ende des ersten Schuljahres die Lehrperson seltener disziplinieren musste, reingeringelte Beiträge der Schüler von der Lehrperson eher ermahnt wurden und die Schüler den Aufforderungen der Lehrperson häufiger nachkamen als in Klassen mit einem geringeren Leistungsstand. Am Ende des zweiten Schuljahres sind die Unterschiede zwischen den Klassen mit ehemals hohem bzw. geringem Leistungsstand bereits deutlich geringer. Beobachter schätzten Klassen mit einem besseren Leistungsstand zusätzlich als ruhiger ein. Außerdem sind diese Klassen durch ein positives Arbeitsklima geprägt, die Lehrpersonen werden von den Beobachtern als engagierter eingestuft und gelten als besser vorbereitet. Gegenüber dem Bereich Lesen, für den der positive Einfluss einer effektiven Klassenführung auf die Leistungen der Schüler für die ersten Schuljahre vielfach bestätigt werden konnte (z. B. Klicpera & Gasteiger-Klicpera, 1993; Pianta et al., 2008), ist die Befundlage für den Bereich Rechtschreiben rar. Nach Klicpera und Gasteiger-Klicpera (1993) kann der Mangel an Studien vor allem darauf zurückgeführt werden, dass das Schreiben im ersten Schuljahr noch keinen so großen Stellenwert hat wie das Lesen, da die Entwicklung des Schreibens langsamer verläuft und die Aneignung von Schreibweisen bzw. orthographischen Charakteristiken oftmals die gesamte Grundschulzeit andauert.

Im Rahmen der ECCE-Studie konnte für das zweite Schuljahr gezeigt werden, dass keines der erfassten Merkmale der Unterrichtsqualität (vgl. Tabelle 1, S. 33) – mit Ausnahme des Anwendungsbezugs – mit den Leistungen der Schüler zusammenhängt. Roßbach (2002b) vermutet, dass „es sein könnte, dass in der Anpassungsphase der Schüler an die Anforderungen der Grundschule ein direktes Unterrichten noch nicht so bedeutsam ist wie in den späteren Jahren der Grundschule“ (S. 243). Ähnliches könnte man auch im Zusammenhang mit der Bedeutung des Klassenmanagements vermuten. Die Befunde von Roßbach (2002b) sind jedoch mit Vorsicht zu betrachten, da die Analysen der Auswirkungen des Klassenmanagements sowie weiterer Unterrichtsqualitätsmerkmale auf die Schulleistungen nur auf Individualenebene möglich waren. Außerdem wurden die Leistungen mit einem Test erfasst, der sich aus den Subskalen Satzverständnis, Rechnen, Lösen angewandter Probleme sowie aus Kenntnissen über die natürliche und soziale Umwelt zusammensetzt. Da für die Berechnung der Analysen ein Gesamtscore pro Schüler gebildet wurde, können auf Basis der Befunde aus der ECCE-Studie keine Aussagen über Effekte der Unterrichtsqualitätsmerkmale in Abhängigkeit von verschiedenen Fächern gemacht werden.

Tabelle 8: Bedeutung der Klassenführung für die Leistungen in der Grundschule – Eine ...

Studie u. Autoren	Schuljahr	Leistungsbereich	Datenquelle z. Erfassung der Klassenführung
<i>Deutschsprachige Studien ...</i>			
SCHOLASTIK-Studie: Helmke & Weinert (1997b)	3. SJ 4. SJ	Leistungszuwachs in: Mathematik Rechtschreiben	Beobachter
EVES-Studie: Roos & Schöler (2009; vgl. auch Hofmann, 2007)	Ende 4. SJ	Leistungsniveau im Rechtschreiben	Beobachter
Klicpera & Gasteiger-Klicpera (1993)	1. SJ 2. SJ	Leistungsniveau Rechtschreiben	Beobachter
ECCE-Studie: Roßbach (2002a; b)	2. SJ	Leistungsniveau (Gesamtskala: Satzverständnis, Rechnen, Lösen angewandter Probleme, Kenntnisse über die natürliche und soziale Umwelt)	Beobachter
<i>Internationale Studien ...</i>			
Pianta et al. (2008)	1. SJ 3. SJ 5. SJ	Leistungen in Mathematik (und Lesen*)  * wird im Rahmen der vorliegenden Arbeit nicht untersucht. Daher werden die Ergebnisse hier nicht präsentiert.	Beobachter
Cadima et al. (2010)	1. SJ	Leistungen in: (1) Wortschatz (Vocabulary) (2) Textverständnis (Print Concepts) (3) Zahlenlesen (Number Identification)	Beobachter

... Gegenüberstellung ausgewählter Studien

Operationalisierung der Klassenführung	Analyse-methode	Ergebnisse**
<i>... Deutschsprachige Studien</i>		
Klassenführung (KF): (1) Intensität der Zeitnutzung für die Stoffbehandlung (2) Effizienz der Unterrichtsorganisation (3) Schnelle und gleitende Übergänge zwischen verschiedenen Unterrichtsphasen	Korrelation	(1) Positiver Zusammenhang zwischen KF und Leistungszuwachs (2) Kein Zusammenhang zwischen KF und Leistungszuwachs
Klassenführung (KF): (1) Regel- und Ritualverwendung (Existenz eines Regelsystems) (2) Klassenmanagement und Kontrolle (effektiver Umgang mit Störungen) (3) Stillarbeit / Kontinuitätsaspekte (Sicherstellung des störungsfreien Ablaufs von Stillarbeitsphasen) (4) Zeitnutzung (Minimierung des Zeitaufwandes für außerschulische Angelegenheiten)	Korrelation	Kein Zusammenhang zwischen KF und Leistungsniveau
Klassenführung (KF): (1) Reaktion auf Unterrichtsunterbrechungen (2) Disziplinierung (3) Reaktion auf Herausrufen einer Lösung (durch Schüler) (4) Lärm und Unruhe in der Klasse	Deskriptive Statistiken	Klassen mit hohem Leistungsstand: selten Disziplinprobleme; weniger Lärm und Unruhe
Klassenmanagement: (1) Fließende Übergänge (2) Klare Regeln über das Verhalten in der Klasse (3) Hohes Schülerengagement (4) Effektiver Umgang mit Störungen (5) Klima wechselseitigen Respekts (6) Soziale Unterstützung (7) Ermutigung schwächerer Schüler	Korrelation / Partialkorrelation	Kein Zusammenhang zwischen Klassenmanagement und Leistungsniveau
<i>... Internationale Studien</i>		
Instructional Support: (1) Effiziente Zeitnutzung (Productive Use of Time) (2) Methodenvielfalt (Richness of Instructional Methods) (3) Quantität der Auseinandersetzung (Quantity of Exposure)	Change Growth Mixture Models (GMM)	Für alle SJ: keine Effekte des Instructional Supports  3./5. SJ: Effekt der effektiven Zeitnutzung (Quantity of Exposure) auf Leistungen
Classroom Organization: (1) Verhaltensmanagement (Behavior Management) (2) Produktivität (Productivity) (3) Lehr-Lern-Formate (Instructional Learning Formats)	Mehrebenen-analyse	(1) / (2) positiver Effekt der Classroom Organization auf die Leistungen (3) kein Effekt der Classroom Organization auf die Leistungen

\*\* Die Nummern in Klammern beziehen sich auf die Spalte Leistungsbereich.

Dass auch für das erste Schuljahr fachspezifische Unterschiede in der Bedeutung einer effektiven Klassenführung angenommen werden können, zeigt die amerikanische Studie von Pianta et al. (2008). Mithilfe des COS (Classroom Observation System, vgl. Kapitel 2.1, Tabelle 1) wurden die zwei Unterrichtsqualitätsmerkmale ‚Emotional Quality‘ und ‚Instructional Quality‘<sup>52</sup> auf Basis von Unterrichtsbeobachtungen erfasst und deren Effekte auf die Leistungen in drei Schuljahren (1., 3., 5. Schuljahr) in den Bereichen Mathematik und Lesen untersucht. Pianta et al. (2008) überprüften, inwieweit eine veränderte Unterstützung (Emotional und Instructional Support) die Leistungen im ersten, dritten und fünften Schuljahr in Mathematik beeinflusst. In Anlehnung an bisherige Studien erwies sich für den Bereich Mathematik die Höhe der zeitlichen Auseinandersetzung mit dem Unterrichtsgegenstand im Sinne einer aktiven Lernzeitnutzung als ein wichtiger Prädiktor für die Leistungen im dritten und fünften Schuljahr. Für das erste Schuljahr konnte dieser positive Effekt nicht bestätigt werden. Der Faktor ‚Instructional Support‘ hatte hingegen in keinem Schuljahr einen signifikanten Effekt auf die Leistungen im Bereich Mathematik.

Eine Studie aus Portugal bestätigt für das erste Schuljahr im Rahmen von Mehrebenenanalysen und nach Kontrolle weiterer Merkmale einen positiven Effekt der ‚Classroom Organization‘<sup>53</sup> für die Bereiche des Wortschatzes (Vocabulary) und des Textverständnisses (Print Concept). Zwar konnten die Autoren keinen Haupteffekt der ‚Classroom Organization‘ im Bereich Zahlenlesen (Number Identification) bestätigen, es zeigte sich jedoch, dass in diesem Bereich vor allem schwächere Schüler von einer effektiven Klassenführung profitieren. Neben dem Zahlenlesen wurden im Rahmen der Studie von Cadima et al. (2010) weitere mathematische Fähigkeiten aus unterschiedlichen Gebieten erhoben: Rechenfertigkeit (Numbering Skills), Zählfertigkeit (Calculation Skills) und Konzeptverständnis / Wissen (Understanding of Concepts) (Cadima et al., 2010) und anschließend zu einer Skala ‚Global Math‘ zusammengefasst. Allerdings zeigte sich bereits auf korrelativer Ebene kein Zusammenhang der allgemeinen Unterrichtsqualität mit der Skala ‚Global Math‘. Aus diesem Grund und weil sich auch im Rahmen der sich anschließenden Mehrebenenregressionsanalysen kein Effekt auf das ‚Global Math‘ zeigte, er-

52 Der Faktor ‚Emotional Support‘ beinhaltet die Dimensionen ‚Overcontrol-Reflected‘, ‚Chaos-Reflected‘, ‚Positive Emotional Climate‘, ‚Negative Emotional Climate-Reflected‘ und ‚Sensitivity‘ (vgl. Kapitel 5.2.1, Tabelle 9). Der Faktor ‚Instructional Support‘ setzt sich aus den Dimensionen ‚Productive Use of Time‘ und ‚Richness of Instructional Methods‘ zusammen (vgl. Tabelle 8). Inwieweit sich der Faktor ‚Emotional Quality‘ als wichtiger Prädiktor für die Leistungen in den beiden Bereichen erwiesen hat, wird im Kapitel 5.2 ausführlich dargestellt, wenn es um die Bedeutung des Unterrichtsklimas für die Leistungsentwicklung der Schüler geht.

53 Der Faktor ‚Classroom Organization‘ beinhaltet die Aspekte ‚Behavior Management‘, ‚Productivity‘ und ‚Instructional Learning Formats‘ (vgl. CLASS-Beobachtungssystem, Kapitel 2.1, Tabelle 1).

folgte keine differenzierte Analyse für die drei Subdimensionen des CLASS (Instructional Support, Emotional Support und Classroom Organization).

### 5.1.2 *Klassenführung und Selbstkonzept*

Mit Blick auf die Multikriterialität von Unterricht können nicht nur Zusammenhänge zwischen einer effektiven Klassenführung und kognitiven, sondern auch mit affektiv-motivationalen Schülermerkmalen bestätigt werden (vgl. Seidel & Shavelson, 2007). Im Rahmen der vorliegenden Arbeit liegt der Fokus auf den fachspezifischen Selbstkonzepten im Bereich Mathematik und Rechtschreiben. Anschließend werden ausgewählte Studienergebnisse dargestellt, die die Vorhersage dieser Variablen durch Unterrichtsqualitätsmerkmale, wie das der Klassenführung, aber auch der Strukturiertheit des Unterrichts, untersuchten. Aufgrund mangelnder Befunde zum Zusammenhang einer effektiven Klassenführung mit fachspezifischen Selbstkonzepten in der Grundschule, werden zusätzlich Ergebnisse aus der Sekundarstufe herangezogen (z. B. Gruehn, 2000; Helmke, 1992; Helmke & Schrader, 1990).

Dass eine intensive Zeitnutzung als Merkmal der Klassenführung und somit eine ausgeprägte Lehrstofforientierung zu einem günstigen Selbstkonzept beiträgt, konnte Helmke (1992) im Rahmen der Münchner Hauptstudie für die Sekundarstufe zeigen. Auch Helmke und Schrader (1990) bestätigen im Kontext ihrer Studie den positiven Zusammenhang zwischen dem ‚Classroom Management‘ und dem Selbstkonzept in Mathematik. Allerdings erweist sich dieser als eher gering. Vor allem die Effektivität von Regeln (hoch inferent erfasst) und die aus Schülersicht wahrgenommene Unterrichtsführung hängen mit einem positiven mathematischen Selbstkonzept zusammen (Helmke et al., 1986). Neben direkten Effekten konnten im Rahmen der Studie von Helmke et al. (1986) auch indirekte Wirkungen über das Schülerengagement bestätigt werden. Je höher die Schüler die Effektivität der Klassenführung wahrnehmen, desto mehr engagieren sie sich im Unterricht, was wiederum positive Auswirkungen auf die affektiven Lernbereiche (Einstellungen gegenüber Mathematik und mathematisches Selbstkonzept) der Schüler zur Folge hat. Das Klassenmanagement wurde in dieser Studie stark in Anlehnung an das Konzept der direkten Instruktion erfasst. Ein hoher Leistungsdruck und ein hohes Interaktionstempo im Unterricht haben in der Sekundarstufe hingegen einen negativen Effekt auf das Selbstkonzept in Mathematik (vgl. Gruehn, 1995).

Für die Grundschule mangelt es bislang an Studien, in denen die Bedeutung einer effektiven Klassenführung für die Entwicklung der fachspezifischen Selbstkonzepte untersucht wurde. In Anlehnung an die Ergebnisse aus der Sekundarstufe kann jedoch auch für die Grundschule angenommen werden, dass verschiedene Merkmale der Klassenführung positive Effekte auf die fachspezifischen Selbstkonzepte haben. Ein gewisses Maß an Regellosig-

keit bzw. Chaos in der Klasse kann zu Verunsicherungen der noch jungen Schüler führen und über die Wahrnehmung der Schüler negative Auswirkungen auf die Entwicklung der Selbstkonzepte haben.

Im Rahmen der KILIA-Studie konnte bislang lediglich gezeigt werden, dass die Effektivität der Klassenführung keinen Effekt auf den Zusammenhang zwischen mathematischem Selbstkonzept und Mathematikleistung hat, unabhängig davon, ob es sich um Extremklassen mit fallendem oder steigendem Zusammenhang<sup>54</sup> handelt (vgl. Kammermeyer & Martschinke, 2003). Die Autorinnen vermuten jedoch, dass dies vor allem an der Erfassung des Merkmals selbst liegt. Denn die Grundlage bildeten nicht nur Stunden, die sich hauptsächlich durch direkte Instruktion auszeichneten, sondern die Daten beruhen auf 30 Unterrichtsstunden, in denen häufig auch offene Unterrichtsphasen stattfanden. Ein Merkmal solcher offenen Unterrichtsphasen ist unter anderem, dass sich die Lehrperson in ihrer direkt lenkenden Funktion zurücknimmt „und eher indirekt durch die Auswahl des Materials bzw. die Gestaltung der Lernumgebung wirkt“ (Kammermeyer & Martschinke, 2003, S. 498).

## 5.2 Unterrichtsklima und multikriteriale Effekte

Neben der Effektivität der Klassenführung wird in der vorliegenden Arbeit auch der Effekt eines positiven Unterrichtsklimas auf die Leistungen und Selbstkonzepte der Schüler untersucht. Im Rahmen einer Metaanalyse von 119 Studien, durchgeführt von Cornelius-White (2007), können insgesamt positive Zusammenhänge des Unterrichtsklimas bzw. der Schülerorientierung mit kognitiven und affektiv-behavioralen Schülermerkmalen bestätigt werden, wobei die Höhe der Zusammenhänge zwischen dem schülerorientierten Unterricht und den Merkmalen im kognitiven Bereich stärker variieren als im affektiv-behavioralen. Höhere Zusammenhänge zeigten sich vor allem für die Qualität der Lehrer-Schüler-Beziehung im Sinne eines schülerorientierten Lehrerverhaltens mit den kognitiven Fähigkeiten sowie mit dem kritischen Denken in Deutsch und Mathematik.

---

54 Im Kontext der KILIA-Studie variiert der Zusammenhang zwischen den mathematischen Leistungen und dem Selbstkonzept zu Beginn des ersten Schuljahres in allen teilnehmenden Klassen zwischen  $r = -.55$  und  $r = .88$ , am Ende des ersten Schuljahres zwischen  $r = -.39$  und  $r = .76$  (vgl. Kammermeyer & Martschinke (2003). Um die Bedeutung der erfassten Lehrervariablen für die Leistungs- und Selbstkonzeptentwicklung im Anfangsunterricht zu untersuchen, wurde für zehn Klassen die Veränderung des Zusammenhangs zwischen Schulanfang und Ende des ersten Schuljahres herangezogen. Es wurden die „fünf Klassen mit dem am meisten steigenden und fünf Klassen mit dem am meisten sinkenden Zusammenhang gegenübergestellt“ (ebd., S. 496).

Im folgenden Kapitel 5.2.1 werden ausgewählte Studien vorgestellt, anhand derer deutlich wird, dass die Bedeutung des Unterrichtsklimas für die Entwicklung kognitiver Schülermerkmale unterschiedlich bewertet wird. In Kapitel 5.2.2 werden ausgewählte Studien zur Bedeutung des Unterrichtsklimas für die Entwicklung fachspezifischer Selbstkonzepte im Grundschulalter präsentiert.

### 5.2.1 *Unterrichtsklima und Leistungen*

Im Gegensatz zu einer effektiven Klassenführung, für die übereinstimmend positive Zusammenhänge mit kognitiven Schülermerkmalen bestätigt werden können, wird die Bedeutung eines positiven Unterrichtsklimas für die Entwicklung kognitiver Schülermerkmale unterschiedlich bewertet. So verdeutlicht Brophy (2000) die Wichtigkeit des Unterrichtsmerkmals unterstützendes Unterrichtsklima für die individuellen Leistungen der Schüler, indem er es an die Spitze seiner Auflistung setzt. Allerdings gelangt Brophy (2000) auf Basis einer ausführlichen Literaturzusammenfassung zu dieser Platzierung. Im Gegensatz zu Brophy (2000) schreiben Wang, Haertel und Walberg (1993b) im Rahmen ihrer Metaanalyse dem Klassenklima eine eher marginale Rolle zur Erklärung des Lernerfolgs zu.

Betrachtet man Studien aus dem angloamerikanischen Bereich, dann weisen diese einen konsistent positiven, jedoch auch sehr geringen Zusammenhang zwischen Klima und Leistungen in der Sekundarstufe auf (zusf. Gruehn, 2000, S. 79f). Dieser Effekt verschwindet allerdings, sobald zusätzlich Merkmale auf Unterrichtsebene untersucht werden. Für Deutschland liegen bislang eher uneinheitliche Befunde vor. So belegen ältere Studien die prädiktive Kraft des Klimas für die Leistungen der Schüler. Beispielsweise fand Dreesmann (1978) auf Klassenebene bei den Rechtschreibleistungen eine Varianzaufklärung von 13 Prozent für die Merkmale des Unterrichtsklimas (zit. n. Dreesmann, 1982). Aktuelle Studien im Fach Deutsch kommen hingegen zu anderen Ergebnissen. So konnte im Kontext der DESI-Studie kein Einfluss des aus Schülersicht berichteten Unterrichtsklimas oder der Unterstützung der Lehrperson auf den Lernzuwachs in den Bereichen Lesekompetenz und Sprachbewusstheit bestätigt werden. Lediglich die aus Schülersicht wahrgenommene Schülerorientierung ging mit einem geringen Zuwachs in dem Bereich Sprachkompetenz einher (vgl. Klieme et al., 2008). Für den Bereich Mathematik konnten Studien für die Sekundarstufe hingegen keine Zusammenhänge des Unterrichtsklimas mit den Schülerleistungen bestätigen (vgl. Helmke & Schrader, 1990), auch nicht nach Kontrolle weiterer individueller und klassenbezogener Variablen (vgl. Gruehn, 2000; Lipowsky et al., 2009).

Im Vergleich zur Sekundarstufe wird der positiven Lehrer-Schüler-Beziehung in der Grundschule theoretisch ein größerer Zusammenhang mit Schülermerkmalen attestiert (vgl. Kapitel 4.2), da junge Schüler durch ihre Bezie-

hungen mit Erwachsenen noch stärker beeinflusst werden als ältere Schüler (vgl. Roorda et al., 2011). Ein Großteil der Studien fokussiert dabei eher den Einfluss auf das Verhalten und Engagement der Schüler (vgl. Kapitel 4.2, z. B. Hamre & Pianta, 2001). Auch der Zusammenhang mit den Leistungen der Schüler wurde bereits mehrfach untersucht (z. B. Ang, 2005; Davis, 2001; Helmke & Weinert, 1997b; Roorda et al., 2011). Allerdings ist auch hier die Befundlage nicht eindeutig (vgl. Tabelle 9). So bestätigt Ang (2005) für das Ende der Grundschulzeit (4. bis 6. Schuljahr, Singapur), dass in Klassen, in denen sich die Schüler trauen, die Lehrperson um Hilfe zu bitten bzw. kaum Konflikte zwischen der Lehrperson und den Schülern vorhanden sind, Schüler bessere Leistungen am Ende des Schuljahres haben. Als Index für die Leistungen der Schüler wurde jedoch ein Gesamtwert für die Fächer Englisch, Mathematik und Physik gebildet, womit fachspezifische Unterschiede in der Bedeutung der Lehrer-Schüler-Beziehung für die Leistungen nicht untersucht werden konnten.

Für den Bereich Mathematik belegen Tiedemann und Billmann-Mahecha (2004) für das dritte Schuljahr einen signifikanten Effekt des Klassenklimas (aggregierte Schülereinschätzungen) auf die kognitiven Fähigkeiten in Mathematik (Textaufgaben und Arithmetik). Obwohl sich dieser als eher gering erweist, bedeutet dies, dass Klassen, in denen die Schüler ein günstiges Klima wahrnehmen, im Schnitt höhere mittlere Leistungen im Fach Mathematik haben. Die Analyse auf Individualebene konnte diesen Effekt nicht bestätigen.

Mit der Studie von Dunn und Harris (1998) können für das vierte Schuljahr Zusammenhänge des durch die Schüler wahrgenommenen Klimas mit den Leistungen in den Fächern Lesen, Mathematik und Sprache bestätigt werden. In der Studie wurde das Klima einer Klasse mithilfe des MCI (My Class Inventory, vgl. Fraser, Anderson & Walberg, 1982) aus Sicht der Schüler erhoben. Es beinhaltet die fünf Subdimensionen Zufriedenheit (Satisfaction), Spannungen / Unstimmigkeiten (Friction), Wettbewerb (Competition), Gruppenzusammenhalt (Cohensiveness) und Schwierigkeiten mit der Arbeit in der Klasse (Difficulty). Auf Basis von multiplen Regressionsanalysen erwies sich in dieser Studie vor allem das Merkmal ‚Difficulty‘ als konstanter Prädiktor für die Vorhersage individueller Leistungen in allen drei Bereichen. Das bedeutet, dass die vermehrte Wahrnehmung von Schwierigkeiten in der Klasse schlechtere Leistungen der Schüler in allen drei Bereichen zur Folge hat. Die Leistungen in Mathematik konnten zusätzlich durch das Ausmaß an wahrgenommenem Wettbewerb in der Klasse und die wahrgenommene Spannung / Unstimmigkeit innerhalb der Klasse erklärt werden. Schüler, die in ihrer Klasse mehr Wettbewerb und Spannungen bzw. Unstimmigkeiten wahrnahmen, zeigten im vierten Schuljahr tendenziell schlechtere Leistungen in Mathematik. Allerdings sind in der Studie von Dunn und Harris (1998) die aufgeklärten Varianzanteile sehr gering, womit – auch unter Bezug auf die Ergebnisse von Tiedemann und Billmann-Mahecha (2004) – für das Ende der Grundschulzeit eher davon ausgegangen werden kann, dass das Klima

aus Schülersicht eine geringe prädiktive Vorhersagekraft für die Leistungen hat. Zu ähnlichen Vermutungen kommt Davis (2001) auch im Rahmen ihrer Studie. Sie konnte für das dritte und vierte Schuljahr keinen Zusammenhang zwischen einer positiven Lehrer-Schüler-Beziehung (erfasst über Aspekte des Vertrauens, der Unterstützung und der Identifikation mit der Lehrperson) und der mittleren Klassenleistung (Notendurchschnitt) bestätigen. Die Autorin sieht einen möglichen Grund für den nicht vorhandenen Zusammenhang in der Erfassung der Lehrer-Schüler-Beziehung durch die Befragung der noch jungen Schüler. Sie vermutet, dass die Grundschüler nicht so kritisch sind wie ältere Schüler, wenn diese ihre Beziehung zur Lehrperson einschätzen.

Im Gegensatz zu Studien, die das Klima aus Schülersicht erfassen, können für das Ende der Grundschulzeit im Rahmen von Beobachtungsstudien keine Zusammenhänge zwischen dem beobachteten Klima und den Leistungen der Schüler bestätigt werden. Dies gilt sowohl für den Bereich Mathematik als auch für den Bereich Rechtschreiben (vgl. Tabelle 9). So zeigt sich in der SCHOLASTIK-Studie kein signifikanter Zusammenhang des sozialen Klimas mit dem mittleren Leistungszuwachs der Klasse in Mathematik (3./4. Schuljahr, vgl. Helmke & Weinert, 1997b). Auch im Rechtschreiben besitzt das Klassenklima in dieser Studie keine prädiktive Vorhersagekraft. Zu ähnlichen Ergebnissen für den Bereich Rechtschreiben kommen Roos und Schöler (2009; vgl. auch Hofmann, 2007) im Rahmen der EVES-Studie. Das beobachtete Arbeitsklima besitzt in ihrer Studie keine prädiktive Vorhersagekraft für die Rechtschreibleistungen der Schüler am Ende des vierten Schuljahres.

Für den Anfangsunterricht mangelt es bislang an Studien, die die Bedeutung des Unterrichtsklimas für die Leistung der Schüler untersuchen. Die wenigen Studien zeigen jedoch, dass, obwohl theoretisch angenommen, das Unterrichtsklima auch im Anfangsunterricht nur selten einen Effekt auf die Leistungen der Schüler hat (siehe Tabelle 9; vgl. Cadima et al., 2010; May, 2001; Perry et al., 2007; Roßbach, 2002b). So erwies sich in der PLUS-Studie das soziale Klima lediglich als ein wichtiger Prädiktor für den Lernzuwachs im Textschreiben im zweiten Schuljahr. Im Rechtschreiben konnte kein Zusammenhang mit Lernzuwachs bestätigt werden (vgl. May, 2001). Auch in der ECCE-Studie ergab sich, nach Kontrolle zusätzlicher Strukturmerkmale, kein Zusammenhang des schülerorientierten sozial-emotionalen Klimas mit den Leistungen der Schüler im zweiten Schuljahr (vgl. Roßbach, 2002b).

Für den Bereich Mathematik ist die Befundlage erneut sehr uneinheitlich. So konnten Cadima et al. (2010) für das erste Schuljahr keinen Effekt der emotionalen Unterstützung (Emotional Support) auf die Leistungen der Schüler bestätigen, während sich in der Studie von Perry et al. (2007) der ‚Emotional Support‘ der Lehrperson als positiver Prädiktor für die Leistungen der Schüler in Mathematik am Ende des ersten Schuljahres erwies. In der Studie von Pianta et al. (2008) ergaben sich für das Fach Mathematik im ersten und dritten Schuljahr tendenziell sogar negative Effekte des ‚Emotional Supports‘ auf die Leistungen der Schüler. Die Autoren interpretieren ihre Ergebnisse

dahingehend, dass es im Fach Mathematik eher um das Lehren und Einüben technischer Fertigkeiten geht und der Fokus weniger auf der Lehrer-Schüler-Interaktion liegt.

Zusammenfassend zeigt sich auch hier – ähnlich wie bei der Bedeutung einer effektiven Klassenführung für die Leistungen der Schüler (vgl. Kapitel 5.1.1) – ein uneinheitlicher Forschungsstand. Die kontroverse Befundlage lässt sich jedoch vermutlich am ehesten auf die in den Studien verwendeten unterschiedlichen Operationalisierungen und Analysemethoden zurückführen (vgl. Tabelle 9).

### 5.2.2 *Unterrichtsklima und Selbstkonzept*

Weniger uneinheitliche Ergebnisse liefern Studien, die sich mit dem Unterrichtsklima als Prädiktor für affektiv-motivationale Merkmale beschäftigen. Hier können deutlich positive Zusammenhänge für die Sekundarstufe bestätigt werden. So untersuchte Pekrun (1985) beispielsweise im Rahmen einer Querschnittsuntersuchung an hessischen Förderschulen den Zusammenhang zwischen der Selbstkonzeptausprägung und den Klimamerkmale Unterrichtsanstrengung der Lehrperson, Unterstützung, Wettbewerb, Kooperation und Affiliation in der Klasse. Zusätzlich unterschied er zwischen dem schulischen und allgemeinen Selbstkonzept, dem sozialen Selbstkonzept gegenüber den Lehrpersonen und den Schülern sowie dem allgemeinen Selbstwertgefühl.<sup>55</sup> Vor allem das schulische Selbstkonzept korrelierte zwischen  $r = .14$  und  $r = .33$  mit den entsprechenden Klimamerkmale. Die Korrelationen mit dem allgemeinen Selbstkonzept und dem Selbstwertgefühl bewegen sich in ähnlicher Höhe. Eder (1996) konnte über einen Zeitraum von zwei Jahren Effekte des wahrgenommenen Klimas auf verschiedene psychosoziale Merkmale, u. a. Zufriedenheit, Interesse und Selbstkonzept bestätigen. Des Weiteren beteiligen sich Schüler mit positiven Klimaerfahrungen nach Eder (2001) „mehr am Unterricht, stören weniger und schaffen sich damit bessere Lern- und Leistungsvoraussetzungen. Positive Klimaerfahrungen sind im Allgemeinen mit einem geringeren Ausmaß an schulischen Belastungen (Schul- angst, Stress, psychovegetative Belastungen) und einem tendenziell positiven (Leistungs-)Selbstkonzept und Selbstwertgefühl verknüpft“ (S. 582).

Im Zusammenhang mit den Wirkungen des Klassenklimas existieren jedoch auch Studien in der Sekundarstufe, die einen negativen bzw. keinen Zusammenhang zwischen Selbstkonzept und Klimavariablen zeigen können (vgl. Dreesmann, 1982; Helmke & Schrader, 1990; Kuliga, 1985; Petillon, 1997). So verweist Dreesmann (1982) auf seine eigene Studie, in der er für den

55 Die Begriffe Selbstkonzept und Selbstwert werden häufig synonym verwendet, obwohl sie unterschiedliche Bedeutungen haben. Manchmal wird der Selbstwert jedoch als die bewertende Facette des Selbstkonzepts angesehen (vgl. Woolfolk & Schönplflug, 2008). Eine klare Trennung fällt jedoch schwer.

Mathematikunterricht im achten Schuljahr (Hauptschule) unter anderem zu dem Ergebnis kam, dass mit dem Erleben eines schwierigen Unterrichts tendenziell eine schlechtere Selbstbewertung der Schüler einhergeht. Die erlebte Disziplin im Unterricht sowie das Wahrnehmen einer Erfolgsaussicht auf Anstrengung und die wahrgenommene Verständlichkeit des Unterrichts hängen wiederum mit einem positiveren Selbstkonzept der Schüler zusammen (vgl. Dreesmann, 1982). Die affektive Qualität der Lehrer-Schüler-Beziehung korreliert hingegen in diesem Alter nicht mit der Selbstkonzeptentwicklung in Mathematik (vgl. Helmke & Schrader, 1990). Auch durch einen hohen Leistungs- und Anpassungsdruck kann das Selbstbewusstsein der Schüler absinken (vgl. Eder, 1996; Fend, 1977).

Im Gegensatz zu den Befunden aus der Sekundarstufe kann für die frühen Grundschuljahre davon ausgegangen werden, dass das Unterrichtsklima Auswirkungen auf das Selbstkonzept der Schüler hat, denn zu Beginn der Schulzeit werden die entscheidenden Weichen für die Entwicklung der verschiedenen Selbstkonzepte und der sie begleitenden Selbstwertgefühle gestellt (vgl. Einsiedler, 1988). Dabei spielen die Wertschätzung durch die Lehrperson, die soziale Integration in die Klasse und damit auch das Wohlbefinden eine große Rolle für die Lernbereitschaft, für das soziale Miteinander und für die Verarbeitung von Erfolgen und Misserfolgen (vgl. Hascher, 2004). Ein negatives Klassenklima gefährdet vermutlich vor allem in der Grundschule – im Vergleich zu späteren Schuljahren – das Selbstkonzept (vgl. Jerusalem & Schwarzer, 1991). So zeigen in der Studie von Jerusalem und Schwarzer (1991) besonders jüngere Schüler in klimanegativen Klassen längerfristig weniger Selbstwertgefühl als Schüler in klimapositiven Klassen. Die Autoren werteten dies als einen Hinweis darauf, „daß jüngere Schüler gegenüber Risikoumwelten noch verwundbarer sind bzw. daß ihr Selbstkonzept noch nicht so stabil ist wie bei den älteren“ (S. 124). Dies bedeutet, dass das Selbstkonzept in der Grundschule stärker als im weiteren Verlauf der Schulzeit durch ein negatives Klima gefährdet sein kann. Nach Jerusalem (1993) ergibt sich quasi „ein klimaspezifischer Schereneffekt“ (S. 10). Allerdings verglichen Jerusalem und Schwarzer (1991) Schüler im Alter zwischen zehn und zwölf Jahren (= jüngere Schüler) mit Schülern im Alter zwischen 14 und 16 Jahren (= ältere Schüler).

Tabelle 9: Bedeutung des Unterrichtsklimas für die Leistungen in der Grundschule – Eine ...

Studie u. Autoren	Schuljahr	Leistungsbereich	Datenquelle z. Erfassung des Klimas
<i>Deutschsprachige Studien ...</i>			
SCHOLASTIK-Studie: Helmke & Weinert (1997b)	3. SJ 4. SJ	Leistungszuwachs in: (1) Mathematik (2) Rechtschreiben	Beobachter
Hannoversche Grundschul- studie: Tiedemann & Billmann-Mahecha (2004)	3. SJ	Leistungen in: (1) Mathematik (Textauf- gaben und Arithmetik) (2) Rechtschreiben	Schüler (Individualwert & Klassenmittel- wert)
EVES-Studie: Roos & Schöler (2009; vgl. auch Hofmann, 2007)	Ende 4. SJ	Leistungsniveau im Rechtschreiben	Beobachter
PLUS-Studie: May (2001)	2. SJ	Lernzuwachs im: (1) Textschreiben (2) Rechtschreiben	Beobachter
ECCE-Studie: Roßbach (2002a; b)	2. SJ	Leistungsniveau (Gesamtskala: Satzver- ständnis, Rechnen, Lösen angewandter Probleme, Kenntnisse über die natür- liche und soziale Umwelt)	Beobachter
– Fortsetzung auf den Seiten 108/109 –			

... Gegenüberstellung ausgewählter Studien

Operationalisierung des Klimas	Analyse- methode	Ergebnisse*
<i>... Deutschsprachige Studien</i>		
Soziales Klima: (1) Akzeptanz und ausdrückliche Thematisierung des affektiven Erlebens der Schüler (2) Bedeutung des Lehrers als persönlicher Ansprechpartner (3) Vertrauensperson über die Rolle des Stoffmittlers hinaus	Korrelation	(1) / (2) Kein Zusammenhang zwischen sozialem Klima und Leistungszuwachs
Klassenklima: (1) Sozial- und Leistungsdruck (2) Schülerzentriertheit (3) Kohäsion (4) Disziplin	Mehr ebenen analyse	Unterschiedliche Effekte des Klassenklimas je nach Erfassungsebene und Leistungsbereich: (a) Kontextprädiktor Klassenklima (aggregierter Mittelwert): (1) Positiver Effekt (2) Kein Effekt (b) Individualprädiktor Klassenklima: (1) / (2) Kein Effekt
Arbeitsklima: (1) Klima (leistungszentriertes Unterrichtsklima) (2) Motivation (die Lehrperson motiviert die Klasse) (3) Schülerengagement (interessierte Mitarbeit der Schüler/-innen) (4) Reaktion auf Misserfolge (keine Herabsetzungen und Bloßstellungen) (5) Schülerbeiträge (Lehrperson nimmt Schülerbeiträge auf) (6) Beziehungsaspekt (die Lehrperson zeigt Wärme und Humor)	Korrelation	Kein Zusammenhang zwischen Arbeitsklima und Leistungsniveau
Soziales Klima: (1) Gegenseitige Unterstützung der Kinder (2) Sozialverhalten der Kinder untereinander (3) Häufigkeit von beobachtbaren Aggressionen (4) Wärme der Atmosphäre (u. a. Umgang mit Regelverletzungen)	Korrelation	(1) Zusammenhang des sozialen Klimas mit Lernzuwachs (2) Kein Zusammenhang des sozialen Klimas mit Lernzuwachs
Schülerorientiertes sozial-emotionales Klima: (1) Anerkennung der Schüler als Personen (2) Raum für ihre privaten Gefühle und Probleme (3) Orientierung des Unterrichts nicht nur auf fachliche Ziele (4) Ablehnung eines leistungsorientierten und kontrollierenden Klimas	Korrelation / Partialkorrelation	Kein Zusammenhang des schülerorientierten sozial-emotionalen Klimas mit Leistungsniveau
* Die Nummern in Klammern beziehen sich auf die Spalte Leistungsbereich.		

Studie u. Autoren	Schuljahr	Leistungsbereich	Datenquelle z. Erfassung des Klimas
<i>Internationale Studien ...</i>			
Ang (2005)	4.–6. SJ	Leistungen in: Englisch, Mathematik und Physik (Mittelwert)	Lehrer
Dunn & Harris (1998)	4. SJ	Leistungen in: (1) Mathematik (2) Sprache (3) Lesen	Schüler
Davis (2001)	3. SJ 4. SJ	Notendurchschnitt (über alle Fächer)	Schüler (Klassenmittelwert)
Cadima et al. (2010)	1. SJ	Leistungen in: (1) Wortschatz (Vocabulary) (2) Textverständnis (Print Concepts) (3) Zahlenlesen (Number Identification)	Beobachter
Pianta et al. (2008)	1. SJ 3. SJ 5. SJ	Leistungen in: Mathematik (und Lesen**)	Beobachter
Perry et al. (2007)	Ende 1. SJ	Leistungen in: Mathematik (und Lesen**)	Beobachter

*Anmerkungen:* Erfassung des Klassenklimas für das 3. SJ i. d. Hann. Grundschulstudie mithilfe einer adaptierten Version des Linzer Fragebogens zum Schul- und Klassenklima (LFSK 4–8, vgl. Eder & Mayr, 2000).

Erfassung des Arbeitsklimas i. d. Studie von Roos und Schöler (2009; vgl. auch Hofmann, 2007) in Anlehnung an das Beobachtungssystem der SCHOLASTIK-Studie. Allerdings sind die Konstrukte teilweise nicht miteinander vergleichbar.

Erfassung des Klimas aus Lehrersicht in der Studie von Ang mithilfe des TSR1 (Teacher-Student Relationship Inventory, vgl. Ang, 2005).

Erfassung des Klassenklimas aus Schülersicht in der Studie von Dunn und Harris mithilfe des MCI (My Class Inventory, vgl. Fraser, Anderson & Walberg, 1982).

Operationalisierung des Klimas	Analyse- methode	Ergebnisse*
<i>... Internationale Studien</i>		
Schüler-Lehrer-Beziehung (Student-Teacher Relationship): (1) Zufriedenheit (Satisfaction) (2) Konflikte (Conflict) (3) Lernhilfe (Instructional Help)	Multiple Re- gressions- analysen	Positiver Effekt d. Zufrie- denheit (Satisfaction) u. in- struktorialen Unterstützung (Instructional Help) Negativer Effekt des Merk- mals Konflikte (Conflict)
Climate: (1) Zufriedenheit (Satisfaction) (2) Spannungen / Unstimmigkeiten (Friction) (3) Wettbewerb (Competition) (4) Gruppenzusammenhalt (Cohensiveness) (5) Schwierigkeiten mit der Arbeit in der Klasse (Difficulty)	Multiple Re- gressions- analysen	(1) / (2) / (3) positiver Effekt des Merkmals Schwierig- keiten mit der Arbeit in der Klasse (Difficulty)
Schüler-Lehrer-Beziehung (Student/Teacher Re- lationship): (1) Vertrauen gegenüber der Lehrperson (LP) (Quality of Affect) (2) Unterstützung durch die LP (Quality of Utilization) (3) Identifikation mit der LP (Quality of Emulation)	Korrelation	Kein Zusammenhang der Qualität der Student/ Teacher Relationship (Schüler-Lehrer-Beziehung) mit Noten
Emotional Support: (1) Positives Klima (Positive Climate) (2) Negatives Klima (Negative Climate (-)) (3) Sensibilität der LP (Teacher Sensitivity) (4) Berücksichtigung der Perspektive der Schüler (Regard for Student Perspectives)	Mehrebe- nenanalyse	(1) / (2) / (3) Kein Effekt des Emotional Supports auf die Leistungen
Emotional Support: (1) Kontrolle (Overcontrol (-)) (2) Chaos (-) (3) Positives emotionales Klima (Positive Emotional Climate) (4) Negatives emotionales Klima (Negative Emotional Climate (-)) (5) Sensibilität (Sensitivity)	Change Growth Mixture Mo- dels (GMM)	1./3. SJ: Tendenziell nega- tiver Effekt des Emotional Supports auf Leistungen (jedoch nicht signifikant) 5. SJ: Positiver Effekt des Emotional Supports auf Leistungen
Emotional Support: (1) Wärme der Lehrperson / Ansprechbarkeit (Teacher Warmth / Responsiveness) (2) Unterstützung sozialer Kompetenzen (Support for Interpersonal Skills) (3) Unterstützung der Kommunikationsfä- higkeit (Support for Communication Skills) (4) Einbezug der Schülererfahrungen (Re- levance of Activities to Children's Experience) (5) Engagement der Schüler (Student Engagement) (6) Individualisierung der Lernaktivitäten (Individualization of Learning Activities)	Mehr- ebenen- analyse	Positiver Effekt des Emotio- nal Supports

\*\* Wird in vorliegender Arbeit nicht untersucht, daher hierfür keine Ergebnisse.

Inwieweit das Unterrichtsklima im Anfangsunterricht eine Bedeutung für die Selbstkonzeptentwicklung hat, soll im Rahmen der vorliegenden Studie überprüft werden. Die Untersuchungsergebnisse von Kammermeyer und Martschinke aus dem Jahr 2004 geben bislang lediglich einen Hinweis darauf, dass mit steigendem Zusammenhang zwischen Leistungen und Selbstkonzept im ersten Schuljahr Lehrpersonen grundsätzlich weniger Freiheitsspielräume einräumen als Lehrpersonen in Klassen mit einem sinkenden Zusammenhang. Die Autorinnen schlussfolgern daraus, dass in Klassen, in denen die Lehrpersonen ihren Schülern viele Freiheitsspielräume einräumen, der soziale Vergleich erschwert wird, der wiederum Ausgangspunkt für die Ausbildung der Selbstkonzepte darstellt. Andererseits konnten die Autorinnen zeigen, dass in Klassen mit einem sinkenden Zusammenhang zwischen den Leistungen und Selbstkonzepten im ersten Schuljahr sich die Lehrpersonen tendenziell mehr am individuellen Fortschritt der Schüler orientieren. Auch die Betonung des individuellen Fortschritts wirkt dem sozialen Vergleich entgegen. Allerdings merken Kammermeyer und Martschinke selbst an, dass die Ergebnisse mit Vorsicht interpretiert werden sollten, da sie ausgehend von der kleinen Stichprobe nicht auf Basis „von komplexen Mehrebenenanalysen, sondern aufgrund von einfachen Extremgruppenvergleichen zustande gekommen sind“ (Kammermeyer & Martschinke, 2004, S. 214). Durch den Extremgruppenvergleich konnten jedoch keine signifikanten unterrichtlichen Einflussfaktoren identifiziert werden, die bei der Selbstkonzeptentwicklung im Laufe des ersten Schuljahres eine Rolle spielen.

Auch aus der Studie von Perry et al. (2007) können keine Rückschlüsse über die Bedeutung des Unterrichtsklimas für fachspezifische Selbstkonzepte gezogen werden, da in ihrer Studie keine domänenspezifische Erfassung der Fähigkeitsselbsteinschätzungen erfolgte. Den Schülern wurden während eines Einzelinterviews lediglich vier Items vorgelegt („I am good at schoolwork“, „I am good at reading“, „I am good at math“, „I am good at artwork“), die anschließend zu einer Gesamtskala zusammengefasst wurden, die eher ein allgemeines schulisches Selbstkonzept darstellt. Das Antwortformat war vierstufig („almost never“, „sometimes“, „a lot“, „almost always“). Perry et al. (2007) bestätigen jedoch, dass Schüler des ersten Schuljahres ihre allgemeinen akademischen Fähigkeiten im Mittel höher einschätzen, wenn sie in Klassen unterrichtet werden, in denen mehr unterstützende Maßnahmen durch die Lehrperson beobachtet werden können, auch nach Kontrolle früherer tatsächlicher Leistungen. Zudem zeigen Schüler in Klassen mit mehr unterstützenden Maßnahmen ein positiveres Sozialverhalten.

## 6 Die Erfassung von Unterrichtsmerkmalen durch Videoanalyse

Im Rahmen der videobasierten Unterrichtsforschung stellt die Entwicklung von Beobachtungsverfahren für die Analyse von Unterricht bzw. die Qualität von Lehr-Lern-Prozessen eine anspruchsvolle Aufgabe dar. Der aktuelle Kenntnisstand zur Analyse von Lehr-Lern-Prozessen im videografierten Unterricht baut hauptsächlich auf Erfahrungen aus der Sekundarstufe auf (vgl. Kapitel 2.2). Bislang liegen kaum fundierte Erfahrungen im Umgang mit Videoaufzeichnungen in der Grundschule vor. Aus diesem Grund werden im Folgenden die Herausforderungen bei der Videoanalyse von Unterrichtsprozessen in der Grundschule näher betrachtet.

### 6.1 Videoanalysen und ihre Herausforderungen in der Grundschule

Als bedeutsamer Unterschied im Vergleich zu Sekundarschulen, muss in der Grundschule bei dem Einsatz von Unterrichtsvideos berücksichtigt werden, dass sich Grundschulunterricht in stärkerem Maße als in anderen Schulstufen durch Methodenvielfalt auszeichnet (vgl. Götz et al., 2005), während sich der Unterricht in der Sekundarstufe vermehrt auf lehrergesteuerten Unterricht im Sinne direkter Instruktion<sup>56</sup> (Frontalunterricht) beschränkt. Der frontale, lehrerzentrierte Unterricht<sup>57</sup> erfolgt in der Grundschule integriert in den offenen Unterricht. Er bietet den Schülern einerseits Entlastung nach den Phasen der Selbstständigkeit, erfüllt andererseits jedoch auch die didaktischen Funktionen des Informierens und Darbietens von Sachverhalten, des Erarbeitens neuen Unterrichtsstoffes oder des Sicherens von Ergebnissen und der Reflexion (vgl. Gudjons, 2003). Besonders im Anfangsunterricht erleichtern offene Unterrichtsformen zusätzlich „den Aufbau von Regeln und Handlungsgewohnheiten, die auf die Selbstständigkeit und Eigenverantwortlichkeit der Kinder zielen“ (Knauf, 2001, S. 33). Zudem kann durch die häufigere Arbeit in Kleingruppen mehr personale Unterstützung sowie Aufmunterung oder Rückmel-

---

56 Waldis, Grob, Pauli und Reusser (2010a) verstehen darunter solche Lernsettings, in denen „die Lernaktivitäten von der Lehrperson eingeleitet und überwacht [werden], sie präsentiert die Lerninhalte, stellt Fragen und führt Aufgaben und Übungen ein“ (S. 177). Der Stundenablauf ist nach Meinung der Autoren in den meisten Fällen sequenziert, die Schüler stehen unter ständiger Beobachtung und bei Bedarf schiebt die Lehrperson kurze Interventionen ein.

57 Häufig werden die Begriffe Lehrerzentrierung und Lehrersteuerung synonym verwendet.

derung durch die Lehrperson erfolgen und somit die soziale Eingebundenheit gefördert werden (vgl. Hidi, Renninger & Krapp, 2004).

Die Methodenvielfalt wird vor allem dadurch sichtbar, dass die Schüler in einem sehr viel größeren Maße selbstständig alleine, zu zweit oder in kleineren Gruppen arbeiten (vgl. Grell, 2002). Götz et al. (2005) zeigten in ihrer Studie, dass in der Sekundarstufe bis zu 57 % des Gesamtunterrichts durch Frontalunterricht realisiert wird, während in der Grundschule nur ungefähr 39 % des Unterrichts in Form von Frontalunterricht stattfindet. Für die PERLE-Stichprobe bestätigten Lotz und Faust (2010) für den Anfangsunterricht in Deutsch und Mathematik eine ähnliche prozentuale Verteilung. Grundlage der Untersuchung bildeten Lehrpersonen, die an beiden Videostudien teilgenommen haben. Während im ersten Schuljahr (Videostudie Deutsch) nur 33 Prozent der untersuchten Lehrpersonen überwiegend schülerorientierte Sozialformen (Einzel-, Partner- und Gruppenarbeit) verwirklichten, waren es im zweiten Schuljahr (Videostudie Mathematik) bereits 50 Prozent der Lehrpersonen.

Die Methodenvielfalt spiegelt sich auch in der Raumgestaltung wider. Während ein Klassenraum in der Grundschule eher eine Lernwerkstatt mit verschiedenen Lernzonen darstellt und eine Vielzahl von Möglichkeiten individuellen Lernens bietet (vgl. Toman, 2007), lässt die Gestaltung vieler Klassenräume in den Sekundarstufen diese Vielzahl von Möglichkeiten individuellen Lernens nicht zu. Durch die eher offene Raumgestaltung mit flexibel veränderbaren Funktionsecken (Schreib-, Mal-, Lese- oder Medienecke) in der Grundschule können die Schüler im Klassenraum unterschiedlichen Tätigkeiten nachgehen bzw. vielfältige Lern- und Arbeitsangebote nutzen. Diese Funktionsvielfalt des Klassenraums, die frei wählbaren Arbeitsplätze für die Schüler sowie die häufige Arbeit in Kleingruppen stellten eine große Herausforderung der Videografie im Grundschulunterricht dar. Je nachdem welches Forschungsinteresse zugrunde liegt, sollte die Klassenraumgestaltung bzw. die hohe Flexibilität der Schüler während Schülerarbeitsphasen berücksichtigt werden.

Der Einsatz verschiedener Unterrichtsmethoden in einer Unterrichtsstunde hat zusätzlich häufigere Sozialformwechsel zur Folge. Auf die selbstständigen, schülerorientierten Arbeitsphasen mit einem Partner oder in der Gruppe folgt oft ein gemeinsames Klassengespräch, teilweise im Sitzkreis. Solche gemeinschaftlichen Situationen – wie der Klassenunterricht, wenn er als gemeinsames Planen, Erklären und Diskutieren von Regeln oder als Austausch von Erfahrungen betrachtet wird – „sind Voraussetzung für die Schaffung eines positiven Klimas der wechselseitigen Akzeptanz und der Anerkennung von Verschiedenheit in der Klasse sowie für den Aufbau von Ansätzen sozialer Mitverantwortung und Solidarität“ (Knauf, 2001, S. 32).

Alle hier genannten Unterschiede zwischen dem Unterricht in Grundschulen und anderen Schulstufen sind entscheidend für die Kamerapositionierung bzw. für die Erstellung des Kameraskripts (vgl. Kapitel 10). Durch die bessere Möglichkeit der Lehrperson zur individuellen Betreuung der Schüler in den

offenen Unterrichtsformen spielt die Lehrer-Schüler-Interaktion in der Grundschule vermutlich eine größere Rolle als in anderen Schulstufen. Inwieweit die unterschiedlichen Vorgaben im Rahmen der beiden Videostudien Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) Auswirkungen auf den Einsatz verschiedener Unterrichtsmethoden haben und inwieweit daraus unterschiedliche Unterrichtsverläufe resultieren, wird in Kapitel 9.3 nochmal ausführlich dargestellt. Zunächst wird die Methode des hoch inferenten Ratings zur Beschreibung der Unterrichtsqualitätsmerkmale differenziert betrachtet.

## 6.2 Hoch inferente Ratings zur Beurteilung der Unterrichtsqualität

Zur Erforschung der Merkmale von Unterrichtsqualität mithilfe von Unterrichtsvideos lassen sich die eingesetzten Methoden zur Gewinnung von Informationen über den Unterricht auf einem Kontinuum von *niedrig inferent* bis *hoch inferent* einordnen (vgl. Borich & Klinzing, 1986; Clausen et al., 2003; Reyer, 2004; Rosenshine, 1970; Seidel, 2003c). Diese Verfahren variieren im Grad der Inferenz (vgl. Hugener, Rakoczy, Pauli & Reusser, 2006; Seidel, 2003c). Nach Rosenshine (1970) beschreibt der Grad der Inferenz „the process intervening between the objective data seen or heard and [...] the coding of those data on an observational instrument“ (S. 281). Damit sind die interpretativen Schlussfolgerungen der Beobachter gemeint, die für die Kodier- bzw. Ratingentscheidungen notwendig sind. Der Unterschied zwischen niedrig inferenten Kodierungen und hoch inferenten Ratings lässt sich wie folgt verstehen:

### 6.2.1 Hoch inferente Ratings versus niedrig inferente Kodierungen

*Niedrig inferente Beobachtungen* beinhalten im Allgemeinen Aspekte spezifisch beobachtbarer Verhaltensweisen, die direkt quantitativ erfasst werden (quantitatives Verfahren) und fokussieren eher auf kurze, im Unterricht vergleichsweise häufig auftretende Aspekte bzw. Verhaltensäußerungen. Die Unterrichtsbeobachtung erfolgt auf Basis disjunkter Kategoriensysteme, in dem die einzelnen Kategorien durch genaue Definitionen und Regeln voneinander abgrenzbar sind. Die Regeln sind für den Beobachter klar definiert, es bedarf keiner größeren Interpretations- und Bewertungsleistung durch den Beobachter. Somit ist der Grad der Schlussfolgerungen für den Beobachter eher gering. Im Rahmen der PERLE-Studie gilt die Basiskodierung als ein Beispiel für eine niedrig inferente Kodierung. Hier wurden unter anderem die einzelnen Sozialformen (Partnerarbeit, Gruppenarbeit etc.) und deren Auftre-

tenshäufigkeiten sowie zeitliche Dauer kodiert, wobei den Kodierungen ein zweistufiges Antwortformat zugrunde lag („kommt vor“, „kommt nicht vor“, vgl. Lotz, 2013b; Lotz, Gabriel & Lipowsky, 2013). Durch solche niedrig inferenten Kodierungen erhält man demnach genaue Informationen über die Häufigkeit und den zeitlichen Anteil verschiedener Phasen oder Ereignisse sowie die zeitliche Abfolge und Sequenzierung des Unterrichts. Aus diesem Grund werden derartige Unterrichtsmerkmale fachübergreifend allgemein auch als *Sichtstruktur* oder *Oberflächenstruktur* von Unterricht bezeichnet (vgl. Hodel & Waldis, 2007; Meyer, 2004; Reyer, 2004; Seidel, Dalehefte & Meyer, 2001; Seidel, 2003a). Der Vorteil dieser niedrig inferenten Kodierungen liegt vor allem in der Genauigkeit und der hohen Reliabilität der Ergebnisse (vgl. Lotz et al., 2013). Allerdings büßen niedrig inferente Kodierungen häufig an Validität ein, „da die für die Kodierungen einzelnen Unterrichtsereignisse oder -komplexe detailliert aufgegliedert und entsprechenden Einzelkategorien zugeordnet werden“ (Seidel, 2003c, S. 106). Zudem lassen sich Aussagen zur Unterrichtsqualität durch ein solches kleinschrittiges Vorgehen nur schwer ableiten.

Um die Qualität von Unterricht bzw. die Qualität einzelner Unterrichtsabschnitte oder -ereignisse zu erfassen, kommen in der videobasierten Unterrichtsforschung aus diesem Grund *hoch inferente Ratings* zum Einsatz, die im Gegensatz zu niedrig inferenten Kodierungen die Tiefenstruktur von Unterricht erfassen können (vgl. Reyer, 2004)<sup>58</sup>. Die Analyseeinheit ist in der Regel die gesamte Lektion bzw. Unterrichtsstunde, da einzelne Anhaltspunkte häufig über die gesamte Lektion verteilt sind und sich nicht an spezifischen Ereignissen festmachen lassen (vgl. Waldis et al., 2010b). Das hoch inferente Rating besteht aus einem Gesamturteil, das die verschiedenen verhaltensnah definierten Indikatoren berücksichtigt (vgl. Rakoczy & Pauli, 2006). In dieses Gesamturteil können nicht nur die Häufigkeit, sondern auch die Intensität der betreffenden Verhaltensweisen sowie die Verteilung dieses Verhaltens innerhalb der beobachteten Klasse eingehen (vgl. Lotz et al., 2013). Im Gegensatz zu niedrig inferenten Kodierungen erfolgt die Einschätzung auf einer abgestuften Skala.

Die Integration und gleichzeitige Bewertung der komplexen und zusammenhängenden Unterrichtsereignisse erfordert seitens des Beobachters verstärkt interpretative Schlussfolgerungen (vgl. Seidel, 2003c), was dazu führt, dass hoch inferente Ratings im Vergleich zu niedrig inferenten Beobachtungen anfälliger für systematische und unsystematische Beurteilungsfehler sind (vgl. Clausen et al., 2003). Aus diesem Grund stellen die Beobachter bei der Erfassung von Unterrichtsmerkmalen durch hoch inferente Ratings eine grö-

---

58 Reyer (2004) unterscheidet zwischen einer lehrerseitigen und einer schülerseitigen Tiefenstruktur, je nachdem, ob die Tiefenstruktur auf der Untersuchungsebene der Lehrperson als Lehrkonzepte, Wissen und Einstellungen oder auf der Untersuchungsebene der Schüler als Kompetenzen und Einstellungen erfasst wird.

bere Fehlerquelle dar als bei der Kodierung niedrig inferenter Sichtstrukturen (vgl. Clausen et al., 2003; Hoyt, 2000; Praetorius et al., 2012), was wiederum zur Folge hat, dass die Beurteilerreliabilität (auch Inter-Rater-Reliabilität) meist geringer ist. Eine gewisse Objektivierung der intersubjektiven Einschätzungen der Rater kann durch ein ausführliches Training erreicht werden, bei dem die Rater ihre Interpretationen ausführlich diskutieren. Dieses Vorgehen ähnelt dem in der qualitativen Unterrichtsforschung. Lotz et al. (2013) vermuten, dass gerade der interpretative Gehalt solcher hoch inferenten Ratings der Grund für die Aussagekraft solcher Urteile in Bezug auf die Erklärung von Leistungs- und Motivationsentwicklung im Unterricht sein könnte. So konnten beispielsweise auch Clausen et al. (2003) zeigen, dass die Zusammenhänge niedrig inferenter Verfahren mit dem Lernerfolg der Schüler schwächer ausfallen als bei hoch inferenten Einschätzungen.

Bezüglich der Validität haben hoch inferente Urteile gegenüber niedrig inferenten Beobachtungen den Vorteil, dass sie oft einen stärkeren Bezug zur Theorie haben, wodurch nach Clausen (2002; vgl. auch Reyer, 2004) die Qualität der Aussagen höher bewertet werden kann. So können beispielsweise Basisdimensionen der Unterrichtsqualität direkt durch hoch inferente Ratings operationalisiert werden. Auch ein Großteil der in Tabelle 3 (vgl. Kapitel 2.2) aufgeführten Videostudien in der Grundschule setzen hoch inferente Ratings zur Erfassung der Unterrichtsqualität ein. Im nachfolgenden Kapitel werden die Bestandteile eines hoch inferenten Ratingverfahrens sowie dessen Beziehungsstruktur genauer dargestellt.

### 6.2.2 Bestandteile eines Ratingverfahrens nach Langer & Schulz von Thun (2007)

In den Sozialwissenschaften zählen Rating-Skalen zu den am häufigsten verwendeten Erhebungsverfahren, gelten jedoch auch als sehr umstritten (vgl. Praetorius et al., 2012). Langer und Schulz von Thun (2007) beschreiben ein Ratingverfahren als ein Beobachtungsverfahren, das dadurch gekennzeichnet ist, „daß die Meßwerte durch Schätzurteile menschlicher Beurteiler zustandekommen“ (S. 14). Diese subjektiven Urteile macht sich auch das hoch inferente Ratingverfahren zunutze (vgl. Kapitel 6.2). Allgemein kennzeichnen Ratingverfahren folgende Bestandteile (adaptiert nach Langer & Schulz von Thun, 2007, S. 14f):<sup>59</sup>

- (1) Das *Objekt*, an dem etwas gemessen werden soll (= Messobjekt). Im Rahmen der vorliegenden Arbeit handelt es sich bei den Objekten um Unterrichtsvideos (= Video) in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr).

---

59 Für die vorliegende Arbeit wurden die Bestandteile eines Ratingverfahrens nach Langer und Schulz von Thun (2007) für die Beurteilung von Unterrichtsvideos adaptiert.

- (2) Dabei ist für die vorliegende Arbeit nicht alles, was diese Objekte charakterisiert, von Interesse. Nach Langer und Schulz von Thun ist das, „was gemessen werden soll, [...] das *Merkmal*“. In der vorliegenden Studie ist beispielsweise von Interesse, ob die Lehrpersonen allgegenwärtig sind („Allgegenwärtigkeit der Lehrperson“ als Merkmal der Klassenführung) oder ob die Schüler sich gegenseitig wertschätzen und respektieren („gegenseitige Anerkennung der Schüler“ als Merkmal des Unterrichtsklimas). Jedes Merkmal wird mithilfe einer Grundidee und durch verhaltensnahe Indikatoren beschrieben (vgl. Kapitel 12.1). Solche Merkmale bzw. Unterrichtsereignisse und -prozesse variieren in ihrer Komplexität (vgl. Clausen, 2002), d. h., die Ereignisse unterscheiden sich darin, wie gut sie am Objekt beobachtbar sind. Merkmale mit einem hohen Grad an Komplexität sind nicht direkt an konkreten Handlungen beobachtbar, sondern können erst durch die Integration verschiedener Verhaltensweisen erschlossen werden, z. B. sind zwischenmenschliche Beziehungen nicht unmittelbar beobachtbar.
- (3) Durch die Definition des Merkmals wird ein Auftrag an die *Rater* formuliert: „Achte auf die einzelnen Indikatoren, konzentriere dich auf diese und auf nichts anderes!“ (ebd., S. 15).
- (4) „Entsprechend der Beobachtung geht vom Objekt [= Video, Erg. v. Verf.] eine Eindruckswirkung auf den Rater aus. Der Rater gibt über diese Eindruckswirkung, über das erlebte Mehr oder Minder, Auskunft in Form eines Urteils auf einer abgestuften *Skala*“ (ebd., S. 15). Im vorliegenden Fall erfolgt die Einschätzung auf einer vierstufigen Skala von 4 (= hohe Übereinstimmung) bis 1 (= niedrige Übereinstimmung) (siehe auch Kapitel 12.1.4). Demnach ist – wie auch Langer und Schulz von Thun schreiben – „mit der Abstufung eine zahlenmäßige Zuordnung verbunden“ (ebd., S. 15).

Das Schema in Abbildung 10 verdeutlicht die Beziehungsstruktur der einzelnen Bestandteile des Ratingprozesses. Es verdeutlicht, dass alles, was am Messobjekt (= Video) auffällt, als Merkmal zugrunde gelegt werden kann.

Das Merkmal seinerseits legt den Betrachtungsaspekt fest, unter dem das Objekt zu sehen ist und spezifiziert entsprechend den Beobachtungsauftrag für den Rater. Danach beobachtet der Rater das Objekt und empfängt einen Eindruck. Über diesen gibt er Auskunft durch Angabe einer Ausprägung auf der Merkmalsskala (Langer & Schulz von Thun, 2007, S. 15).

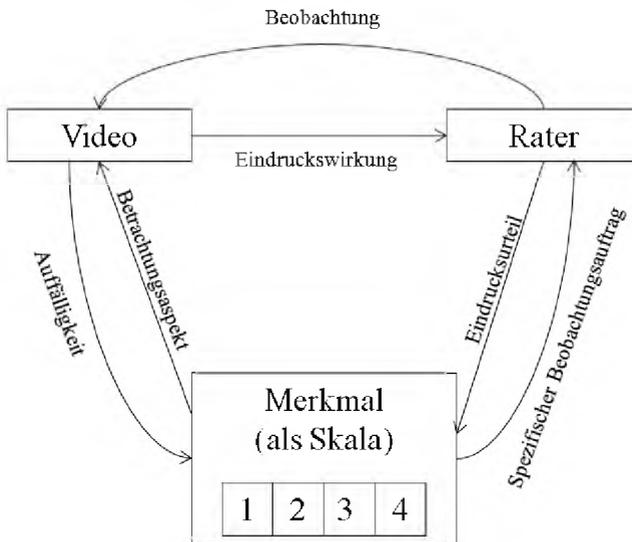


Abbildung 10: Beziehungsstruktur von Video (= Messobjekt), Rater und Merkmal (adaptiert nach Langer & Schulz von Thun, 2007, S. 15).

### 6.3 Gütekriterien zur Qualitätskontrolle bei Videoanalysen

Auch im Zusammenhang mit Videoanalysen müssen verschiedene Güteprüfungen durchgeführt werden (vgl. Aufschnaiter & Welzel, 2001; Ulewicz & Beatty, 2001). Obwohl der Entwicklung des hoch inferenten Ratinginstruments zur Klassenführung und zum Unterrichtsklima unter anderem ein qualitatives Verfahren zugrunde liegt (vgl. Kapitel 12.1.2), kann die Auswertung der Videodaten als ein quantitatives Verfahren angesehen werden. Somit sind die Gütekriterien der klassischen Testtheorie sinngemäß auch auf die Güteprüfung anderer Datenerhebungsmethoden wie beispielsweise der Beobachtung übertragbar (vgl. Rost, 2007). Im Rahmen bisheriger Videostudien wurden aus diesem Grund die drei Gütekriterien der klassischen Testtheorie (Objektivität, Validität, Reliabilität) in teilweise modifizierter Form verwendet (z. B. Hugener, 2006; Vehmeyer, 2009). Im Folgenden wird auf die drei Gütekriterien näher eingegangen.

*Objektivität:* Die Objektivität der Videodaten wird durch eine transparente Darstellung des Vorgehens sowohl während der Aufzeichnung der Videos als auch während des eigentlichen Ratings gewährleistet (vgl. Hugener, 2006). Eine ausführliche Dokumentation des Vorgehens beim hoch inferenten Ra-

ting stellt vor allem sicher, dass die Datenerhebung unabhängig vom Beurteiler ist und weniger anfällig gegenüber Urteilsfehlern. Dadurch wird sichergestellt, dass auch andere Forscher mithilfe des hoch inferenten Ratingsystems zu vergleichbaren Ergebnissen kommen. Bortz und Döring (2006) sprechen in diesem Zusammenhang von der sogenannten *Durchführungsobjektivität*. Diese Form der Objektivität wird meist durch schriftlich festgelegte Durchführungsanweisungen gesichert. Im Rahmen der vorliegenden Arbeit wurde das Vorgehen in Form eines Ratingmanuals festgehalten, das unter anderem die präzise Definition und Operationalisierung der zu beurteilenden Merkmale der Klassenführung und des Unterrichtsklimas sowie die Aufstellung von Ratingregeln beinhaltet (vgl. Kapitel 19). Durch eine ausführliche Schulung der Rater mithilfe dieses Manuals (vgl. Kapitel 12.1.1) wurde zusätzlich dem Auftreten verschiedener Urteilsfehler entgegengewirkt (vgl. Kapitel 12.1.4). Da vor allem bei hoch inferenten Ratings individuelle Deutungen in die Interpretationen der Ausprägungen einzelner Merkmale einfließen, gilt es in diesem Zusammenhang, zusätzlich ein gewisses Maß an *Interpretationsobjektivität* zu sichern.

*Validität*: Die Validität (= Gültigkeit) stellt das wichtigste Kriterium für die Qualität von Datenerhebungen dar. „Die Validität gibt an, ob ein Test das misst, was er messen soll bzw. was er zu messen vorgibt [...]“ (Bortz & Döring, 2006, S. 200). Eine fehlerhafte Vorstellung eines Konzepts kann die Richtigkeit der Beurteilungen beeinträchtigen (Langer & Schulz von Thun, 2007). Die Validität von Videoanalysen bemisst sich demnach daran, inwieweit ein beobachtetes Ereignis eindeutig einem Merkmal bzw. einer Kategorie zugeordnet werden kann. Als wichtigstes Kriterium für die Videoanalyse gilt in diesem Zusammenhang die interpersonale Konsensbildung (oder *konsensuelle Validierung*, vgl. Bortz & Döring, 2006, S. 328). Sie beschreibt das gemeinsame Einigen der Rater auf die Zuordnung beobachtbaren Verhaltens zu einem hoch inferenten Merkmal der Klassenführung und des Unterrichtsklimas. Eine erste konsensuelle Validierung erfolgte während der Schulung. Erst wenn sich die Rater und die Entwicklerin des Ratingsystems auf ein gemeinsames theoretisches Verständnis geeinigt hatten, galt das Training als abgeschlossen (vgl. Kapitel 12.1.1). Weitere Aspekte, die möglicherweise gegen die Validität der Videodaten sprechen, sind auftretende Verzerrungen durch Reaktivität sowie der eingeschränkte Beobachtungszeitraum, der einer einmaligen Unterrichtsbeobachtung zugrunde liegt (vgl. Kunter, 2005). Inwieweit die Lehrpersonen berichten, dass ihr aufgezeichneter Unterricht eine typische Stunde darstellt, wird in Kapitel 11.3 beschrieben. In der vorliegenden Studie können Verzerrungen durch Reaktivität als minimal angesehen werden, da in beiden Videostudien ein Großteil der Lehrpersonen berichtete, dass der aufgezeichnete Unterricht eine nahezu alltägliche Unterrichtssituation widerspiegelte. Es kann jedoch nicht ausgeschlossen werden, dass sich die Lehrpersonen für diese aufgezeichnete Stunde besonders vorbereitet haben, um einen aus ihrer Sicht „guten“ Unterricht zu halten. Für die vorliegende Ar-

beit von zentraler Bedeutung ist die Frage nach der *prognostischen Validität* bzw. *Vorhersagevalidität* der hoch inferent erfassten Klassenführung und des Unterrichtsklimas. Die Daten, die mithilfe des hoch inferenten Ratingsystems für den Anfangsunterricht erhoben werden, sollten entsprechend der Theorie (vgl. Kapitel 5) die Leistungs- und Selbstkonzeptentwicklung in den ersten beiden Schuljahren vorhersagen können.

*Reliabilität*: Die Reliabilität (= Zuverlässigkeit) gibt die Messgenauigkeit eines Instruments an (vgl. Bortz & Döring, 2006). In der videobasierten Unterrichtsforschung wird hauptsächlich der Durchschnitt aus den Raterurteilen ermittelt, um eine ausreichende Reliabilität und Validität der Urteile zu gewährleisten. Ziel solcher Verfahren im Allgemeinen ist es, das Ausmaß der Übereinstimmungen (= Konkordanzen) der Einschätzungsergebnisse bei unterschiedlichen Beobachtern oder Ratern zu berechnen. Die Zusammenfassung solcher Urteile setzt jedoch eine hinreichende Übereinstimmung der Urteile voraus. Zur Überprüfung der Qualität der Beurteilungen wurde im Rahmen der vorliegenden Studie auf die Generalisierbarkeitstheorie zurückgegriffen (Cronbach et al., 1972; Shavelson & Webb, 1991; vgl. Kapitel 6.3.2). Mithilfe dieses Ansatzes kann überprüft werden, inwieweit individuelle Fehler bzw. charakteristische Unterschiede in der Beurteilung durch mehrere Rater Auswirkungen auf den Beurteilungsprozess haben (vgl. Waldis et al., 2010b).

### 6.3.1 Mögliche Urteilsfehler bei dem Einsatz des Ratingsystems

Obwohl Beobachtungen häufig als beste Methode zur Erfassung von Unterrichtsmerkmalen angesehen werden (vgl. Clare et al., 2001; Praetorius et al., 2012), unterliegen sie Fehlerquellen (vgl. Hoyt, 2000; Praetorius et al., 2012). Besonders das Problem des *Rater-Bias*<sup>60</sup> (= Urteilsfehler) wird häufig als Nachteil der Methode angesehen. Die wesentliche Kritik gründet sich hauptsächlich auf die Subjektivität der Urteile. Nach Hoyt (2000) wird als Rater-Bias eine mangelnde Übereinstimmung der Rater verstanden, die einerseits auf unterschiedliche Interpretationen der Ratingskala und andererseits auf rater-spezifische Wahrnehmungen, beispielsweise der Unterrichtsvideos (= Beurteilungsobjekte, vgl. Kapitel 6.2.2), zurückgeführt werden können.<sup>61</sup> Eine solche Nicht-Übereinstimmung stellt eine Einschränkung der Reliabilität der Ratings dar.

Verschiedene Studien konnten zeigen, dass die Messergebnisse durch Beurteilereffekte verzerrt werden können bzw. bei Beurteilungen durch Drit-

---

60 Unter Rater-Bias wird der Anteil an den Ratings verstanden, der durch den Rater und nicht durch das zu beurteilende Material selbst bestimmt ist (vgl. Hoyt, 2000). Dabei werden die Begriffe Rater-Bias, Rater-Effekte oder Rater-Fehler häufig synonym verwendet.

61 Aus diesem Grund wird der Einfluss der Rater auch als Messfehler angesehen.

te mit Messfehlern zu rechnen ist (vgl. Bortz & Döring, 2006; Hoyt & Kerns, 1999; Hoyt, 2000; Praetorius et al., 2012). So bestätigen beispielsweise Hoyt und Kerns (1999) nach einer Metaanalyse von 79 Ratingskalen, dass Differenzen zwischen Urteilen häufig durch Unterschiede in der Tendenz zur Strenge bzw. Milde der Rater erklärt werden kann. Als Folge dieser Differenzen in den Beurteilungen ergibt sich, „dass sich die merkmalsunabhängige Varianz erhöht und somit die Validität der Daten gemindert wird“ (Pietsch & Tosana, 2008, S. 433).

In der Sozialpsychologie werden neben dem Milde-Strenge-Fehler weitere Arten sogenannter Verzerrungseffekte genauer beschrieben und untersucht (vgl. von Eye, 2006). Folgende relevante Fehlertypen, die beim Einsatz von Bewertungsskalen auftreten können, werden voneinander unterschieden (z. B. Bortz & Döring, 2006; Pietsch, 2008; Pietsch & Tosana, 2008): Halo-Effekt, zentrale Tendenzen, Rater-Ratee-Interaktionen und Primacy-Recency-Effekte. Was unter den jeweiligen Beurteilungsfehlern verstanden wird und welche Maßnahmen in diesem Zusammenhang zur Vermeidung solcher Fehler im Rahmen der vorliegenden Arbeit getroffen wurden, wird in Kapitel 12.1.4.2 ausführlich dargestellt.

Zur Überprüfung der Güte des hoch inferenten Ratingsystems und um den Einfluss des Rater-Bias auf die Messergebnisse zu ermitteln, wurde in der vorliegenden Arbeit zu mehreren Messzeitpunkten die Inter-Rater-Reliabilität anhand des relativen Generalisierbarkeitskoeffizienten (G-Koeffizienten) berechnet. Der Ansatz der Generalisierbarkeitstheorie wird im Folgenden ausführlich beschrieben.

### *6.3.2 Generalisierbarkeitsstudien zur Berechnung der Inter-Rater-Reliabilität*

In der videobasierten Unterrichtsforschung gilt der Ansatz der Generalisierbarkeitstheorie (G-Studien, Cronbach et al., 1972; Shavelson & Webb, 1991) als weit verbreitete Methode zur Berechnung der Beobachterübereinstimmung bei hoch inferenten Daten (zusf. Praetorius et al., 2012; Clausen et al., 2003; Rakoczy, 2008; Renkl & Helmke, 1993). Wie Praetorius et al. (2012) jedoch feststellen, bilden Beobachtungen des Unterrichts nur in einem geringen Teil der Studien die Datengrundlage. Der restliche Teil fokussiert eher Einschätzungen der Unterrichtsqualität durch die Schüler.

Im Kern solcher G-Studien geht es darum, eine gemessene Variation auf verschiedene potenzielle Varianzquellen zurückzuführen und deren relative Anteile zu bestimmen, anhand derer wiederum die Güte der Messung abgeschätzt werden kann (vgl. Rakoczy & Pauli, 2006; Stumpp & Großmann, 2009). Im Gegensatz zur klassischen Testtheorie können Unterschiede in den Beurteilungen somit nicht nur auf eine globale Fehlerquelle zurückgeführt werden (wie in der KTT der Fall, vgl. Shavelson, Webb & Rowley, 1989), son-

dern es kann überprüft werden, (1) welcher Anteil der vorliegenden Variation tatsächliche Unterschiede zwischen dem Unterricht (Videos) abbildet (= „wahre“ Varianz), (2) welcher Anteil auf charakteristische Unterschiede in der Beurteilung der Unterrichtsstunden durch die zwei Rater zurückzuführen ist (systematische Fehlervarianz) und (3) wie viel unsystematische Variation in das Maß einfließt. Eine solche Varianzanalyse gibt Antwort auf folgende Fragen (in Anlehnung an Langer & Schulz von Thun, 2007, S. 85f):

- Spiegeln die Unterschiede in den Werten Unterschiede in den Videos wider (= konstruktbezogene oder „wahre“ Varianz)?
- Oder spiegeln die Unterschiede in den Werten Uneinigkeiten zwischen den Ratern wider (= systematische Fehlervarianz)?

Uneinigkeiten zwischen den Ratern können beispielsweise bezüglich ihrer Skalenhandhabung bestehen, d. h. der eine Rater vergibt in der Regel höhere oder niedrigere Werte als der andere (Rater-Bias, z. B. Milde-Streng-Fehler). Unterscheiden sich die Rater jedoch in ihrer Reaktion auf die Besonderheiten des Videos, d. h. besteht unter den Ratern Uneinigkeit hinsichtlich der Relationen und Abstände zwischen den Objekten (= Video), schlägt sich dies bei der Varianzanalyse als sogenannte Wechselwirkungsvarianz (Interaktionseffekt zwischen Ratern und Videos) nieder, welche als unsystematischer Messfehler bezeichnet wird.

Nach Pietsch (2008; vgl. auch Clausen et al., 2003; Langer & Schulz von Thun, 2007) sollte der Anteil der Varianz, der auf tatsächliche Unterschiede zwischen Unterrichtsvideos zurückzuführen ist (= konstruktbezogene oder „wahre“ Varianz) am größten sein.<sup>62</sup> Der Anteil der Varianz, der auf individuelle Urteilsfehler der Rater zurückzuführen ist (systematische Varianz), sollte deutlich geringer sein.<sup>63</sup> Auch der Anteil der Wechselwirkungsvarianz (un-

---

62 In einer Studie von Pietsch und Tosana (2008) lag die Varianz der Facette Unterricht (= „wahre“ Varianz) bei rund 25 Prozent, was bedeutet, dass die Unterschiede in der Messung von Unterrichtsqualität vom beobachteten Unterricht selbst abhängen. Der Anteil ist relativ gering, lässt sich jedoch durch den vergleichsweise hohen Varianzanteil erklären, der auf den mit dem Messfehler konfundierten Interaktionseffekt Unterricht x Item, Residuum (ca. 55 %), entfällt. „Dies bedeutet, dass die beobachtete Qualität von Unterricht in Abhängigkeit davon, auf welchen Items hohe Werte markiert werden, unterschiedlich hoch ausfällt. Es lässt sich also eine Wertigkeit von Items bei der Bestimmung von Unterrichtsqualität nachweisen“ (ebd., S. 442). Der Anteil, der durch die Beobachtungitems selbst erklärt wird, liegt bei 11,5 %. Zusätzlich lässt sich ein Anteil in der Variation bei der Messung von rund 8 % auf die Interaktion zwischen Beobachter und Items zurückführen.

63 Im Rahmen der Metaanalyse von Hoyt und Kerns (1999) zeigte sich, dass circa 37 Prozent der Varianz bei den Bewertungen durch Beurteiler auf individuelle Urteilsfehler zurückzuführen sind. Auch nach der Schulung von Ratern zeigt sich noch eine Fehlervarianz von 14 Prozent (vgl. Hoyt, 2000). Pietsch (2008) kommt jedoch zu einem deutlich niedrigeren Ergebnis von rund neun Prozent Fehlervarianz auf der Varianzkomponente Beobachter/ Rater, wobei in seiner Studie diese vor allem auf die unterschiedliche Interpretation von Unterrichtsqualitätsitems durch die bewertenden Unterrichtsbeobachter zurückzuführen

systematische Varianz) sollte so gering wie möglich sein. Bei einer sehr hohen „wahren“ Varianz kommt es jedoch zwangsläufig zu einem günstigen Verhältnis zu diesem unsystematischen Fehleranteil.

Als Reliabilitätsmaß wird im Rahmen solcher G-Studien der G-Koeffizient berechnet. Dieser gilt als Maß für Messgenauigkeit. Analog zum Reliabilitätskoeffizienten der klassischen Testtheorie wird der G-Koeffizient als der Anteil der universalen Varianz ( $\sigma^2p$ ) an der Gesamtvarianz (universale + relative Fehlervarianz:  $\sigma^2p + \sigma^2\text{fehler}$ ) definiert und kann einen Wert zwischen Null und Eins annehmen. Wie oben bereits dargestellt, ergibt sich die Gesamtvarianz aus der Summe der Varianzkomponenten sowie deren Interaktion (vgl. Gleichung 1; Renkl & Helmke, 1993; Stumpp & Großmann, 2009).

$$E(\rho^2) = \frac{\sigma^2p}{\sigma^2p + \sigma^2\text{fehler}} \quad (1)$$

Man unterscheidet zwischen einem absoluten und einem relativen G-Koeffizienten. Beim absoluten G-Koeffizienten wird in den Fehlerterm sowohl die systematische als auch die unsystematische Fehlervariation mit einbezogen, während der relative G-Koeffizient nur die unsystematische Fehlervarianz in die Berechnung einschließt (vgl. Rakoczy & Pauli, 2006). Der absolute G-Koeffizient ist vor allem dann bedeutsam, wenn nicht das gesamte Material von allen Ratern beurteilt wird, da trotzdem sichergestellt werden kann, dass die absolute Höhe der Urteile übereinstimmt und nicht einer der Rater beispielsweise deutlich strenger urteilt als der andere (vgl. Brennan & Kane, 1977). Der relative G-Koeffizient wird dagegen eher für Entscheidungen herangezogen, bei denen die Rangreihe der Beurteilungsgegenstände von Bedeutung ist (vgl. Rakoczy & Pauli, 2006). Da im Rahmen der vorliegenden Arbeit die Rangreihe der Urteile zwischen den Ratern von Relevanz ist, wird der relative G-Koeffizient als Maß für die Güte der Messung herangezogen. Dieser gibt an, inwieweit die Einschätzungen der jeweils betrachteten hoch inferenten Merkmale der Klassenführung und des Unterrichtsklimas durch die beiden Rater ausreichende Generalisierbarkeit (Zuverlässigkeit) aufweisen (vgl. Gabriel & Lipowsky, 2013). Die Berechnungen der relativen Anteile der Varianzkomponenten und des relativen G-Koeffizienten erfolgen im Rahmen der vorliegenden Arbeit mit dem Programm GT (vgl. Ysewijn, 1997).

Da der relative G-Koeffizient als Pendant zum Reliabilitätskoeffizienten aus der klassischen Testtheorie angesehen werden kann (vgl. Clausen et al., 2003), gelten für ihn die gleichen Beurteilungsrichtlinien (vgl. Webb, Shavelson & Haertel, 2007; Wirtz & Caspar, 2002). In der vorliegenden Arbeit wird ein relativer G-Koeffizient von  $\geq .65$  als Kriterium für eine zufriedenstellende Qualität der Daten festgelegt. Dieser muss für die Verwendung der Daten in

---

ren sind. Bei diesem geringen Varianzanteil, so Pietsch (2008), gilt die Komponente Beobachter / Rater aber als zu vernachlässigende Fehlerquelle bei der Beurteilung der Unterrichtsqualität.

weiterführenden Analysen gewährleistet sein. Inwieweit die einzelnen Merkmale der Klassenführung und des Unterrichtsklimas diesem Kriterium standhalten, wird in Kapitel 13 überprüft.

## 7 Forschungsfragen der Arbeit

Auf Grundlage des in den vorherigen Kapiteln dargelegten Forschungsstandes gilt es an dieser Stelle, die Forschungsfragen der vorliegenden Arbeit zu entwickeln. Um die gewonnenen Daten des hoch inferenten Ratings für weitere Hypothesenprüfungen nutzen zu können, müssen die gängigen Gütekriterien erfüllt sein (vgl. Kapitel 6.3). Aus diesem Grund werden anschließend die Fragen zur Güte (Inter-Rater-Reliabilität) des hoch inferenten Ratingsystems vorgestellt (Kapitel 7.1). Danach folgen die Forschungsfragen zur Ausprägung (Kapitel 7.2) und zur Struktur (Kapitel 7.3) der Unterrichtsqualitätsmerkmale in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr). Dem schließen sich die Fragen zu multikriterialen Effekten einer effektiven Klassenführung und eines positiven Unterrichtsklimas im Anfangsunterricht an, wobei erneut beide Fächer getrennt voneinander betrachtet werden (Kapitel 7.4).

### 7.1 Inter-Rater-Reliabilität

Obwohl die beiden Basisdimensionen Klassenführung und Unterrichtsklima als fachübergreifend gültige Unterrichtsqualitätsmerkmale – auch in der Grundschule – gelten (vgl. Kapitel 2), wurden diese im Rahmen von Videostudien bislang hauptsächlich in naturwissenschaftlichen Fächern untersucht (vgl. Kapitel 2.2). Für diesen Bereich liegen hoch inferente Ratingsysteme vor, mit deren Hilfe Merkmale der Unterrichtsqualität (mit unterschiedlicher Schwerpunktsetzung) im Kontext von Beobachtungs- bzw. Videostudien erfasst werden können. Obwohl die Merkmale der Unterrichtsqualität als allgemeine Merkmale betrachtet werden, existieren m. W. n. bislang keine Videostudien, in denen ein einmal entwickeltes hoch inferentes Ratingsystem für verschiedene Fächer eingesetzt wurde. Vor dem Hintergrund, dass ein einmal entwickeltes Ratingssystem zu Merkmalen der Unterrichtsqualität auch in anderen Studien und unabhängig vom Fach valide und reliable Daten erfassen sollte, liegt das Hauptaugenmerk der vorliegenden Arbeit auf der Operationalisierung eines *fachunspezifischen* hoch inferenten Ratingsystems zur Erfassung der Klassenführung und des Unterrichtsklimas im Anfangsunterricht.

Fachunspezifisch operationalisiert bedeutet in diesem Zusammenhang, dass weder die Indikatoren der einzelnen Merkmale des hoch inferenten Rating-systems auf fachspezifische Besonderheiten Bezug nehmen, noch das Rating spezifische fachwissenschaftliche oder fachdidaktische Kenntnisse von den Ratern erfordert (vgl. Kapitel 12.1). Das hoch inferente Ratingsystem erhebt somit den Anspruch, in verschiedenen Fächern im Anfangsunterricht eingesetzt werden zu können.

In einem ersten Schritt wird überprüft, ob mithilfe des hoch inferenten Ratingsystems – unabhängig vom Fach – reliable Daten für den Anfangsunterricht generiert werden können. In der videobasierten Forschung wird zur Überprüfung der Reliabilität der Videodaten häufig die Beobachterübereinstimmung bzw. Inter-Rater-Reliabilität berechnet (vgl. Kapitel 6.3). Eine hinreichende Übereinstimmung der Urteile der zwei Rater ist Voraussetzung für die spätere Zusammenfassung der Urteile der Rater und für die Verwendung der Daten in weiteren Analysen. Zur Überprüfung der Beobachterübereinstimmung wird in der vorliegenden Arbeit auf den Ansatz der Generalisierbarkeitstheorie (vgl. Cronbach et al., 1972; Shavelson & Webb, 1991) zurückgegriffen und der relative G-Koeffizient als ein Übereinstimmungsmaß für hoch inferente Ratings herangezogen. Hierfür werden für jedes der hoch inferent erfassten Merkmale Generalisierbarkeitsstudien (G-Studien) durchgeführt (ausführlich in Kapitel 6.3.2). In einem ersten Schritt erfolgt eine fachunspezifische Berechnung der G-Studien, also auf Basis aller Unterrichtsvideos der Fächer Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr). Folgende Forschungsfrage bildet den Ausgangspunkt für alle weiteren Analysen der vorliegenden Arbeit:

*Forschungsfrage Nr. 1:*

Lassen sich die fachunspezifisch operationalisierten Merkmale der Klassenführung und des Unterrichtsklimas unabhängig vom Fach reliabel mithilfe von hoch inferenten Schätzverfahren erfassen?

In einem weiteren Schritt wird der Frage nachgegangen, ob das fachunspezifisch operationalisierte hoch inferente Ratingsystem für verschiedene Fächer reliable Daten liefert. Dabei stellt sich die Frage, ob die zwei Rater mithilfe des hoch inferenten Ratingsystems zu ähnlichen Einschätzungen in den beiden Fächern gelangen und somit eine hohe fachspezifische Beobachterübereinstimmung erzielen, oder ob sich durch den Einsatz des fachunspezifischen operationalisierten Ratingsystems Schwierigkeiten für die Einschätzung der Merkmale in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) ergeben. Durch die zusätzliche Durchführung von fachspezifischen G-Studien wird der Frage nachgegangen, ob sich Unterschiede in der Höhe der Inter-Rater-Reliabilität für die Merkmale der Klassenführung und des Unterrichtsklimas in Abhängigkeit vom Fach aufzeigen lassen (*Forschungsfrage Nr. 2*). Dies wird in Kapitel 13.2 explorativ untersucht. Unterschiede in

der Beurteilbarkeit der Merkmale könnten ein Hinweis darauf sein, dass eine fachspezifische Beurteilung der Merkmale nötig ist bzw. allgemeine Merkmale der Klassenführung und des Unterrichtsklimas im Anfangsunterricht je nach Fach bzw. Schuljahr unterschiedlich operationalisiert werden müssen. Es werden keine Hypothesen formuliert, da erstmalig für die Grundschule ein und dasselbe hoch inferente Ratingsystem für verschiedene Fächer eingesetzt wurde.

*Forschungsfrage Nr. 2:*

Zeigen sich Unterschiede in der Inter-Rater-Reliabilität in Abhängigkeit vom Unterrichtsfach?

Neben dem relativen G-Koeffizienten stellt die eigentliche Information, die im Rahmen einer G-Studie gewonnen wird, die Höhe der geschätzten Varianzkomponenten (VK) dar (vgl. Kapitel 6.3.2). Im Kontext solcher G-Studien kann die Variation in der Reliabilität auf drei potenzielle Varianzquellen zurückgeführt werden: (1) auf die Videos selbst, (2) auf die Rater sowie (3) auf die Interaktion zwischen Video und Rater. Bislang wurden nur in einzelnen Studien, die die Unterrichtsqualität auf Basis von Beobachtungen erfassen, sowohl der G-Koeffizient als auch die prozentuale Verteilung der Varianzkomponenten berichtet (zusf. Praetorius et al., 2012). Da hoch inferente Ratings meist mehr oder weniger starken Verzerrungen unterliegen (vgl. Kapitel 6.3.1), gilt es vor allem, den Einfluss des Rater-Bias bei der Einschätzung der Merkmale der Klassenführung und des Unterrichtsklimas so niedrig wie möglich zu halten. Da die Durchführung der Videoanalysen auf Basis einer intensiven Schulung von zwei Ratern erfolgte (vgl. Kapitel 12.1.4), sollte eine gewisse intersubjektivität der Urteile gegeben und somit der relative Anteil der systematischen Varianz an der beobachteten Varianz gering sein. Folgende Forschungsfragen sollen in diesem Kontext beantwortet werden:

*Forschungsfrage Nr. 3.1:*

Wie groß ist der Varianzanteil, der auf Unterschiede in den Beurteilungen durch die zwei Rater zurückzuführen ist (= systematische Fehlervarianz) bzw. ist die Erfassung der Merkmale der Klassenführung und des Unterrichtsklimas durch die Subjektivität der Rater systematisch verfälscht?

*Forschungsfrage Nr. 3.2:*

Zeigen sich Unterschiede in der Höhe der systematischen Fehlervarianz in Abhängigkeit vom Fach?

Da im Rahmen der vorliegenden Arbeit verschiedene Verfahren zur Minimierung dieser Ratereffekte (Subjektivität) angewendet wurden, hierzu zählt beispielsweise die Schulung der Rater (vgl. Kapitel 12.1.4.2), sollte der relative Anteil der systematischen Fehlervarianz in beiden Fächern eher gering ausfallen.

Nach der Überprüfung der Forschungsfragen Nr. 1 bis 3.2 werden zusätzliche Fragestellungen innerhalb von drei Teilstudien untersucht. In Abhängigkeit von der Güte der Daten des hoch inferenten Ratings bzw. einer zufriedenstellenden Inter-Rater-Reliabilität für die zwei Videostudien stehen in der Studie 1 zunächst die deskriptiven Auswertungen zur Ausprägung der Merkmale der Klassenführung und des Unterrichtsklimas in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) im Vordergrund der Analysen. Im nachfolgenden Kapitel 7.2 werden die in diesem Zusammenhang relevanten Forschungsfragen erläutert. Neben Analysen zur Ausprägung wird die Faktorenstruktur der hoch inferent erfassten Merkmale überprüft. In Kapitel 7.3 werden die Forschungsfragen zur Faktorenstruktur hergeleitet. Abschließend werden die Forschungsfragen formuliert, die sich auf die Bedeutung der Klassenführung und des Unterrichtsklimas für multikriteriale Schülermerkmale beziehen (vgl. Kapitel 7.4).

## 7.2 Ausprägung der Unterrichtsqualitätsmerkmale (Studie 1)

Für den Anfangsunterricht lassen sich durch die in PERLE implementierten Videostudien in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) eine Vielzahl von Erkenntnissen über die tatsächliche Ausgestaltung der Klassenführung und des Unterrichtsklimas im Anfangsunterricht gewinnen. Um der Frage der Ausprägung der beobachteten Merkmale der Klassenführung und des Unterrichtsklimas nachzugehen, werden die aus dem hoch inferenten Rating gewonnenen Daten deskriptiv ausgewertet. Die in Kapitel 3.2 und 4.2 theoretisch beschriebenen Merkmale dienen als Grundlage. In Abhängigkeit von einer zufriedenstellenden Inter-Rater-Reliabilität der einzelnen Merkmale der Klassenführung und des Unterrichtsklimas (siehe *Forschungsfragen Nr. 1 bis 3.2*) soll in der Studie 1 der folgenden Frage nachgegangen werden:

*Forschungsfrage Nr. 4:*

Welche Ausprägungen zeigen die einzelnen Merkmale der Klassenführung und des Unterrichtsklimas in den videografierten Unterrichtsstunden im ersten (Fach Deutsch) und zweiten (Fach Mathematik) Schuljahr?

Studien im ersten Schuljahr belegen, dass Beobachter sowohl die Klassenführung als auch das Unterrichtsklima ähnlich positiv einschätzen (vgl. Hamre & Pianta, 2007; NICHD ECCRN, 2002; Eckerth et al., 2012). So konnte auf Basis von Unterrichtsbeobachtungen mit dem COS-1 (Classroom Observation System for First Grade) bestätigt werden, dass ein Großteil der beobachteten Klassen im ersten Schuljahr mittlere bis hohe Ausprägungen auf den

Aspekten positives Klima und Management aufweisen. Vor allem eine hohe Ausprägung der ‚Sensitivity‘ der Lehrperson ist für das erste Schuljahr charakteristisch (vgl. Hamre & Pianta, 2007). Ähnlich hohe Ausprägungen der Klassenführung und der individuellen Unterstützung konnten auch Eckerth et al. (2012) für den schriftsprachlichen und mathematischen Unterricht im ersten Schuljahr belegen.

Befunde aus dem zweiten Schuljahr präsentieren sich eher uneinheitlich. Roßbach (2002a; b) zeigt beispielsweise im Rahmen der ECCE-Studie Unterschiede in der Höhe der Ausprägungen der beiden Merkmale Klassenmanagement und schülerorientiertes sozial-emotionales Klima. In der Studie schätzen externe Beobachter auf einer fünfstufigen Skala das Klassenmanagement im Mittel mit 3.6 ein, während sich für das schülerorientierte sozial-emotionale Klima eher ein niedriger Mittelwert von 2.3 ergibt. Im Vergleich zu der ECCE-Studie konnte im Kontext des FiS-Projekts ein signifikanter Anstieg der Werte der Klassenführung und der individuellen Unterstützung vom ersten zum zweiten Schuljahr bestätigt werden (vgl. Eckerth et al., 2012). Auch in der Studie von Helmke und Renkl (1993) lassen sich für beide beobachteten Unterrichtsvariablen ähnlich hohe Ausprägungen auf einer fünfstufigen Skala bestätigen.

Die dargestellten Befunde auf Basis von Beobachtungsstudien lassen folgende Rückschlüsse für die Analyse des Anfangsunterrichts mittels Videografie zu:

*Hypothese 1:*

Für das erste (Fach Deutsch) und zweite Schuljahr (Fach Mathematik) schätzen die Rater die Merkmale der Klassenführung und des Unterrichtsklimas im Mittel deutlich positiv ein.

Die Angaben über die mittlere Ausprägung der einzelnen Merkmale der Klassenführung und des Unterrichtsklimas werden im Rahmen der Arbeit durch Unterschiedsberechnungen für die beiden Videostudien ergänzt. Festzustellende Unterschiede werden mithilfe von Varianzanalysen auf ihre statistische Signifikanz geprüft. Diese dienen der Beantwortung der folgenden Frage:

*Forschungsfrage Nr. 5:*

Lassen sich im Hinblick auf die Ausprägungen der einzelnen Merkmale der Klassenführung und des Unterrichtsklimas signifikante Unterschiede zwischen den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) identifizieren?

Die Beantwortung der Forschungsfrage Nr. 5 erfolgt lediglich anhand einer kleineren Stichprobe von 30 Lehrpersonen, für die gleichzeitig Angaben für das Fach Deutsch (1. Schuljahr) und das Fach Mathematik (2. Schuljahr) vorliegen. Somit lassen sich signifikante Unterschiede zwischen den beiden Messzeitpunkten auf die unterschiedlichen Fächer / Schuljahre zurückfüh-

ren. Der Frage nach Unterschieden in der Höhe der Ausprägungen der Merkmale der Klassenführung und des Unterrichtsklimas in verschiedenen Fächern wurde bislang kaum nachgegangen. Dabei können Aspekte dieser beiden Basisdimensionen je nach Fach unterschiedliche Bedeutungen haben und somit eine unterschiedliche Rolle spielen. Im Kontext des FiS-Projekts zeigte sich jedoch kein Unterschied in der Höhe der Ausprägungen der Klassenführung und individuellen Unterstützung im schriftsprachlichen und mathematischen Anfangsunterricht (vgl. Eckerth et al., 2012). Inwieweit dieses Ergebnis auf Basis der PERLE-Daten bestätigt werden kann, soll explorativ überprüft werden.

### 7.3 Struktur der Unterrichtsqualitätsmerkmale (Studie 2)

Die Grundlage für die in der Studie 2 durchgeführten Analysen bildet die Faktorenanalyse als ein multivariates Analyseverfahren. Dies ist ein gängiges Verfahren in der Unterrichtsforschung, da somit Zusammenhänge zwischen Items (im vorliegenden Fall hoch inferent erfasste Merkmale) auf latente Faktoren zurückgeführt werden können (vgl. Bühner, 2006; Moosbrugger & Schermelleh-Engel, 2006). Bisherige Studien in der Sekundarstufe konnten auf Basis hoch inferenter Daten je nach Forschungsinteresse verschiedene Unterrichtsaspekte faktoranalytisch bestätigen (vgl. Clausen et al., 2003; Klieme et al., 2001; Lipowsky et al., 2009; Rakoczy, 2006). Aspekte der Klassenführung oder der direkten Instruktion sowie des Unterrichtsklimas / Schülerorientierung<sup>64</sup> wurden dabei übereinstimmend identifiziert.

Eine solche faktoranalytische Prüfung auf Grundlage hoch inferenter Daten fand m. W. n. bislang nicht für den Anfangsunterricht statt, da häufig lediglich einzelne Merkmale eines Konstrukts im Rahmen videobasierter Studien erhoben wurden (z. B. IGEL-Studie, vgl. Kapitel 2.2). Auf Basis von Daten der Unterrichtsbeobachtung konnten jedoch im Rahmen der Validierung des CLAS-Systems (vgl. Pianta et al., 2008; Hamre & Pianta, 2007) drei Faktoren der ‚Classroom Quality‘ bestätigt werden: ‚Emotional Support‘, ‚Classroom Organization‘ und ‚Instructional Support‘ (vgl. Kapitel 2.1). Da diese Faktoren den drei Basisdimensionen von Klieme et al. (2001) (vgl. Kapitel 1.2) inhaltlich stark ähneln, kann angenommen werden, dass die hier hoch inferent erfassten Merkmale der Klassenführung und des Unterrichtsklimas auch im Anfangsunterricht separate Konstrukte abbilden. Die Beantwortung der Forschungsfragen in Studie 2 erfolgt in zwei Analyseschritten:

---

64 Auf den Unterschied zwischen den Begriffen Unterrichtsklima und Schülerorientierung wurde bereits in Kapitel 4.1.1 eingegangen.

*Erster Analyseschritt:* Diese erste Analyse zielt darauf ab, die Struktur der Klassenführung und des Unterrichtsklimas für den Anfangsunterricht allgemein zu untersuchen. Das bedeutet, dass die Berechnung auf Basis aller 101 Videos (Deutsch / Mathematik) erfolgt. Ziel ist es, ein Modell zu finden, das die Struktur in den empirischen Daten für den Anfangsunterricht abbildet. Auf Grundlage der gesamten Stichprobe von Unterrichtsvideos erfolgt die Beantwortung der Forschungsfrage Nr. 6 im ersten Analyseschritt mithilfe von konfirmatorischen Faktoranalysen in *MPlus* (vgl. Muthén & Muthén, 1998–2007). Ausgehend von einem eindimensionalen Modell, in dem alle manifesten Merkmale der Klassenführung und des Unterrichtsklimas auf einem gemeinsamen latenten Generalfaktor laden (= eindimensionales Modell), wird ein 2-Faktor-Modell spezifiziert, von dem angenommen wird, dass dieses die Struktur der Klassenführung und des Unterrichtsklimas besser abbildet.

*Forschungsfrage Nr. 6:*

Lässt sich die theoretisch angenommene 2-Faktoren-Struktur bestätigen?

Auf Basis der theoretischen Herleitung der einzelnen Merkmale der Klassenführung und des Unterrichtsklimas kann davon ausgegangen werden (vgl. Kapitel 3.2 und 4.2), dass diese mindestens zwei differenzierbare Teilbereiche der Unterrichtsqualität abbilden (= zweidimensionales Modell). Aus diesem Grund wird folgende Hypothese aufgestellt:

*Hypothese 2:*

Ein zweidimensionales Modell passt besser auf die zugrunde liegenden, hoch inferenten Daten als ein eindimensionales Modell.

Bezüglich der manifesten hoch inferenten Merkmale der Klassenführung und des Unterrichtsklimas werden zusätzlich folgende Vorannahmen für das zweidimensionale Modell getroffen: Auf Grundlage der Theorie wird davon ausgegangen, dass ein Merkmal immer nur Ausdruck eines latenten Konstrukts ist. Des Weiteren wird bei dem zweidimensionalen Modell eine latente Korrelation der zwei Faktoren angenommen.

*Zweiter Analyseschritt:* Da die Merkmale der Klassenführung und des Unterrichtsklimas in zwei verschiedenen Fächern erhoben wurden, ergibt sich die Möglichkeit, die faktorielle Struktur in Abhängigkeit vom Unterrichtsfach und somit die fachbezogene Konstruktstabilität zu überprüfen. Um der Frage nachzugehen, ob sich zwischen den Fächern Unterschiede in der faktoriellen Struktur identifizieren lassen, werden in einem zweiten Analyseschritt fachspezifische explorative Faktorenanalysen mit SPSS Version 20 berechnet. Ein Modellvergleich mithilfe konfirmatorischer Faktorenanalysen in *MPlus* ist aufgrund der kleinen Stichprobe pro Fach nicht möglich (vgl. Kapitel 11). Es lässt sich in diesem Zusammenhang folgende Forschungsfrage formulieren:

*Forschungsfrage Nr. 7:*

Zeigen sich Unterschiede in der faktoriellen Struktur in Abhängigkeit vom Unterrichtsfach?

An dieser Stelle werden erneut keine Hypothesen formuliert. Inwieweit sich Unterschiede innerhalb der zwei Konstrukte zeigen lassen, wird explorativ überprüft.

## 7.4 Effekte der Unterrichtsqualitätsmerkmale auf multikriteriale Schülermerkmale (Studie 3)

Grundlage für die in der Studie 3 durchgeführten Analysen bildet das Kapitel 5 des Theorieteils. Wie die Ausführungen zeigen, berichten bisherige Studien in der Grundschule Ergebnisse zum Zusammenhang einer effektiven Klassenführung und eines positiven Unterrichtsklimas mit multikriterialen Schülermerkmalen häufig auf Basis korrelativer Analyseverfahren. Studien, die den Einfluss der Unterrichtsqualität mehrebenenanalytisch untersuchen, sind eher selten. Bevor die Fragestellungen zur prognostischen Validität der hoch inferenten Daten für die Leistungs- und Selbstkonzeptentwicklung formuliert werden, wird in einem ersten Schritt die Forschungsfrage zum Zusammenhang einer effektiven Klassenführung mit einem positiven Unterrichtsklima im Anfangsunterricht entwickelt.

*Forschungsfrage zum Zusammenhang zwischen Klassenführung und Unterrichtsklima:* Aufgrund der bisherigen Kenntnisse ist davon auszugehen, dass eine effektive Klassenführung eng mit dem Unterrichtsklima zusammenhängt. Dabei kann angenommen werden, dass sich auf Ebene der einzelnen Merkmale höhere Zusammenhänge zeigen als auf Skalenebene (vgl. Kapitel 1.2.2). Inwieweit Unterschiede in der Höhe des Zusammenhangs in Abhängigkeit vom Unterrichtsfach bestehen, wurde bislang nicht untersucht. Aus diesem Grund erfolgt diese Analyse explorativ. Im Kontext der Mehrebenenregressionsmodelle der Studie 3 sollen die folgenden Forschungsfragen beantwortet werden:

*Forschungsfrage Nr. 8.1:*

Zeigt sich der theoretisch angenommene Zusammenhang zwischen der Klassenführung und dem Unterrichtsklima auch empirisch?

*Forschungsfrage Nr. 8.2:*

Unterscheidet sich die Höhe des Zusammenhangs in Abhängigkeit vom Fach?

*Hypothese 3:*

Die Klassenführung hängt eng mit dem Unterrichtsklima zusammen.

Für die Forschungsfrage Nr. 8.2 wird erneut keine Hypothese formuliert. Auch diese Analyse erfolgt explorativ. Die Korrelation zwischen einer effektiven Klassenführung und einem positiven Unterrichtsklima sollte jedoch nicht größer als  $r = \pm .80$  sein (vgl. Urban & Mayerl, 2011), da sonst nicht mehr von zwei differenzierbaren Konstrukten ausgegangen werden kann. Damit wird das Problem der Multikollinearität der Unterrichtsmerkmale angesprochen. Je höher einzelne Merkmale miteinander korrelieren, desto eher führt es dazu, „dass die entsprechende Statistiksoftware gar nicht funktioniert oder zu sachlich unsinnigen Ergebnissen führt (etwa in der Weise, dass sich nur eines der Qualitätsmerkmale signifikant auf Leistung auswirkt und alle anderen Qualitätsmerkmale irrelevant zu sein scheinen)“ (Helmke, et al., 2007, S. 63).

*Klassenführung und Leistungen in Rechtschreiben und Mathematik:* Die in Kapitel 5.1 dargestellten Befunde lassen darauf schließen, dass in der Grundschule die Klassenführung, je nach Leistungsbereich und Schuljahr, unterschiedliche Effekte auf die Leistungsentwicklung der Schüler hat. So können Helmke und Weinert (1997b) im Rahmen der SCHOLASTIK-Studie für das Ende der Grundschulzeit positive Zusammenhänge einer effektiven Klassenführung mit den mathematischen Leistungen bestätigen. Für den Anfangsunterricht liegen bislang nur wenige Studien vor, die für das Fach Mathematik einen positiven Zusammenhang belegen (z. B. Cadima et al., 2010).

Für den Bereich Rechtschreiben kann anhand der dargestellten Befunden von Querschnittstudien im Kapitel 5.1.1 vermutet werden, dass die Rolle einer effektiven Klassenführung im Laufe der Grundschulzeit abnimmt und am Ende der Grundschulzeit kaum mehr eine Rolle spielt (vgl. Roos & Schöler, 2009; Helmke & Weinert, 1997b). So können Studien für das dritte und vierte Schuljahr keinen signifikanten Zusammenhang einer effektiven Klassenführung mit der Leistungsentwicklung im Rechtschreiben bestätigen (z. B. Helmke & Weinert, 1997b; Hofmann, 2007; Roos & Schöler, 2009), während sich beispielsweise im Kontext der Studie von Klicpera und Gasteiger-Klicpera (1993) für das erste und zweite Schuljahr positive Auswirkungen einer effektiven Klassenführung auf die Leistungen der Schüler im Rechtschreiben ergaben, wobei der Zusammenhang im zweiten Schuljahr bereits deutlich geringer ausfällt. Im Rahmen der Studie 3 soll folgenden Forschungsfragen nachgegangen werden:

*Forschungsfrage Nr. 9:*

Welche Bedeutung hat die Klassenführung für die Leistungen der Schüler in den Bereichen Rechtschreiben (1. Schuljahr) und Mathematik (2. Schuljahr)?

*Hypothese 4:*

Für den Bereich Rechtschreiben (1. Schuljahr) wird angenommen, dass die Klassenführung die Leistungen der Schüler am Ende des ersten Schuljahres positiv beeinflusst.

*Hypothese 5:*

Für den Bereich Mathematik (2. Schuljahr) wird angenommen, dass die Klassenführung die Leistungen der Schüler am Ende des zweiten Schuljahres positiv beeinflusst.

*Klassenführung und Selbstkonzepte im Schreiben und Mathematik:* Unter Berücksichtigung der Multikriterialität von Unterricht soll im Rahmen der vorliegenden Arbeit zusätzlich der Einfluss der Klassenführung auf die bereichsspezifischen Selbstkonzepte im Schreiben (1. Schuljahr) und Rechnen (2. Schuljahr) untersucht werden. Für die Sekundarstufe konnte bereits bestätigt werden, dass vor allem eine intensive bzw. effektive Zeitnutzung als Merkmal der Klassenführung und somit eine ausgeprägte Lehrstofforientierung zu einem günstigeren Selbstkonzept beiträgt (vgl. Helmke, 1992). Zudem bestätigen Studien aus dem Sekundarschulbereich, dass ein gewisses Maß an Disziplin sowie die Strukturiertheit des Unterrichts positive Auswirkungen auf das Selbstkonzept der Schüler haben (vgl. Kapitel 5.1). Für den Anfangsunterricht mangelt es bislang an Studien, die die Bedeutung einer effektiven Klassenführung für die Selbstkonzeptentwicklung der Schüler untersuchen. Da vor allem in den ersten Schuljahren das Selbstkonzept der Schüler noch wenig stabil ist – unabhängig von der Domäne, in der es erfasst wird – kann ein gewisses Maß an Regellosigkeit bzw. Chaos in der Klasse zu Verunsicherungen der noch jungen Schüler führen und über die Wahrnehmung der Schüler negative Auswirkungen auf die Entwicklung der Selbstkonzepte haben (vgl. Kapitel 5.1.2). Dabei stellt sich die Frage, ob Unterschiede in der Bedeutung der Effektivität der Klassenführung für die Selbstkonzepte in Abhängigkeit vom Fach angenommen werden können. Folgende Forschungsfrage Nr. 10 soll beantwortet werden:

*Forschungsfrage Nr. 10:*

Welche Bedeutung hat die Klassenführung für die Selbstkonzepte der Schüler in den Bereichen Schreiben (1. Schuljahr) und Rechnen (2. Schuljahr)?

Inwieweit sich Unterschiede in der Bedeutung einer effektiven Klassenführung für die fachspezifischen Selbstkonzepte im Schreiben und Rechnen zeigen lassen, wird explorativ untersucht. Aufgrund mangelnder Forschung in diesem Kontext werden erneut keine Hypothesen formuliert. Ausgehend von der Tatsache, dass das Selbstkonzept im frühen Grundschulalter noch nicht so eng mit den tatsächlichen Leistungen der Schüler zusammenhängt (z. B. Guay, Marsh & Boivin, 2003; Helmke, 1998; Kammermeyer & Martschinke,

2006), kann angenommen werden, dass die Organisation und Gestaltung des Unterrichts mehr Erklärungskraft für die Selbstkonzeptentwicklung haben als im Sekundarschulalter.

*Unterrichtsklima und Leistungen in Rechtschreiben und Mathematik:* Die Befundlage zur Bedeutung des Unterrichtsklimas für die Leistungen der Schüler in der Grundschule gestaltet sich eher uneinheitlich (vgl. Kapitel 5.2.1). Für den Anfangsunterricht mangelt es an Studien, die die Bedeutung des Unterrichtsklimas für die Leistungsentwicklung in Mathematik und Rechtschreiben untersuchen. Für den Bereich Rechtschreiben weisen die Befunde vom Ende der Grundschule (vgl. Kapitel 5.2.1) jedoch darauf hin, dass das Unterrichtsklima keinen Zusammenhang mit dem Lernzuwachs der Schüler im Rechtschreiben hat. May (2001) kommt für das zweite Schuljahr zum gleichen Ergebnis. Für den Bereich Mathematik zeigt die Studie von Pianta et al. (2008) im ersten und dritten Schuljahr tendenziell sogar negative Effekte des ‚Emotional Supports‘ auf die Leistungen der Schüler. Allerdings gestaltet sich auch für das Fach Mathematik die Befundlage eher uneinheitlich. Theoretisch kann jedoch angenommen werden, „dass ein Minimum an Wertschätzung und Respekt und eine positive Lernatmosphäre basale Voraussetzungen dafür sind, dass es überhaupt zu einer inhaltlich vertieften Auseinandersetzung mit dem Unterrichtsgegenstand kommen kann“ (Lipowsky, 2007b, S. 27). Inwieweit sich im Rahmen der vorliegenden Studie Unterschiede in der Bedeutung des Unterrichtsklimas für die Leistungsentwicklung der Schüler in Abhängigkeit vom Unterrichtsfach aufzeigen lassen, soll untersucht werden. In diesem Zusammenhang wird die folgende Forschungsfrage Nr. 11 formuliert:

*Forschungsfrage Nr. 11:*

Welche Bedeutung hat das Unterrichtsklima für die Leistungen der Schüler in den Bereichen Rechtschreiben (1. Schuljahr) und Mathematik (2. Schuljahr)?

*Hypothese 6:*

Für den Bereich Rechtschreiben wird angenommen, dass das Unterrichtsklima keinen Effekt auf die Leistungen der Schüler am Ende des ersten Schuljahres hat.

Aufgrund der sehr uneinheitlichen Befundlage für das Fach Mathematik und dem Mangel an Studien für den Anfangsunterricht erfolgt diese Analyse explorativ. Es werden für das Fach Mathematik keine Hypothesen formuliert.

*Unterrichtsklima und Selbstkonzepte im Schreiben und Mathematik:* Wie aus Kapitel 5.2.2 deutlich wurde, existieren bislang m. W. n. keine Studien im Anfangsunterricht, die die Bedeutung des Unterrichtsklimas für die Selbstkonzeptentwicklung der Schüler fachspezifisch untersuchten. So erfassten Perry et al. (2007) zwar verschiedene fachspezifische Fähigkeitsselbstkonzepte von Schülern des ersten Schuljahres, jedoch bekamen die Schüler für jede Domäne nur ein Item vorgelegt. Durch die Bildung einer Gesamtskala

blieb die Frage nach domänenspezifischen Wirkungen unterstützender Maßnahmen in der Studie offen. Perry et al. (2007) konnten jedoch bestätigen, dass Schüler in Klassen mit mehr unterstützenden Maßnahmen durch die Lehrperson, nach Kontrolle früherer Leistungen, ihre akademischen Fähigkeiten allgemein höher einschätzen.

Da allgemein angenommen wird, dass vor allem junge Schüler von einem positiven Unterrichtsklima bzw. von einer positiven Lehrer-Schüler-Beziehung profitieren (vgl. Kapitel 5.2.2), sollte sich unabhängig vom Fach ein positiver Effekt des Unterrichtsklimas auf die Selbstkonzepte der Schüler bestätigen lassen. Durch die geringe Stabilität des Selbstkonzepts in den ersten Schuljahren sind besonders junge Schüler anfälliger gegenüber „Risikoumwelten“ (Jerusalem & Schwarzer, 1991, S. 124) und somit stärker als im weiteren Verlauf der Schulzeit durch ein negatives Klima gefährdet. Im Umkehrschluss sollten Schüler in klimapositiven Klassen tendenziell eher positive Selbstkonzepte entwickeln. Folgende Forschungsfrage gilt es in diesem Kontext zu beantworten:

*Forschungsfrage Nr. 12:*

Welche Bedeutung hat das Unterrichtsklima für die Selbstkonzepte der Schüler in den Bereichen Schreiben (1. Schuljahr) und Rechnen (2. Schuljahr)?

*Hypothese 7:*

Es wird angenommen, dass das Unterrichtsklima die Selbstkonzepte der Schüler positiv beeinflusst. Dies gilt für die Bereiche Schreiben (1. Schuljahr) und Rechnen (2. Schuljahr) gleichermaßen.

### III Methodisches Vorgehen

Die folgenden Kapitel geben Auskunft über das methodische Vorgehen im Rahmen der vorliegenden Arbeit. Zunächst wird die PERLE-Studie (**P**ersönlichkeits- und **L**ernentwicklung von Grundschulkindern) vorgestellt, in die die hier dokumentierte empirische Untersuchung eingebettet ist (Kapitel 8). Anschließend werden die Untersuchungsdesigns der beiden Videostudien Deutsch (1. Schuljahr, Kapitel 9.1) und Mathematik (2. Schuljahr, Kapitel 9.2) dargestellt. Kapitel 10 umfasst in diesem Zusammenhang das Vorgehen bei der Aufzeichnung der Unterrichtsvideos. Die Stichprobenbeschreibung erfolgt im Kapitel 11. Das Ratingsystem für die videobasierte Erfassung der Unterrichtsqualitätsmerkmale Klassenführung und Unterrichtsklima wird in Kapitel 12.1 beschrieben, indem einerseits auf die Entwicklung des hoch inferenten Ratingsystems eingegangen wird (Kapitel 12.1.1) und andererseits der Aufbau beispielhaft beschrieben wird (Kapitel 12.1.2). Eine ausführliche Darstellung des hoch inferenten Ratingsystems erfolgt im Anhang (Kapitel 19). Welche Analyseeinheit dem hoch inferenten Rating zugrunde liegt und wie der Ratingprozess abgelaufen ist, wird in den Kapiteln 12.1.3 und 12.1.4 beschrieben. In Kapitel 12.2 werden die für die Studie relevanten Zielvariablen und deren Operationalisierungen im Rahmen der PERLE-Studie skizziert.

## 8 Kontext der Untersuchung – Die PERLE-Studie

Die Überprüfung der entwickelten Fragestellungen (vgl. Kapitel 7) erfolgt im Rahmen der PERLE-Studie. Die Studie wird gemeinsam von den Universitäten Kassel und Bamberg sowie dem Deutschen Institut für Internationale Pädagogische Forschung (DIPF) unter Leitung von Prof. Frank Lipowsky und Prof. Gabriele Faust (†) durchgeführt. Das Projekt wird vom Bundesministerium für Bildung und Forschung (BMBF) gefördert.

Ziel der kombinierten Querschnitt- und Längsschnittstudie ist es, die Persönlichkeits- und Lernentwicklung von Grundschulkindern über die gesamte Grundschulzeit zu erklären. Als Zielvariablen der Längsschnittstudie (PERLE I und II) wurden sowohl die Lernfortschritte in den Fächern Mathematik und Deutsch als auch personale Faktoren wie Kreativität, Motivation und Selbstkonzept über vier Schuljahre erfasst (vgl. Greb et al., 2007; Lipowsky et al., 2009). Während in PERLE I (Laufzeit des Längsschnitts 2006 bis 2008) der Fokus auf den Videostudien in den drei Fächern Deutsch, Kunst

und Mathematik und den Lehrerbefragungen zu unterrichtlichen Bedingungen lag (vgl. Abbildung 11), lag der Schwerpunkt der PERLE-II-Studie (Laufzeit 2008 bis 2010) zusätzlich zu der Analyse der individuellen Voraussetzungen der Schüler auf der Analyse der familiären Bedingungen, unter denen die Schüler lernen und sich entwickeln, sowie der qualitativen Studie.<sup>65</sup> Diese beinhaltete die Evaluation der speziellen Förderumwelt der sogenannten BIP-Kreativitätsgrundschulen in den neuen Bundesländern.

Mit den curricular vergleichbaren Videostudien wird in PERLE unter anderem das Ziel verfolgt, Unterrichtsqualitätsmerkmale in der Grundschule zu erfassen und Merkmale zu identifizieren, die die schülerspezifischen Entwicklungsverläufe beeinflussen (vgl. Greb et al., 2009a). Die PERLE-Videostudien knüpfen an Videostudien an, die bereits zahlreich für die Sekundarstufe durchgeführt wurden (vgl. Hiebert et al., 2003; Hugener, Pauli & Reusser, 2006; Petko, 2006; Seidel, Prenzel, Duit & Lehrke, 2003). Im Folgenden wird kurz auf die BIP-Kreativitätsgrundschulen und ihr Schulkonzept eingegangen (ausführlich bei Lipowsky, Faust, Kastens & Post, 2013).

*Die BIP-Kreativitätsgrundschulen:* Das Akronym BIP steht für Begabung, Intelligenz und Persönlichkeit. Bei den BIP-Kreativitätsgrundschulen handelt es sich um überwiegend in den neuen Bundesländern angesiedelte private Grundschulen, die auf die umfassende Förderung der Entwicklung von Begabung, Intelligenz, Persönlichkeit und Kreativität ausgerichtet sind. Die Basis hierfür bildet ein spezielles komplexes Schulkonzept (vgl. Mehlhorn, 2001; Mehlhorn & Mehlhorn, 2003). Das Bildungsangebot staatlicher Schulen wurde u. a. um künstlerisch-ästhetische, sprachliche und mathematisch-naturwissenschaftliche Inhalte ergänzt (vgl. Greb, Lipowsky & Faust, 2009a)<sup>66</sup>. Auf Ebene des Unterrichts zeichnen sich die BIP-Schulen durch klassen- und jahrgangsübergreifenden Unterricht aus. In den Fächern Deutsch und Mathematik werden die Klassen zudem nach unterschiedlichen Kriterien (z. B. Leistungsstärke) geteilt und getrennt von zwei Lehrpersonen unterrichtet (Teamenteaching). Diese Teilung ist allerdings nicht starr, sondern kann am Ende oder im Verlauf eines Schuljahres entsprechend der Leistungsentwicklung der Schüler angepasst werden (vgl. Lipowsky et al., 2013). Die Schüler können so individuell gefördert werden. Wie mit diesem ‚Teamenteaching‘ im Rahmen der vorliegenden Studie umgegangen wurde, wird an entsprechender Stelle erläutert. Zum Beispiel hatte dies teilweise Folgen für die Ratingprozedur (vgl. Kapitel 12.1.4). Im folgenden Kapitel wird auf das Design der Videostudien Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) näher eingegangen.

---

65 Das Design der PERLE-II-Studie (Laufzeit 2008–2010, Anfang drittes Schuljahr bis Ende viertes Schuljahr) wird an dieser Stelle nicht abgebildet, da diese Messzeitpunkte für die vorliegende Arbeit nicht von Bedeutung sind. Es kann aber unter folgendem Link abgerufen werden: <http://www.perle-projekt.de/> (Zugriff: 24.01.2013).

66 Die Spezifika der BIP-Schulen werden unter anderem von Lotz, Berner, Gabriel, Post, Faust und Lipowsky (2011; vgl. auch Lipowsky et al., 2013) detailliert beschrieben.

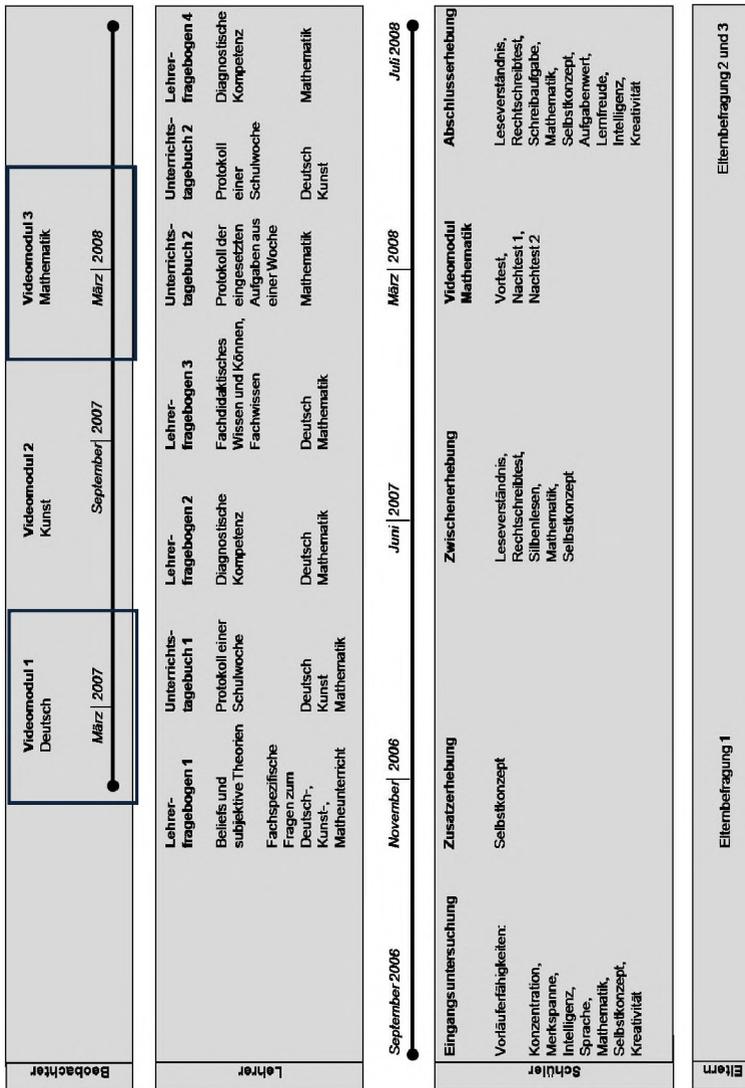


Abbildung 11: Design der PERLE-I-Studie (Laufzeit 2006 bis 2008, Einschulung bis Ende zweites Schuljahr).

*Anmerkung:* Die vorliegende Arbeit bezieht sich nur auf die Videostudien in Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr). Die Videostudie Kunst (1. Schuljahr) geht nicht in die Auswertungen mit ein.

## 9 Die PERLE-Videostudien Deutsch und Mathematik

Um die Vergleichbarkeit der einzelnen Lektionen in Deutsch und Mathematik zu gewährleisten, wurden im Vorfeld der Videoaufzeichnungen sowohl (1) die AufnahmeprozEDUREN durch ein Kameraskript (vgl. Kapitel 10) als auch (2) die Themen der Unterrichtsstunden festgelegt. Außerdem wurde den Lehrpersonen ein Zeitrahmen von 90 Minuten vorgegeben. Auf Basis dieser Vorgaben sollte von den Lehrpersonen ein für die Klasse möglichst typischer und alltäglicher Unterricht realisiert werden, wobei den Lehrpersonen für die weitere Gestaltung des Unterrichts keine weiteren Vorgaben gemacht wurden.

Die jeweiligen Inhalte der Videostudien Deutsch und Mathematik wurden in Absprache mit Fachdidaktikern und aufgrund einer Curriculumsanalyse festgelegt. Darüber hinaus wurde für die Videostudie Mathematik (2. Schuljahr) als didaktische Zielsetzungen der Stunde die Vorgabe gemacht, dass es sich um eine Einführungsstunde in die Multiplikation handeln sollte. Um ein umfangreiches Bild des Deutsch- und Mathematikunterrichts zu erhalten, wurden die Videoaufzeichnungen durch Lehrerinterviews ergänzt. Zusätzlich wurden in Abhängigkeit vom Fach verschiedene Leistungsdaten der Schüler erhoben.

Im Folgenden werden die Designs der Videostudien Deutsch und Mathematik beschrieben (vgl. Lotz, Berner, Gabriel, Post, Faust & Lipowsky, 2011).

### 9.1 Die Videostudie Deutsch – Vorgaben und Ablauf

Im März 2007 (erstes Schuljahr) fand die Videostudie im Fach Deutsch statt. In Abbildung 12 ist das Design der Videostudie Deutsch dargestellt. Zusätzlich zu der Aufzeichnung des Unterrichts wurden im Rahmen der Videostudie Deutsch sowohl Lehrer- und Schülerinterviews als auch ein Gruppentest (Gruppenbefragung) mit den Schülern durchgeführt.

Um zu gewährleisten, dass alle aufgezeichneten Unterrichtsstunden der Videostudie Deutsch eine gemeinsame thematische Basis haben, wurden den Lehrpersonen etwa zehn Tage vor der Videoaufzeichnung Planungsvorgaben für die 90-minütige Deutschstunde gemacht. Dabei wurden den Lehrpersonen drei obligatorische und eine fakultative Aufgabe vorgegeben. Die Reihenfolge und der zeitliche Umfang der einzelnen Aufgaben konnten von den Lehrpersonen selbst bestimmt werden.

Bei der ersten obligatorischen Aufgabe handelte es sich um die Vorstellung des ersten Teils des Bilderbuches „Lucy rettet Mama Krok“ (vgl. Doucet & Wilsdorf, 2005). Das Bilderbuch ist ein für das erste Schuljahr sehr an-

spruchsvolles Buch und handelt von einem Mädchen namens Lucy, das durch einen Sturm von ihren leiblichen Eltern getrennt und zu einer Krokodilfamilie „geweht“ wird. Lucy wird von der Krokodilmutter freundlich aufgenommen und wächst mit vielen Krokodilgeschwistern auf. Bald aber merkt sie, dass sie bestimmte Dinge nicht so gut beherrscht wie die Krokodile. Als sie deswegen gehänselt wird, beschließt Lucy, ihre „Adoptivfamilie“ zu verlassen (vgl. Lotz et al., 2011; Faust, Lipowsky & Gleich, 2011). Abgesehen von dem Thema Identität werden in dem Bilderbuch zwei weitere wichtige Themen behandelt: das Anderssein und die Ausgrenzung. In einer, auf die Vorstellung des ersten Buchkapitels folgenden, produktionsorientierten Anschlussaufgabe sollten die Schüler aus der Perspektive von Lucy einen Abschiedsbrief schreiben, den Lucy ihrer Krokodilmutter hinterlässt (zweite obligatorische Aufgabe, Schreibaufgabe).

Für die Durchführung der Leseübung (dritte obligatorische Vorgabe) bekamen die Lehrpersonen im Vorfeld drei Lesetexte zugeschickt, von denen sie entweder alle oder eine Auswahl einsetzen sollten. Die Lehrpersonen hatten jedoch die individuellen Freiheiten, eigene Texte zu ergänzen, wenn sie es für notwendig hielten. Für die fakultative Wortschatzübung wurde von PERLE lediglich vorgegeben, dass sie aus Wörtern des Bilderbuches bestehen sollte.

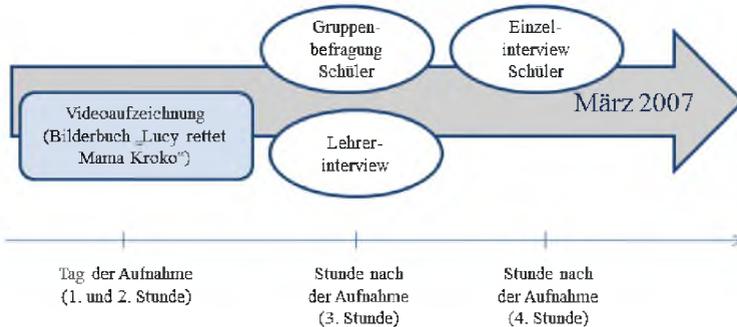


Abbildung 12: Design der Videostudie Deutsch (2. Halbjahr 1. Schuljahr, März 2007)

Direkt nach der aufgezeichneten Unterrichtsstunde fand ein Gruppentest mit den Schülern statt. Hierbei wurde in Form von bildlich dargestellten Multiple-Choice-Aufgaben das Textverständnis der Schüler überprüft (vgl. Abbildung 13). Der Gruppentest bestand aus sechs Fragen zum Bilderbuch (z. B. „Wie hat Lucy ihre Eltern verloren?“ oder „Was tut Lucy, als sie Mama Krok das erste Mal sieht?“). Die Schüler hatten die Möglichkeit, zwischen vier Antwortalternativen die richtige herauszufinden. Die Antwortalternativen waren in Form von Bildern dargestellt (vgl. Abbildung 13). Der Test wurde von geschulten Datenerhebenden unter standardisierten Bedingungen durchgeführt. In dem anschließenden Einzelinterview wurden die Schüler zu ihrer Meinung zum Buch („Hat dir die Geschichte gefallen?“), zum Textverständnis („Geht

Lucy weg, weil sie Angst vor Krokodilen hat?“) und zur Textinterpretation („War Lucy eine gute Krokodilsschwester?“) befragt (vgl. Lotz & Corvacho del Toro, 2013).

Parallel zur Gruppenbefragung der Schüler wurden die Lehrpersonen von einer PERLE-Mitarbeiterin interviewt. Im Fokus des Interviews standen Fragen zu den Zielen und Schwerpunkten der Unterrichtsstunde, zur Art der Vorbereitung sowie der eigenen Zufriedenheit mit der Unterrichtsgestaltung. Diese Interviews liegen in transkribierter Form vor.



Abbildung 13: Beispiel für das Antwortformat (Gruppenbefragung Schüler Videostudie Deutsch). Die Schüler mussten auf eine ihnen gestellte Frage das richtige Bild ankreuzen.

## 9.2 Die Videostudie Mathematik – Vorgaben und Ablauf

Im März 2008, im zweiten Halbjahr des zweiten Schuljahres, wurden die Unterrichtsvideos im Fach Mathematik aufgezeichnet. Aus Abbildung 14 wird deutlich, welche zusätzlichen Daten im Kontext der Videoaufzeichnungen erfasst wurden.

Den Lehrpersonen, die an der Videostudie Mathematik teilnahmen, wurde bereits im ersten Schulhalbjahr mitgeteilt, dass die zu videografierende Unterrichtseinheit die Multiplikation beinhalten sollte. Sie wurden daher gebeten, diese Einheit noch nicht zu unterrichten. Ungefähr zehn Tage vor der Aufnahme erhielten die Lehrpersonen die genaue Vorgabe, dass sie sowohl die Multiplikation als auch das Malzeichen und den Begriff „Malnehmen“ einführen sollen. Die Vorgaben konnten von den Lehrpersonen selbstständig

umgesetzt werden. Wie in der Videostudie Deutsch hatten die Lehrpersonen im Anschluss an die aufgezeichnete Stunde die Möglichkeit, in einem Leitfadeninterview Stellung zu ihrem Unterricht zu beziehen. Die Lehrpersonen wurden unter anderem zu ihren Lehrzielen und Unterrichtsvorbereitungen für die Doppelstunde befragt. Der Fokus des Lehrerinterviews lag hauptsächlich auf der Übungsphase. Abschließend schätzten die Lehrpersonen auch hier den aufgezeichneten Unterricht in Bezug auf seine Repräsentativität ein. Das Gespräch wurde auf Tonband aufgezeichnet und im Anschluss anonymisiert transkribiert.

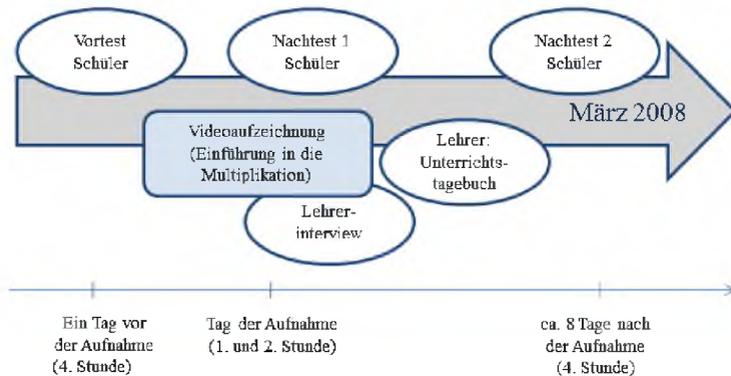


Abbildung 14: Design der Videostudie Mathematik (2. Halbjahr 2. Schuljahr, März 2008)

In Ergänzung zur Aufzeichnung des Unterrichts wurden im Rahmen der Videostudie Mathematik Leistungsdaten der Schüler zu drei Messzeitpunkten erhoben. Eine standardisierte Durchführung dieser Tests wurde durch geschulte Datenerheber gewährleistet. Einen Tag vor der Videoaufzeichnung wurde ein standardisierter Gruppentest mit den Schülern durchgeführt, der das Vorwissen zur Multiplikation erfasste. Innerhalb des sogenannten Vortests mussten die Schüler beispielsweise aus Plusaufgaben Malaufgaben bilden ( $4 + 4 + 4 = 3 \times 4$ ) oder Plus- und Malaufgaben lösen ( $5 + 5 + 5 = \dots$  oder  $5 \times 6 = \dots$ ). Insgesamt bearbeiteten die Schüler im Vortest 55 Aufgaben.

Am Tag der Videoaufzeichnung wurden im Rahmen des Nachtests 1 Aufgaben administriert, die primär das konzeptuelle Verständnis der Schüler überprüften. Etwa acht Tage nach Abschluss der Videoaufzeichnungen wurde erneut der Leistungsstand der Schüler mithilfe eines Nachtests 2 (Follow-up-Test) erhoben. Der Test war breiter angelegt als der Nachtest 1 und beinhaltete Aufgaben, die das konzeptionelle Verständnis, aber auch Rechenfertigkeiten erfassen. Der Nachtest 1 umfasste insgesamt 19 Aufgaben, der Nachtest 2 hingegen 47 Aufgaben.

Zusätzlich zu dem Leitfadeninterview sollten die Lehrpersonen zwischen dem Nachtest 1 und dem Nachtest 2 ein Unterrichtstagebuch führen, das Aus-

kunft zur Unterrichtsorganisation und zu den eingesetzten Aufgaben zwischen Nachtest 1 und Nachtest 2 sowie zur Vergabe von Hausaufgaben geben sollte.

Da die Videoaufzeichnungen den Kriterien einer systematischen Beobachtung Rechnung tragen sollten, wurden in den beiden Videostudien Deutsch und Mathematik nicht nur inhaltliche Vorgaben gemacht, sondern zusätzlich einheitliche Richtlinien für die Aufzeichnung des Unterrichts festgelegt. Diese standardisierten Verfahren zur Aufzeichnung sind notwendig, da – trotz der zahlreichen Vorteile – Unterrichtsvideos nicht in der Lage sind, sämtliche Unterrichtsprozesse zu dokumentieren (vgl. Petko et al., 2003; Seidel & Prenzel, 2003a). Zwar haben sie im Gegensatz zu Befragungen eine größere Prozessnähe, jedoch muss berücksichtigt werden, dass die entstandenen Unterrichtsvideos vor allem das Ergebnis einer bestimmten Kameraführung sowie der Positionierung der verwendeten Mikrofone sind und somit immer nur einen Teil der Unterrichtsrealität darstellen. Auch die Vorgaben, die den Lehrpersonen im Kontext der PERLE-Videostudien gemacht wurden, haben Einfluss auf die Strukturierung des aufgezeichneten Unterrichts. Dies wird im folgenden Kapitel kurz verdeutlicht. Dem schließt sich eine Beschreibung des Vorgehens bei der Videoerhebung und des eingesetzten Kameraskripts zur Standardisierung der Kameraprozeduren an (vgl. Kapitel 10).

### 9.3 Bedeutung der Vorgaben für den Unterrichtsverlauf

Durch die Vorgabe des Themas für die jeweilige Unterrichtsstunde in den Videostudien Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) ergeben sich unterschiedliche Zielsetzungen für die beiden Videostudien. Aus diesen Vorgaben resultieren unterschiedliche Unterrichtsverläufe bzw. Inszenierungsformen, die sich als Abfolge verschiedener Phasen im Unterricht widerspiegeln. In der Videostudie Deutsch impliziert die Vorgabe der drei obligatorischen Aufgaben beispielsweise die jeweilige Unterrichtsmethode bzw. die Sozialform. Die Vorstellung des Bilderbuchs erfolgt in einem Großteil der Stunden im öffentlichen Unterrichtsgespräch, während die Phase des Briefeschreibens eher zur Einzelarbeit anregt. Bei der Umsetzung der Leseübung hatten die Lehrpersonen mehr Freiheiten. Einerseits konnten die Schüler alleine, in Partner- oder Gruppenarbeit Lesen üben, andererseits konnte dies auch im Klassengespräch erfolgen. Erste Analysen zeigen jedoch, dass nur in einigen wenigen Klassen die Leseübung größtenteils im öffentlichen Unterrichtsgespräch stattfand, während ein weitaus größerer Teil der Lehrpersonen die Schüler alleine üben lässt (vgl. Lotz, 2010).

Auch die Einführung in die Multiplikation (Videostudie Mathematik) folgt einem festen Verlaufsplan, kann jedoch je nach Lehrperson unterschiedlich ausgestaltet und strukturiert werden. Im Zentrum steht die Erarbeitung

eines für die Schüler neuen Konzepts – die Multiplikation. Aus diesem Grund folgen Einführungsstunden einem anderen Verlaufsschema als beispielsweise nachfolgende Vertiefungs- bzw. Übungsstunden. Für die Sekundarstufe konnte eine große Spannweite an Inszenierungsmustern in Einführungs- und Übungslektionen aufgezeigt werden (vgl. Hugener & Krammer, 2010). In Anlehnung an Studien aus der Sekundarstufe (z. B. Pauli & Reusser, 2010) und unter Bezug auf Aebli (1994) erfolgt die Konstruierung eines neuen Konzepts in einem ersten Schritt durch die Aktivierung des Vorwissens der Schüler durch die Lehrperson. Dem schließt sich die Phase der Begriffsbildung an. Die Erarbeitung des neuen Stoffs anhand einer Problemstellung kann in der Grundschule fragend-entwickelnd eingeführt werden, d. h. das neue Konzept wird gemeinsam „im praktischen Umgang“ (Aebli, 1994, S. 265) entwickelt. Eine solche Phase ist stark lehrerzentriert. Dem gemeinsamen Entwickeln des neuen Begriffsinhalts folgt meist eine Phase, in der die Teilergebnisse festgehalten werden. Nachdem die neue mathematische Operation eingeführt und festgehalten wurde, sollte diese in einer Übungsphase vertieft werden, sodass die Schüler in selbstständiger Schülerarbeit ein vertieftes Verständnis von der Multiplikation erhalten. Für die Sekundarstufe konnten zusätzlich Inszenierungsmuster ermittelt werden, in denen der neue Stoff explorierend-entdeckend eingeführt wird (Pauli & Reusser, 2010). Ausgehend von einer Problemstellung erfolgt der Versuch der Schüler, in Partner- oder Gruppenarbeit das Problem zu lösen. Dem folgt typischerweise eine Phase, in der die von den Schüler entwickelten Lösungen im Klassengespräch besprochen oder diskutiert werden (ebd.).

## 10 Das Vorgehen bei der Videoerhebung

Um die Vergleichbarkeit der in PERLE aufgenommenen Unterrichtsvideos zu gewährleisten, bedurfte es im Vorfeld eines Kameraskripts sowie der Schulung des Kamerapersonals. In der PERLE-Studie erfolgte die Standardisierung der Kameraprozeduren in Anlehnung an die Pythagorasstudie (vgl. Hugener et al., 2006; Petko, 2006) unter Verwendung eines standardisierten Kameraskripts (vgl. Corvacho del Toro, Berner & Mösko, 2013). Mithilfe dieses Kameraskripts wurden einheitliche Richtlinien beschrieben, wie der Unterricht in der Grundschule aufgezeichnet werden soll. Das Kameraskript erläutert neben Fragen zur technischen Ausrüstung auch die Funktion und Handhabung der Audioquellen sowie der zwei eingesetzten Kameras und deren Führung. Anhand dieses Kameraskripts wurden die Datenerheber (bzw. das Filmpersonal) in einem mehrtägigen Training intensiv geschult.

Die Aufzeichnungen des Deutsch- und Mathematikunterrichts wurden von insgesamt zwei studentischen Hilfskräften und einer Mitarbeiterin des

PERLE-Teams durchgeführt. Für die Videoaufzeichnungen wurden zwei Kameras sowie vier Mikrofone pro Klassenzimmer eingesetzt. Die angestrebte Aufstellung der Kameras und Mikrofone ist in Abbildung 15 am Beispiel des Frontalunterrichts dargestellt.

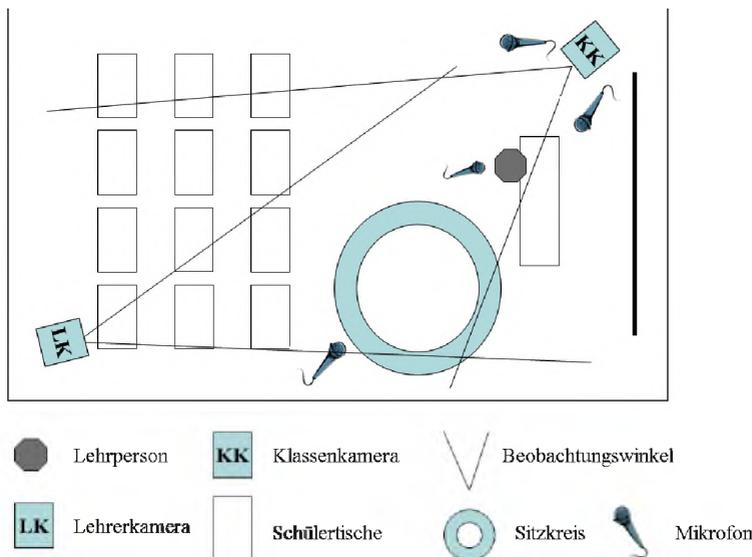


Abbildung 15: Ein Beispiel für den möglichen Aufbau der Kameras und Mikrofone im Klassenraum

Bei den beiden Kameras handelt es sich zum einen um eine *dynamische Lehrerkamera* (LK), die hauptsächlich die Interaktionen zwischen der Lehrperson und einzelnen Schülern erfasst, jedoch auch Interaktionen zwischen Lehrperson und Schülergruppen bzw. der gesamten Klasse (öffentliches Unterrichtsgespräch) aufzeichnen sollte. Diese Lehrerkamera wurde im hinteren Teil des Klassenzimmers positioniert. Zum anderen ist im Rahmen der PERLE-Videostudien eine *stationäre Klassenkamera* (KK) eingesetzt worden, die das gesamte Klassengeschehen überblicksartig fokussieren sollte. Mithilfe der Klassenkamera konnten weitestgehend alle Schüler aufgenommen werden. Diese Überblickskamera wurde in einer der vorderen Ecken des Klassenzimmers neben der Tafel platziert und konnte somit mögliche Informationsverluste der Lehrerkamera ausgleichen. Um möglichst viel über die beiden Kameraperspektiven vom Klassengeschehen einzufangen, waren beide Kameras mit einem Weitwinkelobjektiv ausgestattet. Weil bei der Positionierung der Klassen- und Lehrerkamera sowohl die Lichtverhältnisse als auch die Raumgestaltung (Anordnung der Tische, Klassenraumgröße etc.) ei-

ne Rolle spielten, wurde erst vor Ort entschieden, wie die zwei Kameras am besten positioniert werden.

Neben den integrierten *Mikrofonen* in den zwei Kameras wurden zusätzliche Mikrofone eingesetzt. Der Ton der Klassenkamera wurde durch zwei Stereomikrofone (neben der Klassenkamera, auf die Klasse gerichtet) verstärkt. Des Weiteren wurde auch die Lehrperson mit einem Funkmikrofon ausgestattet (Lavaliermikrofon). Da in der Grundschule häufig Sitzkreise durchgeführt werden, wurde auch der Bereich des Sitzkreises mit einem weiteren Klassenmikrofon ausgerüstet (vgl. Abbildung 15). Die Aufnahme der separaten Tonspuren wurde von einer weiteren geschulten studentischen Hilfskraft überwacht.<sup>67</sup>

Da in der vorliegenden Arbeit sowohl das Verhalten der Lehrperson als auch das der Schüler zur Beurteilung der Klassenführung und des Unterrichtsklimas relevant ist, wurden beide Kameraperspektiven für die Auswertung herangezogen. Für eine gleichzeitige Betrachtung beider Kameraperspektiven wurde auf das Programm Videograph (vgl. Rimmel, 2002) zurückgegriffen (vgl. Kapitel 12.1.4).

Durch die hier beschriebene Aufnahme-prozedur entstanden insgesamt 101 Unterrichtsvideos für die Fächer Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr). Wie sich die Lehrer- und Schülerstichprobe in den beiden Videostudien zusammensetzt, ist Gegenstand des nachfolgenden Kapitels.

## 11 Stichprobenbeschreibung

Insgesamt setzt sich die Stichprobe aus 37 Schulklassen aus den neuen Bundesländern zusammen (Sachsen, Berlin, Mecklenburg-Vorpommern und Thüringen). Neben Klassen von staatlichen Grundschulen wurden zusätzlich Klassen aus den sogenannten BIP-Kreativitätsgrundschulen in die Stichprobe aufgenommen (vgl. Kapitel 8). Die Auswahl der staatlichen Schulen erfolgte nicht zufällig. Da ein Ziel der PERLE-Studie die Evaluation der BIP-Schulen ist, erfolgte vom Sächsischen Staatsministerium für Kultus bzw. den zuständigen Schulverwaltungsbehörden eine Vorauswahl an staatlichen Grundschulen, die zu einer Teilnahme an der Studie motiviert wurden. Demnach handelt es sich bei der PERLE-Stichprobe um eine anfallende Klumpenstichprobe und nicht um eine Zufallsstichprobe, da ganze Schulklassen einbezogen wurden und nicht zufällig gewählte Schüler bzw. Lehrpersonen. Die Auswahl der Schulen und damit auch Klassen erfolgte nur in den neuen Bundeslän-

---

<sup>67</sup> In den BIP-Klassen wurde für den Teilungsunterricht dasselbe Equipment im zweiten Klassenraum aufgebaut, um die Interaktionen der zweiten Lehrpersonen mit der zweiten Hälfte der Schüler aufzunehmen.

dern. Demnach sind die ausgewählten Teilmengen bereits vorgruppiert (vgl. Bortz, 1999).

Zur Verdeutlichung der Datengrundlage gibt die Tabelle 10 einen Überblick über die Anzahl der teilnehmenden Schulen und Klassen sowie die der vorliegenden Unterrichtsvideos. Diese variieren zwischen den Videostudien Deutsch und Mathematik. Die Darstellung der Kennwerte der Lehrer- und Schülerstichprobe erfolgt anschließend (vgl. Kapitel 11.1 und 11.2). An der Videostudie Deutsch nahmen insgesamt 17 BIP-Klassen und 20 Klassen aus staatlichen Grundschulen teil (vgl. Tabelle 10). Insgesamt verteilen sich die hier einbezogenen Klassen auf 13 staatliche Grundschulen und sieben BIP-Kreativitätsgrundschulen. An der Videostudie Mathematik partizipierten insgesamt 17 BIP-Klassen und 19 Klassen aus staatlichen Grundschulen, wobei diese aus elf staatlichen Grundschulen und sieben BIP-Kreativitätsgrundschulen stammen. In Tabelle 10 ist die Anzahl der teilnehmenden Schüler der Videostudien Deutsch (VSS, 1. Schuljahr) und Mathematik (VSM, 2. Schuljahr) dargestellt. Die durchschnittliche Anzahl der Schüler pro Klasse variiert zwischen den zwei Videostudien (VSS:  $\bar{X}$  = 16.2 BIP-Schüler / 18.9 Schüler aus staatlichen Grundschulen; VSM:  $\bar{X}$  = 16.0 BIP-Schüler / 20.3 Schüler aus staatlichen Grundschulen).

Tabelle 10: Verteilung der teilnehmenden Schulen / Klassen und Anzahl der Unterrichtsvideos nach Videostudie

	BIP-Kreativitätsgrundschulen		Staatliche Grundschulen		Gesamt	
	VSS	VSM	VSS	VSM	VSS	VSM
Anzahl Schulen	7	7	13	11	20	18
Anzahl Klassen	17	17	20	19	37	36
Anzahl Klassen mit KT (gesamte Zeit oder zeitweise)	13	12	–	–	–	–
Anzahl Klassen mit KKT	4	5	–	–	–	–
$\bar{X}$ Anzahl Schüler pro Klasse	16.2	16.0	18.9	20.3	17.7	18.3
Unterrichtsvideos	30	32	20	19	50	51

*Anmerkung:* VSS = Videostudie Deutsch / Sprache, VSM = Videostudie Mathematik,  $\bar{X}$  = Durchschnitt, KT = Klassenteilung, KKT = keine Klassenteilung

Für die Videostudie Deutsch (1. Schuljahr) liegen 50 Unterrichtsvideos und für die Videostudie Mathematik (2. Schuljahr) 51 Unterrichtsvideos vor. Wie aus Tabelle 10 ersichtlich wird, deckt sich diese Anzahl von Unterrichtsvideos nicht mit der Anzahl an Klassen, da für die BIP-Klassen aufgrund der Klassenteilung zwei Unterrichtsvideos pro Klasse vorliegen (vgl. Kapitel 12.1.4). In 13 Klassen, die an der Videostudie Deutsch teilgenommen haben, wurde die Klasse für die gesamte Zeit oder zeitweise in zwei Gruppen aufgeteilt. In

vier Klassen fand keine Klassenteilung statt. In der Videostudie Mathematik wurden zwölf teilnehmende Klassen für die gesamte Zeit oder zeitweise in zwei Gruppen aufgeteilt, während in fünf Klassen keine Klassenteilung stattfand. In den staatlichen Grundschulen liegt pro Klasse und pro Videostudie nur ein Unterrichtsvideo vor. Aus diesem Grund sind in Tabelle 10 für die staatlichen Schulen keine Werte in dieser Zeile angegeben.

Da für die vorliegende Arbeit sowohl die Videodaten der Lehrpersonen als auch Leistungs- und Fragebogendaten der Schüler herangezogen werden, erfolgt eine getrennte Beschreibung der Stichproben. Auch bei der Beschreibung der Kennwerte für die beiden Stichproben wird getrennt für die Fächer Deutsch und Mathematik vorgegangen. Im Anschluss wird auf die Stichprobe der teilnehmenden Lehrpersonen detaillierter eingegangen (vgl. Kapitel 11.1). Danach folgt die Beschreibung der Schülerstichprobe (vgl. Kapitel 11.2).

## 11.1 Die Lehrerstichprobe

Insgesamt nahmen an den beiden Videostudien (Deutsch, 1. SJ. und Mathematik, 2. SJ) 66 Lehrpersonen teil, wobei 30 Lehrpersonen sowohl an der Videostudie Deutsch als auch an der Videostudie Mathematik teilgenommen haben. Für beide Videostudien ergibt sich ein  $N$  von 50 teilnehmenden Lehrpersonen<sup>68</sup> (vgl. Tabelle 11). Sowohl die Stichprobe der Videostudie Deutsch als auch die der Videostudie Mathematik setzt sich aus BIP-Lehrpersonen und aus Lehrpersonen staatlicher Grundschulen zusammen. An der Videostudie Deutsch nahmen 30 BIP-Lehrpersonen<sup>69</sup> und 20 Lehrpersonen von staatlichen Grundschulen teil, während sich in der Videostudie Mathematik 31 BIP-Lehrpersonen und 19 staatliche Grundschullehrpersonen videografieren ließen.

Da es sich bei PERLE um eine Evaluationsstudie des BIP-Schulkonzepts handelt, waren fast alle BIP-Lehrpersonen des betreffenden Jahrgangs an der Studie beteiligt. Die Lehrpersonen der staatlichen Grundschulen haben hingegen freiwillig an der PERLE-Studie (einschließlich Videostudien) teilgenommen. Aus diesem Grund muss bei der Analyse der Unterrichtsstunden

---

68 In der Videostudie Mathematik liegen für eine Lehrperson zwei Unterrichtsvideos vor, da sie in zwei unterschiedlichen Klassen die Multiplikation eingeführt hat. In der anschließenden Stichprobenbeschreibung wird diese Lehrperson jedoch nur einmal berücksichtigt. Aus diesem Grund weicht die Anzahl der Unterrichtsvideos von der Anzahl der teilnehmenden Lehrpersonen in der Videostudie Mathematik ab (vgl. Tabelle 10).

69 Die Anzahl der BIP-Lehrpersonen liegt deutlich über der Anzahl an staatlichen Grundschullehrern, da in einem Großteil der BIP-Klassen zwei Lehrpersonen während des aufgezeichneten Unterrichts (sowohl Deutsch als auch Mathematik) anwesend sind. In jeder dieser doppelt besetzten Klasse gibt es sowohl eine Hauptlehrperson als auch eine Nebenlehrperson. Welche der zwei Lehrpersonen hauptverantwortlich ist, wurde durch eine schriftliche Nachfrage (Anfang 2009) in den Schulen erfasst.

beachtet werden, dass es sich bei den staatlichen Lehrpersonen mit hoher Wahrscheinlichkeit um eher engagierte Lehrpersonen handelt. Zudem handelt es sich bei den teilnehmenden Lehrpersonen um eine geschachtelte Stichprobe, da 30 von ihnen sowohl an der Videostudie Deutsch (1. Schuljahr) als auch an der Videostudie Mathematik (2. Schuljahr) teilgenommen haben.

Zur näheren Beschreibung der teilnehmenden Lehrpersonen wird der Lehrerfragebogen 1 (LFB 1) als Datengrundlage herangezogen. Dieser wurde im November 2006 allen Lehrern vorgelegt, d. h. ungefähr vier Monate vor der Videostudie Deutsch und circa einhalb Jahre vor der Videostudie Mathematik (vgl. Kapitel 8, Abbildung 11). Unabhängig davon, ob die Lehrpersonen Deutsch, Kunst oder Mathematik unterrichteten, beantworteten die Lehrpersonen zu Beginn des Fragebogens allgemeine fachunabhängige Fragen. Beispielsweise wurden die Lehrpersonen nach ihrem Geschlecht und Alter sowie nach ihrer Berufserfahrung und Ausbildung gefragt. Zusätzlich bekamen die Lehrpersonen im LFB 1 drei fachspezifische Teile für die Fächer Deutsch, Bildende Kunst und Mathematik vorgelegt, die sie, je nachdem ob sie das jeweilige Fach unterrichteten, ausgefüllt haben (vgl. Poloczec, Greb, Faust & Lipowsky, 2009a). Für die Beschreibung der Stichprobe sind nur die Angaben aus dem allgemeinen fachunabhängigen Teil von Bedeutung. Allerdings liegen nicht für alle an den Videostudien teilnehmenden Lehrpersonen Angaben aus dem LFB 1 vor. Für diejenigen Lehrpersonen, die an der Videostudie Deutsch teilgenommen haben, liegen nur von 43 diese Angaben vor. Von den 50 teilnehmenden Lehrpersonen der Videostudie Mathematik füllten nur 42 den LFB 1 aus. Die jeweiligen Kennwerte der Lehrpersonen sind in Tabelle 11 separat für jede Videostudie dargestellt.

*Geschlechterverteilung:* Wie aus Tabelle 11 ersichtlich wird, herrscht bei der Geschlechterverteilung ein Ungleichgewicht zugunsten der weiblichen Lehrpersonen. An der PERLE-Studie beteiligten sich insgesamt mehr Lehrerinnen (95,2 %) als Lehrer (4,3 %).<sup>70</sup> An der Videostudie Deutsch nahmen 42 Lehrerinnen und 1 Lehrer und an der Videostudie Mathematik 38 Lehrerinnen und 4 Lehrer teil. Für die übrigen Lehrpersonen der Videostudie Deutsch und der Videostudie Mathematik konnten auf Grundlage des LFB 1 keine Angaben zum Geschlecht gemacht werden. Da jedoch in den Unterrichtsvideos die Geschlechter der Lehrpersonen eindeutig identifizierbar waren, wurde für die fehlenden Lehrpersonen das Geschlecht in Tabelle 11 ergänzt. Auf Grund-

---

70 Da der Männeranteil an Grundschulen jedoch im Vergleich zu anderen Schulformen generell deutlich geringer ist, kann diese Verteilung durchaus als repräsentativ für die Grundschule angesehen werden. Laut dem Statistischen Bundesamt sind fast 86 % der Lehrpersonen an Grundschulen Lehrerinnen (Statistisches Bundesamt, 2004; <http://www.bmfsfj.de/Publikationen/genderreport/1-Bildung-ausbildung-und-weiterbildung/1-4-Schulische-bildung/1-4-4-lehrkraefte.html>, Zugriff: 5.12.2012). Ähnliche Verteilungen zeigen sich auch im Rahmen der Internationalen Grundschul-Lese-Untersuchung (IGLU; Lankes, Bos, Mohr, Plafmeier & Schwippert, 2003). Der Anteil der Grundschüler, die von weiblichen Lehrpersonen unterrichtet werden, liegt in Deutschland bei 82 Prozent (ebd.).

lage der Unterrichtsvideos konnte im Nachhinein festgestellt werden, dass es sich bei allen fehlenden Fällen ( $N_{VSS+VSM} = 15$ ) um Lehrerinnen handelte.

Tabelle 11: Kennwerte für die Lehrerstichprobe je nach Videostudie

		VSS (1. SJ)	VSM (2. SJ)
$N_{\text{Gesamt}}$		50	50
$N_{\text{BiP}}$		30	31
$N_{\text{staatl.}}$		20	19
Berufserfahrung (in Jahren)	MW / SD	15.17 / 10.98	15.07 / 10.61
	Min / Max	1.0 / 35.0	0.5 / 35.0
	fehlend	11	13
Alter (in Jahren)	bis 35	6	7
	35–45	26	23
	46–55	9	11
	über 55	2	1
	fehlend	7	8
Geschlecht*	♀	49	46
	♂	1	4

Anmerkung: VSS = Videostudie Deutsch / Sprache, VSM = Videostudie Mathematik, SJ = Schuljahr, MW = Mittelwert, SD = Standardabweichung, Min = Minimum, Max = Maximum, ♀ = Lehrerinnen, ♂ = Lehrer

\* Fehlende Angaben zum Geschlecht der Lehrpersonen wurden anschließend mithilfe der Unterrichtsvideos ermittelt. Ähnlichkeiten zwischen den Lehrpersonen der Videostudie Deutsch und Mathematik resultieren aus den 30 Lehrern, die an beiden Videostudien teilgenommen haben.

*Altersverteilung:* Neben der Benennung des Geschlechts sollten die Lehrpersonen Angaben zu ihrem Alter machen. Die Frage „Wie alt sind Sie?“ beantworteten 43 Lehrpersonen der Videostudie Deutsch und 42 Lehrpersonen, die an der Videostudie Mathematik teilgenommen haben. Die Altersverteilung ist in Tabelle 11 dargestellt. Hierbei wird deutlich, dass die Lehrpersonen in beiden Videostudien überwiegend auf eine Altersgruppe verteilt sind. Sowohl die Lehrpersonen der Videostudie Deutsch als auch die Lehrer der Videostudie Mathematik sind zwischen 35 bis 45 Jahre alt ( $N_{VSS} = 26/60.5\%$ ;  $N_{VSM} = 23/54.8\%$ ). Sechs Lehrpersonen der Videostudie Deutsch sind unter 35 Jahre (14%), während die übrigen Lehrpersonen 46 Jahre oder älter sind ( $N = 11/25.6\%$ ). Im Fach Mathematik gaben sieben Lehrpersonen an, dass sie unter 35 Jahre (16.7%) sind, die restlichen Lehrpersonen befinden sich hingegen in den Altersgruppen 46–55 Jahre und älter als 55 Jahre ( $N = 12/28.6\%$ ).

*Berufserfahrung:* Zusätzlich zum Alter machten die Lehrpersonen im LFB 1 Angaben zu ihrer Berufserfahrung. Die Frage „Wie viele Jahre sind

Sie bereits als Lehrer nach Abschluss Ihrer Ausbildung (ggf. nach dem Referendariat) tätig?“ beantworteten 39 Lehrpersonen der Videostudie Deutsch und 37 Lehrpersonen der Videostudie Mathematik. Die Lehrer der Videostudie Deutsch haben im Durchschnitt 15.17 Jahre Berufserfahrung, während die Lehrer, die an der Videostudie Mathematik teilgenommen haben, durchschnittlich 15.07 Jahre in ihrem Beruf tätig sind.

## 11.2 Die Schülerstichprobe

Im Allgemeinen kommen die Schüler der PERLE-Stichprobe aus Familien mit einem mittleren bis hohen sozioökonomischen Status (vgl. Gabriel et al., 2010). Als Indikator für den sozioökonomischen Status wurde der HISEI (Highest International Socio Economic Index) herangezogen. Der HISEI ( $N = 422$ ) beträgt im Mittel 62.7 Punkte ( $SD = 15.4$ ). Damit liegt der HISEI in der PERLE-Stichprobe 0.85 Standardabweichungen höher als in der deutschen PISA-2003-Stichprobe ( $MW = 49.2$ ;  $SD = 15.9$ ; vgl. Ehmke, Siegle & Hohensee, 2005). Es handelt sich demnach auch bei der Schülerstichprobe um keine repräsentative Stichprobe. Viele Eltern verfügen über einen akademischen Bildungsabschluss.

Wie zur Beschreibung der Lehrerstichprobe wird auch zur Beschreibung der Schülerstichprobe fachspezifisch vorgegangen. Im Fach Deutsch umfasst die Stichprobe 874 Schüler. Von diesen 874 Schülern besuchten 367 (42.0 %) eine BIP-Kreativitätsgrundschule, 507 (58.0 %) gingen auf eine staatliche Grundschule (vgl. Tabelle 12). 418 der Schüler sind männlichen Geschlechts, die Gruppe der Schülerinnen ist etwas größer ( $N = 452$ ). Dies entspricht einer prozentualen Verteilung von 48 % Schülern und 52 % Schülerinnen. Im Durchschnitt waren zum Zeitpunkt der Videostudie Deutsch die Schüler sieben Jahre und einen Monat alt ( $SD = .46$ ), wobei der jüngste knapp über fünf Jahre und der älteste knapp neun Jahre alt war. Das Durchschnittsalter der BIP-Schüler liegt bei fast sieben Jahren ( $SD = .47$ ), die staatlichen Schüler sind im Mittel sieben Jahre und zwei Monate ( $SD = .42$ ). Der jüngste staatliche Grundschüler war zum Zeitpunkt der Videostudie Deutsch fünf Jahre und acht Monate alt, während der älteste Schüler der staatlichen Stichprobe acht Jahre und acht Monate alt war.

Die Schülerstichprobe im Fach Mathematik setzt sich aus insgesamt 772 Schülern zusammen, wobei es sich um 401 Schülerinnen (52.0 %) und 370 Schüler (48.0 %) handelt. Von einem Schüler bzw. einer Schülerin liegen keine Angaben zum Geschlecht vor. Insgesamt besuchen mehr Schüler eine staatliche Grundschule ( $N = 442$ , ♂ = 209, ♀ = 233), während 330 (♂ = 161, ♀ = 168) Schüler auf eine BIP-Kreativitätsgrundschule gehen. Zum Zeitpunkt der Videostudie Mathematik, die genau ein Jahr nach der Videostudie

Deutsch im März 2008 stattfand, waren die Schüler im Durchschnitt acht Jahre und einen Monat alt ( $SD = .42$ ).

Tabelle 12: Kennwerte der Schülerstichprobe nach Schultyp und Fach

	BIP-Kreativitäts- grundschulen		Staatliche Grundschulen		Gesamt	
	Deutsch	Mathe- matik	Deutsch	Mathe- matik	Deutsch	Mathe- matik
Schülerzahl	367	330	507	442	874	772
Geschlecht*	♂ = 177 ♀ = 188	♂ = 161 ♀ = 168	♂ = 241 ♀ = 264	♂ = 209 ♀ = 233	♂ = 418 ♀ = 452	♂ = 370 ♀ = 401
Ø Alter (Jahre; Mo- nate) / $SD$	6; 9 / .47	8; 0 / .46	7; 2 / .42	8; 1 / .38	7; 1 / .46	8; 1 / .42
Min / Max (Jahre; Monate)**	5; 3 / 9; 3	6; 3 / 10; 3	5; 8 / 8; 8	6; 9 / 9; 6	5; 3 / 9; 3	6; 3 / 10; 3

Anmerkung: ♂ = Jungen, ♀ = Mädchen, Ø = Durchschnitt, Min = Minimum, Max = Maximum

\* Fehlende Angaben zum Geschlecht von einem BIP-Schüler / einer BIP-Schülerin im Fach Mathematik. Im Fach Deutsch liegen für vier Schüler (jeweils zwei staatliche und zwei BIP-Schüler) keine Angaben zum Geschlecht vor.

\*\* Im Fach Deutsch liegen für neun Schüler (vier BIP- / fünf staatliche Schüler) und im Fach Mathematik für vier Schüler (zwei BIP- / zwei staatliche Schüler) keine Angaben zum Alter vor.

### 11.3 Repräsentativität des gezeigten Verhaltens in den Videostudien

Die Anwesenheit von Kameras während der Stunde kann die Unterrichts- bzw. Beobachtungssituation beeinflussen. Sowohl Lehrpersonen als auch Schüler können aufgeregt sein und sich dadurch anders verhalten als im alltäglichen Unterricht. Aus diesem Grund wird im folgenden Kapitel kurz auf die möglichen Effekte der Erhebungssituation und die Repräsentativität des in den Videostudien gezeigten Verhaltens eingegangen.

Wenn über die Repräsentativität des Lehrer- und Schülerverhaltens bei Videostudien gesprochen wird, fallen häufig die Begriffe Kameraeffekt oder / und Invasivität (vgl. Petko et al., 2003; Hodel, Waldis, Gautschi & Reusser, 2006). Der Begriff invasiv (von lat. *invadere* = einfallen, eindringen) wird dabei oft im Zusammenhang mit der Rolle des Kamerateams im Klassenraum verwendet. Diese haben während des Filmens die Aufgabe, das alltägliche Unterrichtsgeschehen nicht durch ihre Anwesenheit oder durch das Aufstellen des Kameraequipments zu stören. Solche Invasivitätsmindernden Maßnahmen sind wichtig, da die videobasierte Forschung den Anspruch erhebt,

einen alltäglichen Unterricht zu filmen. Dies wird unter anderem zusätzlich dadurch ermöglicht, dass die Aufzeichnungen im Klassenraum stattfinden und somit die natürliche Lernumgebung beibehalten wird.

Wie bereits Petko et al. (2003) oder auch Kleinbub (2010) berichten, scheint das Problem des Kameraeffekts insgesamt weniger gewichtig zu sein, als man vermuten könnte. Ein Großteil der Lehrpersonen, die beispielsweise an der TIMSS- oder VERA-Videostudie teilgenommen haben, gab an, dass die gefilmten Stunden nahezu alltägliche Unterrichtssituationen widerspiegeln. Auch ein Großteil des Videomaterials aus PERLE kann als repräsentativ für den alltäglichen Grundschulunterricht der teilnehmenden Lehrpersonen angesehen werden. Auf die Frage „War das eine typische Deutsch- / Mathestunde für Sie?“ antwortete der weitaus größte Teil der teilnehmenden Lehrpersonen im Rahmen des Lehrerinterviews<sup>71</sup> mit „ja“ (VSS = 84.8 %; VSM = 80.4 %; vgl. Abbildung 16 und Abbildung 17).<sup>72</sup>

Dies war jedoch auch zu erwarten, da das Unterrichtshandeln von Lehrpersonen vielfach von Handlungsrouinen (unbewusst) abhängig und somit stark änderungsresistent ist (vgl. Hodel et al., 2006). Eine Lehrperson kann sich zwar für die aufgezeichnete Unterrichtsstunde sorgfältiger als sonst vorbereiten, viele Aspekte der Unterrichtsgestaltung, wie beispielsweise die Lehrer-Schüler-Beziehung, die unter anderem im Rahmen der vorliegenden Arbeit untersucht wird, sind jedoch einmalig schwer veränderbar (vgl. Stigler et al., 1999). Neben Aspekten des Unterrichtsklimas kann auch die Klassenführung als ein Unterrichtsmerkmal angesehen werden, das in hohem Maße Automatismen folgt. Der Gestaltung einer Unterrichtsstunde liegt beispielsweise die Einführung eines klaren Regelsystems zugrunde. Das Fehlen eines solchen Regelsystems kann innerhalb einer aufgezeichneten Lektion nicht ausgeglichen werden und schlägt sich in der Effektivität der Klassenführung nieder.

---

71 Die Lehrerinterviews wurden in beiden Videostudien im Anschluss an die videografierten Unterrichtsstunden durchgeführt (vgl. Kapitel 9). Im Fach Deutsch (1. Schuljahr) liegen Angaben von 46 Lehrpersonen vor, vier Lehrpersonen nahmen an dem Lehrerinterview nicht teil. Somit weicht die Anzahl der Lehrpersonen im Fach Deutsch (1. Schuljahr) von den ursprünglich 50 teilnehmenden Lehrpersonen ab (vgl. Kapitel 11.1). In Mathematik (2. Schuljahr) wurden alle 51 Lehrpersonen interviewt.

72 Im Kontext der vorliegenden Arbeit wurden in Anlehnung an bisherige Videostudien die Selbstauskünfte der Lehrpersonen zur Repräsentativität der aufgezeichneten Stunde als Indikator herangezogen, um von der videografierten Unterrichtsstunde und der dort beobachteten Unterrichtsqualität auf die allgemeine Ausprägung der Unterrichtsqualität im jeweiligen Fach schließen zu können. Allerdings ist es fraglich, ob ausgerechnet Lehrpersonen, dessen Selbstauskünfte häufig selbst verzerrt sind – in Bezug auf den eigenen Unterricht spielt hier vor allem die soziale Erwünschtheit eine große Rolle – die Repräsentativität ihrer Stunde adäquat beurteilen können. Schließlich wird häufig argumentiert, dass Videostudien durchgeführt werden, um Verzerrungen durch Selbstauskünfte der Lehrpersonen zu umgehen.

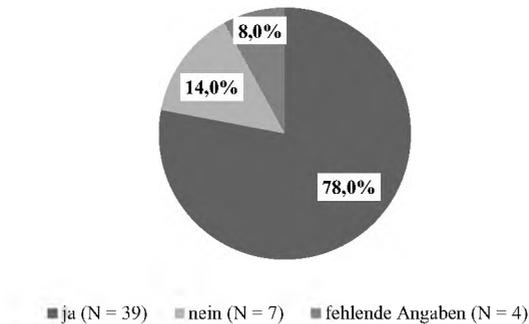


Abbildung 16: Repräsentativität der Unterrichtsstunden in Deutsch (VSS, 1. Schuljahr)

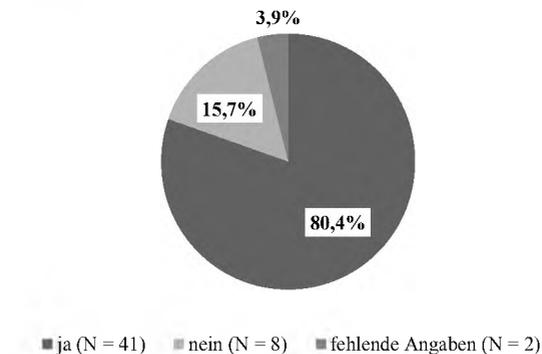


Abbildung 17: Repräsentativität der Unterrichtsstunden in Mathematik (VSM, 2. Schuljahr)

*Anmerkung:* Bei den fehlenden Angaben handelt es sich um Nebenlehrpersonen. Die Hauptlehrperson gibt in diesen Fällen an, dass es sich um eine typische Stunde handelt, womit angenommen werden kann, dass auch die Nebenlehrperson, für die kein Lehrerinterview vorliegt, zu diesem Urteil kommen würde.

Diejenigen Lehrpersonen, die die Frage nach der Repräsentativität der Stunde im Fach Deutsch mit „nein“ beantworteten, gaben als Gründe beispielsweise folgende an (Auszug aus dem Transkript zum Lehrerinterview):

- 1115\_11111: „Die war in dem Sinne nicht typisch, dass wir wirklich gerade in Deutsch ganz wenige solche frontalen Einführungen machen, weil wir wirklich fast individuell nur arbeiten, deswegen war sie doch untypisch. Aber sonst vom Arbeitsaufwand ist es nichts anderes gewesen.“
- 1213\_12107: „Diese Buchvorstellung haben wir ja noch nicht gemacht. Ich kann mir aber vorstellen, dass ein Grundgerüst in Zukunft so sein wird, dass eben auch, wenn die Kinder was machen, mit

einspielen oder erzählen können. So in der Form schon und na ja mit dem Lesen... Wenn die Kinder das vorstellen fällt das weg. // Es war schon für die Kinder eine besondere Situation. Die vielen Leute, die Kameras vielleicht weniger, aber wir waren ja ziemlich viele Personen.“

1312\_13105: „Es war schon dadurch, dass es jetzt eine Doppelstunde war, man hat dann doch schon gemerkt, dass es zum Schluss von der Konzentration her dann nachließ. Nein, es war doch irgendwo außergewöhnlich.“

Im Fach Mathematik gaben die Lehrpersonen beispielsweise folgende Gründe dafür an, dass ihre Stunde weniger repräsentativ war (sinngemäße Wiedergabe der Antworten aus dem Lehrerinterview<sup>73</sup>):

- 1115\_11110: Einführungsstunden werden normalerweise mit Frontalunterricht durchgeführt, in Übungsstunden wird aber häufig Werkstattarbeit durchgeführt.
- 2231\_22302: Normalerweise dauert die Einführung länger; die Gruppe wird manchmal bei der Einführung geteilt.
- 2322\_23205: Normalerweise haben die Schüler morgens 20 Minuten Zeit, um sich auszutauschen, dann werden normalerweise Hausaufgaben besprochen.

Obwohl die Schüler selbst nicht nach der Repräsentativität der videografierten Stunde befragt wurden, kann davon ausgegangen werden, dass sie sich tendenziell typisch verhielten. Ein Auszug aus einem Mathematikvideo soll das verdeutlichen (L = Lehrperson + ID, S = Schüler + ID):

**2343\_23404 (Mathematik) 01:14:20 - 01:14:23:**

L23404 [zu S11]: „Tasche rein! Sofort!“ [S11 isst etwas und steckt es sich in den Mund, nachdem LPS weggegangen ist.]

Schüler würden ein solches Verhalten nicht an den Tag legen, wenn sie sich der Anwesenheit einer Kamera dauerhaft bewusst wären. Im Projekt DESI konnten lediglich minimale Unterschiede zwischen einer „regulären“ und videografierten Englischstunde bestätigt werden, wenn man die Schüler befragt. In der videografierten Stunde verwendete die Lehrperson im Vergleich zur „regulären“ mehr Material und es ging etwas ruhiger zu (z. B. Fach Englisch, vgl. Helmke et al., 2008). Auch in der TIMSS-Studie zeigte sich, dass die Aufmerksamkeit der Schüler durch die Anwesenheit einer Kamera im Klassenzimmer nicht beeinträchtigt wurde (vgl. Bos et al., 2008). Zwar wurde die Anwesenheit der Kamera als zum Teil störend empfunden, nach einer gewissen Zeit bemerkten die Schüler diese jedoch gar nicht mehr. Im Rahmen der PERLE-Studie konnten sogar einzelne Situationen identifiziert werden, in denen die Kamera in das Unterrichtsgeschehen eingebunden wurde, ohne

73 Die Lehrerinterviews aus der Videostudie Mathematik liegen bislang nicht in transkribierter Form vor.

dass es den Anschein erweckt, dass diese als störend empfunden wurden (siehe Transkriptausschnitte Mathematik, 2133\_21307 und 2141\_21404):

**2133\_21307 (Mathematik)** (Schüler bildet aus der Anzahl der Kameras eine Malaufgabe)

**00:48:38 - 00:49:06: S09:** „Ich hab’ zuerst die beiden Kameras genommen. Dahinten und hier. Das war 1 plus 1 ist gleich 2. Und die Malaufgabe ist 2 mal 1 ist gleich 2. [...]“

**2141\_21404 (Mathematik)** (Schüler bildet aus der Anzahl der Kameras eine Malaufgabe)

**01:01:35 - 01:01:43: S14:** „Äh, das sind drei Mikrofone und zwei Kameras, 2 mal 3.“

## 11.4 Hierarchische Datenstruktur

Bei der Analyse der vorliegenden Daten muss beachtet werden, dass z. B. die Leistung oder das Selbstkonzept eines Schülers sowohl durch Einflüsse der Individualebene (individuelle Unterschiede zwischen den Schülern) als auch durch Einflüsse der Klassenebene (Unterschiede zwischen den Schulklassen) geprägt sind. Die Schülerdaten können demnach nicht unabhängig vom Klassenkontext betrachtet werden. Schüler derselben Klassen sind gemeinsamen Einflüssen ausgesetzt (gemeinsamer Unterricht, soziale Interaktionen innerhalb der Klasse etc.) bzw. sind sich in Bezug auf bestimmte Merkmale ähnlicher (vgl. Geiser, 2010; Hofman & Gavin, 1998). Die Merkmale der Klasse sind in diesem Fall für alle Schüler gleich, können aber von Schulklasse zu Schulklasse variieren. Man spricht bei einer solchen, auf mehreren unterschiedlichen Ebenen geschachtelten, Datenstruktur von einer Mehrebenenstruktur (vgl. Ditton, 1998; Geiser, 2010; Lüdtke, Robitzsch, Trautwein & Köller, 2007).

Diese Abhängigkeit der Daten nicht zu beachten, kann zu einer Verzerrung der Ergebnisse führen. Insbesondere besteht die Gefahr, dass es zu einer Unterschätzung der Standardfehler der Modellparameter (z. B. der Regressionskoeffizienten) und damit zu einer Unterschätzung des Alpha-Risikos (= verzerrte statistische Inferenz) kommt (vgl. Ditton, 1998; Geiser, 2010; Raudenbush & Bryk, 2002). Außerdem würden die Konfidenzintervalle inkorrekt geschätzt werden. Auch die p-Werte (z. B. für die Regressionsgewichte) würden zu gering geschätzt werden (vgl. Geiser, 2010).

Ein weiterer Grund für die Berücksichtigung der Mehrebenenstruktur liegt in der vorliegenden Arbeit darin, dass die Modellierung von Variablen auf mehreren Ebenen von inhaltlichem Interesse ist, da zur Vorhersage der Leistungen bzw. Selbstkonzepte in Studie 3 (vgl. Kapitel 16) nicht nur Schülerdaten, sondern auch Unterrichtsqualitätsmerkmale auf Klassenebene he-

rangezogen werden. In der vorliegenden Arbeit stellen die Klassenführung und das Unterrichtsklima Merkmale des Unterrichts (= Klassenebene) dar, deren Ausprägungen sich je nach Klasse unterscheiden. Schüler der Klasse A lernen zum Beispiel in einer Klasse, in der es der Lehrperson sehr gut gelingt, eine störungsfreie und angenehme Lernatmosphäre zu schaffen, während die Schüler der Klasse B häufiger gestört werden. Da diese Unterrichtsmerkmale unter anderem bei der Vorhersage von Variablen auf Schülerebene interessant sind (z. B. Selbstkonzept oder Leistungen), wird in dieser Arbeit die Mehrebenenstruktur berücksichtigt.

Ein gängiges Auswertungsverfahren für hierarchische Daten stellt die Mehrebenenregressionsanalyse dar. Mithilfe dieser speziellen Form der Regressionsanalyse können Phänomene gleichzeitig auf unterschiedlichen Analyseebenen betrachtet werden (z. B. auf Schüler- und Klassenebene), d. h. es wird in der Analyse adäquat betrachtet, dass mögliche Einflussprädiktoren auf verschiedenen Ebenen angesiedelt sind (vgl. Geiser, 2010). Außerdem findet in den Mehrebenenanalysen eine Trennung der Merkmalsvariation zwischen den verschiedenen Ebenen statt.

Um die Bedeutung der beiden Unterrichtsqualitätsmerkmale (Klassenführung und Unterrichtsklima) für multikriteriale Zielvariablen zu überprüfen, wird auf das Programm *MPlus* (vgl. Muthén & Muthén, 1998–2007) zurückgegriffen. Um der Mehrebenenstruktur gerecht zu werden, wurde in *MPlus* der Befehl *type is twolevel* verwendet und die Klassen-ID als Clustervariable definiert. Die Verwendung der Klassen-ID als Clustervariable hat zur Folge, dass Ähnlichkeiten innerhalb der Klasse und Unterschiede zwischen den Klassen nicht überschätzt werden. Die Abhängigkeiten der Werte innerhalb eines solchen Clusters werden während der Schätzung der einzelnen Pfadkoeffizienten berücksichtigt, indem *MPlus* eine Korrektur des Standardfehlers vornimmt (vgl. Muthén & Muthén, 1998–2007). Um den Grad der Abhängigkeit der Beobachtungen festzustellen, werden in der vorliegenden Arbeit die Intraklassenkorrelationen (ICC) berechnet (Kapitel 16). Durch die Intraklassenkorrelation (ICC) wird der Anteil der Varianz an der Gesamtvarianz bestimmt, der zwischen den Klassen liegt. Eine hohe Intraklassenkorrelation bedeutet, dass sich Schüler einer Klasse in dem untersuchten Merkmal ähnlich sind. Nach Geiser (2010) können bereits Intraklassenkorrelationen von .05 oder .01 zu Verzerrungen der Ergebnisse von Signifikanztests konventioneller regressionsanalytischer Verfahren führen, wenn die Abhängigkeit der Daten nicht berücksichtigt wird.

## 11.5 Umgang mit fehlenden Werten

Fehlende Werte (Missing Values) stellen für die sozialwissenschaftliche und psychologische Forschung eine besondere Herausforderung dar. Nach Lüdtkе et al. (2007) ist das Auftreten fehlender Werte mit drei Problemen verbunden (S. 103):

Erstens führt es auf Grund der eingeschränkten Stichprobengröße zu einem Verlust an Effizienz bei der Schätzung der Parameter. Zweitens wird der Umgang mit den Daten erschwert, weil die statistischen Standardverfahren vollständige Datenmatrizen erwarten. Drittens besteht auf Grund systematischer Unterschiede zwischen den beobachteten und den fehlenden Werten die Gefahr verzerrter Parameterschätzungen [...].

Bevor darauf eingegangen wird, wie in der vorliegenden Arbeit mit fehlenden Werten umgegangen wurde, ist es wichtig zu wissen, wie Lücken in den entsprechenden Datensätzen entstanden sind. Lüdtkе et al. (2007) klassifizieren drei Typen von fehlenden Werten. Bei „vollständig zufällig“ entstandenen ‚Missings‘ (Missing completely at Random, MCAR) kann davon ausgegangen werden, dass das Auftreten fehlender Werte weder mit der Ausprägung der Variable selbst noch mit der Ausprägung anderer Variablen zusammenhängt. Dies ist u. a. dann der Fall, wenn z. B. ein Leistungstest in Mathematik von allen Schülern ausgefüllt wird, das Antwortverhalten jedoch weder von der Leistungsstärke der Schüler noch von anderen Kovariaten (z. B. dem Geschlecht oder dem Selbstkonzept in Mathematik) beeinflusst wird. In diesem Fall stellen die Schüler mit fehlenden Werten eine Zufallsstichprobe dar. Bei „zufällig“ entstandenen ‚Missings‘ (Missing at Random, MAR) hängt das Fehlen eines Wertes mit anderen Variablen zusammen (z. B. mit der Höhe des Selbstkonzepts oder dem Geschlecht), jedoch nicht mit der Ausprägung der Variable selbst. Beispielsweise kann angenommen werden, dass Schüler mit einem niedrigeren mathematischen Selbstkonzept einzelne Aufgaben in einem Leistungstest nicht lösen. Ein zufälliges Fehlen von Werten liegt also erst dann vor, wenn weitere Variablen berücksichtigt werden. Ist das Fehlen eines Wertes weder „vollständig zufällig“ noch „zufällig“, spricht man von „nicht zufälligen“ ‚Missings‘ (Missing not at Random, MNAR). Von MNAR muss ausgegangen werden, wenn das Auftreten fehlender Werte von der Ausprägung der Variable selbst abhängt, wenn beispielsweise leistungsschwache Schüler systematisch schwierige Aufgaben in einem Leistungstest nicht lösen.

Für die vorliegende Stichprobe sind in den Schülerdaten (vgl. Kapitel 12.2) ‚Missings‘ vor allem dadurch entstanden, dass nicht alle Schüler über die zwei Schuljahre hinweg an der Studie teilgenommen haben, sei es, dass sie krankheitsbedingt an einer Erhebung nicht teilnehmen konnten oder dass sie im Laufe der ersten beiden Schuljahre die Klasse gewechselt haben. Sol-

che, durch äußere Bedingungen hervorgerufene ‚Missings‘ sind aufgrund des längsschnittlichen Designs der PERLE-Studie jedoch nicht verwunderlich.

Bei der Einschätzung der Klassenführung und des Unterrichtsklimas durch die Rater gab es nur für einzelne BIP-Lehrpersonen bei einzelnen Merkmalen ‚Missings‘. Bei diesen Lehrpersonen handelt es sich um sogenannte Nebenlehrpersonen, die eine mehr oder weniger wichtige Rolle im gesamten Unterrichtsgeschehen spielen.<sup>74</sup> Da diese Nebenlehrpersonen aufgrund ihrer Rolle für die Studie 3 jedoch nicht berücksichtigt werden, sind für Prädiktoren auf Ebene 2 (Klassenführung und Unterrichtsklima) in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) keine fehlenden Werte vorhanden.

Es gibt unterschiedliche Möglichkeiten, mit fehlenden Werten umzugehen (vgl. Madson, Becker & Thomas, 1977; Lüdtke et al., 2007; Hu & Bentler, 1999). Für die Analysen in der Studie 3 (vgl. Kapitel 16) wird – wie bereits erwähnt – das Programm *MPlus* (vgl. Muthén & Muthén, 1998–2007) verwendet. Im Fall der Berechnung der Mehrebenenregressionsmodelle (kurz *Twolevel*-Modelle) führen fehlende Werte automatisch zu ‚Listwise Deletion‘ (fallweiser Ausschluss), d. h. es werden nur diejenigen Personen in die Analysen einbezogen, die für alle Variablen, die im Modell verwendet werden, gültige Werte aufweisen. Je nach Modell schwankt somit die Größe der Stichprobe. Eine Alternative würde die multiple Imputation darstellen, bei der jeder fehlende Wert durch eine bestimmte Anzahl geschätzter Werte ersetzt wird. Auf diese Methode wurde in der vorliegenden Arbeit allerdings nicht zurückgegriffen.

## 12 Operationalisierung und Erfassung der Unterrichtsqualitäts- und Schülermerkmale

In diesem Kapitel wird detailliert auf die Operationalisierung und Erfassung der Klassenführung und des Unterrichtsklimas sowie der Schülermerkmale (Selbstkonzept und Leistungen in Deutsch und Mathematik) eingegangen. Während das Kapitel 12.1 die Konzeption des hoch inferenten Ratingsystems und die Ratingprozedur umfasst und somit Ausgangspunkt für Studie 1 und 2 darstellt, widmet sich Kapitel 12.2 den Erhebungsverfahren, mit deren Hilfe die Untersuchung der Fragestellungen zur Studie 3 erfolgt.

---

74 Auf die Rolle der Nebenlehrperson in den BIP-Klassen wird im Rahmen des Kapitels 12.1.4.1 ausführlicher eingegangen.

## 12.1 Das hoch inferente Ratingsystem

Da die Merkmale der zwei Basisdimensionen guten Unterrichts nicht direkt beobachtbar sind, sondern eher als abstrakt gelten (vgl. Helmke & Schrader, 2008), müssen verschiedene beobachtbare Indikatoren herangezogen werden. Solche Indikatoren beziehen sich auf bestimmte Regelmäßigkeiten im beobachteten Verhalten der Lehrperson und Schüler bzw. auf allgemeine Aspekte des Unterrichtsgeschehens und wurden für jedes Merkmal der Klassenführung und des Unterrichtsklimas separat beschrieben und im Rahmen eines hoch inferenten Ratingsystems zusammenfassend dargestellt. Beide Basisdimensionen wurden als fachunspezifische Unterrichtsmerkmale operationalisiert, d. h., weder nahmen die Indikatoren auf fachspezifische Besonderheiten Bezug noch erforderte das Rating spezifische fachwissenschaftliche oder fachdidaktische Kenntnisse von den Ratern. Das hoch inferente Ratingssystem wurde im Rahmen der vorliegenden Arbeit erstmals in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) eingesetzt.<sup>75</sup> Nachfolgend wird zunächst auf die Entwicklung (Kapitel 12.1.1) und den Aufbau des Ratingsystems (Kapitel 12.1.2) eingegangen, wobei der Fokus auf der Operationalisierung der im Theorieteil vorgestellten Merkmale der Klassenführung (Kapitel 3.2) und des Unterrichtsklimas (vgl. Kapitel 4.2) liegt. Anschließend werden die zugrundeliegende Analyseeinheit (Kapitel 12.1.3) und die Ratingprozedur (Kapitel 12.1.4) beschrieben.<sup>76</sup>

### 12.1.1 Die Entwicklung des hoch inferenten Ratingsystems

Die Entwicklung des Ratingsystems basiert auf einem Verfahren, bei dem sowohl deduktive als auch induktive Herangehensweisen kombiniert wurden (vgl. Hodel & Waldis, 2007; Hugener et al., 2006; Jacobs et al., 1999; Seidel, 2003c). Ausgangspunkt für die Entwicklung des hoch inferenten Ratingsystems bildeten theoretische Grundlagen bezüglich der Unterrichtsqualitätsmerkmale Klassenführung und Unterrichtsklima in der Grundschule (Kapitel 3.2 und 4.2). Dabei wurde einerseits auf bereits bestehende Beobachtungsinstrumente – aus der Sekundarstufe – zurückgegriffen (vgl. DESI-Konsortium, 2008; Rakoczy & Pauli, 2006; Seidel, 2003c; Seidel, 2005; Seidel & Prenzel, 2004), andererseits wurden auf Grundlage der Theorie neue

---

75 Neben den Fächern Deutsch und Mathematik wurde das hoch inferente Ratingsystem auch im Fach Kunst eingesetzt. Die Ratings der Kunstvideostudie sind allerdings nicht Teil der vorliegenden Arbeit.

76 Das hoch inferente Ratingsystem zur Erfassung der Klassenführung und des Unterrichtsklimas sowie die Ratingprozedur und die Analyseeinheit wurde im Rahmen des „Technischen Berichts der PERLE-Videostudien“ in ähnlicher Weise publiziert (vgl. Gabriel & Lipowsky, 2013a; b).

Ratingdimensionen speziell für die Grundschule entwickelt. Beispielsweise wurde nicht nur das Zeitmanagement allgemein betrachtet, sondern das Merkmal „Übergangsmanagement“ ergänzend aufgenommen, da Studien im Vorfeld zeigen konnten, dass Übergänge zwischen einzelnen Arbeitsphasen und Sozialformen etc. im Vergleich zur Sekundarschule in der Grundschule mehr Zeit beanspruchen (Kapitel 3.2). Während der Entwicklung des hoch inferenten Ratingsystems fanden immer wieder Prozesse der Klassifikation, Selektion und Abstraktion statt (vgl. Brückmann, 2009; Hugener et al., 2006; Seidel, 2003c). Die Vorgehensweise bei der Entwicklung des hoch inferenten Ratingsystems ist in Abbildung 18 dargestellt.

Auf Basis der Sichtung einzelner Lektionen und des aktuellen Forschungsstandes wurden erste Vermutungen und Hypothesen dahingehend formuliert, welche Merkmale der Klassenführung und des Unterrichtsklimas für die Grundschule relevant sind. In einem nächsten Schritt wurde das Ratingssystem entwickelt. Zur Entwicklung des Ratingsystems wurden Videos aus der PERLE-Stichprobe herangezogen. Somit konnte der Ausprägungsgrad der Klassenführungsmerkmale und der Merkmale des Unterrichtsklimas anhand der Stichprobe abgeschätzt und eine ausreichende Differenzierung zwischen den einzelnen Klassen hergestellt werden. Während dieses Vorgehensschrittes erfolgten zusätzlich die Wahl der Analyseeinheit (vgl. Kapitel 12.1.3) und die Operationalisierung der theoretisch relevanten Merkmale der Klassenführung und des Unterrichtsklimas für die Grundschule (vgl. Kapitel 3.2 und 4.2).

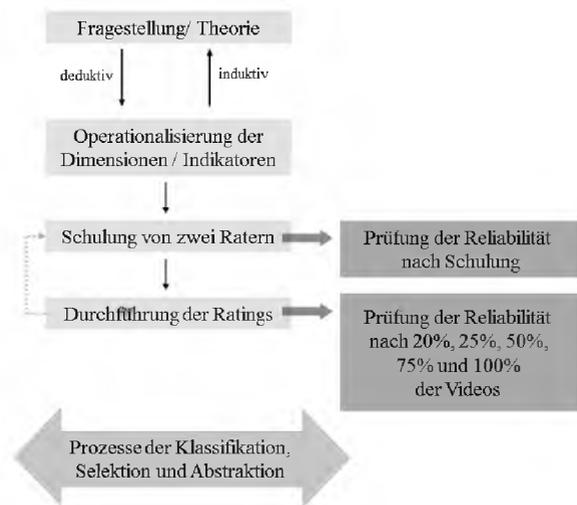


Abbildung 18: Vorgehen bei der Entwicklung des hoch inferenten Ratingsystems für die Analyse der Unterrichtsvideos in PERLE (in Anlehnung an Hugener et al., 2006; Seidel, 2003c).

Nach der Entwicklung eines vorläufigen Ratingsystems wurden zwei Rater (Lehramtsstudenten der Universität Kassel) im Rahmen einer zweitägigen Schulung anhand von zwei Unterrichtsvideos (ein Deutsch- und ein Mathematikvideo) der PERLE-Stichprobe in dem vorläufigen Ratingsystem geschult. Während des Trainings wurden die Rater zuerst mit dem Aufbau des Ratingsystems (Kapitel 12.1.2) vertraut gemacht und Ankerbeispiele und Videoausschnitte für jedes zum Rating anstehende Merkmal vorgestellt. Ziel war es, den Ratern die gesamte Bandbreite des Videomaterials bewusst zu machen. Anschließend wurde im Rahmen eines weiteren Ratings von zwei Unterrichtsstunden (eine Deutsch- und eine Mathematikstunde) überprüft, ob die zwei Rater unabhängig voneinander zu ähnlichen Urteilen gelangen.<sup>77</sup> Wie die zwei Rater zu ihren einzelnen Urteilen gekommen sind, wurde daraufhin erneut in der Gruppe (Entwicklerin des Ratingsystems + zwei Rater) diskutiert. Bei dieser ersten Anwendung des Ratingsystems zeigte sich, dass das Antwortformat zu wenig differenzierte (es gab zu wenige Varianz) bzw. einzelne Indikatoren nicht beobachtbar oder schwer übereinstimmend einzuschätzen waren. Für das endgültige Ratingsystem ergaben sich dadurch nochmals minimale Veränderungen, größere Änderungen in den Ratingdimensionen wurden nicht vorgenommen. Die Änderungen im Ratingsystem wurden mit den zwei Ratern im Rahmen einer Nachschulung besprochen. In Anlehnung an Seidel (2003c; vgl. auch Kobarg & Seidel, 2003) galt das Training ab dem Zeitpunkt als abgeschlossen, als sich die Rater und die Entwicklerin einig waren, dass die Urteile auf Basis eines gemeinsamen theoretischen Verständnisses gefällt wurden. Dieses gemeinsame theoretische Verständnis der einzelnen Merkmale ist Voraussetzung für das Rating aller weiteren Unterrichtsstunden.

Auf Grundlage des überarbeiteten Ratingsystems schätzten die zwei Rater anschließend alle Deutsch- und Mathematikvideos unabhängig voneinander ein. Dabei wurde bei der Reihenfolge der Unterrichtsvideos darauf geachtet, dass sich sowohl die Fächer (Deutsch versus Mathematik) als auch BIP- und staatliche Klassen abwechselten. Somit wurden im Vorfeld mögliche Urteilsfehler (Kapitel 6.3.1) minimiert. Nach dem Rating aller Unterrichtsvideos und der Berechnung der Beobachterübereinstimmungen (Kapitel 13) folgten statistische Analysen (z. B. Faktorenanalysen, vgl. Kapitel 15). Es konnten Rückschlüsse auf die theoretisch hergeleiteten Fragestellungen getroffen werden (Kapitel 7).

---

77 Nachdem alle übrigen Deutsch- und Mathematikvideos eingeschätzt worden sind, wurden die vier Schulungsvideos von den Ratern erneut eingeschätzt. Die Daten der vier Schulungsvideos wurden bei der Berechnung der Inter-Rater-Reliabilität nicht berücksichtigt. Für weitere Analysen wurden die Daten allerdings verwendet.

### 12.1.2 Der Aufbau des hoch inferenten Ratingsystems

Das hoch inferente Ratingsystem umfasst insgesamt 18 Merkmale (Items) zu den in Kapitel 3.2 und 4.2 hergeleiteten Merkmalen der Klassenführung und des Unterrichtsklimas in der Grundschule. Tabelle 13 gibt einen Überblick über die einzuschätzenden Merkmale. Acht Merkmale beziehen sich auf die Klassenführung (vgl. Kapitel 19.1) und zehn auf das Unterrichtsklima (vgl. Kapitel 19.2). Das Verfahren hat die Form eines Fragebogens (vgl. Kapitel 20). Das Ratingsystem beinhaltet sowohl deskriptive (quantitative) als auch evaluative (qualitative) Einschätzungen. Als eher deskriptive Einschätzung ist das Merkmal „Einsatz von Lob (Lob)“ zu erachten. Dieses gibt lediglich Auskunft über die eingeschätzte Häufigkeit von Lob während der aufgeführten Unterrichtseinheit (vgl. Lotz et al., 2013). Im Gegensatz zu der eher deskriptiven Einschätzung der Häufigkeit erfassen die übrigen 17 Merkmale zusätzlich beispielsweise die Intensität eines gezeigten Verhaltens.

Tabelle 13: Überblick über die hoch inferenten Merkmale des Ratingsystems zur Erfassung der Klassenführung und des Unterrichtsklimas in der Grundschule

Basisdimensionen der Unterrichtsqualität	Hoch inferente Merkmale des Ratingsystems
Klassenführung	(1) Allgegenwärtigkeit der Lehrperson (Allg)
	(2) Gruppenfokus (Grup)
	(3) Zeitmanagement / effiziente Zeitnutzung (Zeit)
	(4) Übergangmanagement (Managing Transition) (Ueb)
	(5) Regelklarheit bzw. -verwendung (Regl)
	(6) Störungsfreiheit (Disz)
	(7) effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)
	(8) Vorbereitung des Klassenraums (Vorb)
Unterrichtsklima	(9) humorvolle Lernatmosphäre (Humo)
	(10) Anerkennung der Schüler durch die Lehrperson (Anes)
	(11) gegenseitige Anerkennung der Schüler (Aneg)
	(12) Fürsorglichkeit / Herzlichkeit und Wärme (Fuer)
	(13) Einsatz von Lob (Lob)
	(14) Ermutigung durch die Lehrperson (Emut)
	(15) sachlich-konstruktive Rückmeldung (Rück)
	(16) positiver Umgang der Lehrperson mit Fehlern (Fehl)
	(17) positives Fehlerklima zwischen den Schülern (Fekl)
	(18) Wettbewerb und Leistungsdruck (Wett)

Die *Operationalisierung* der einzelnen hoch inferenten Merkmale erfolgte in Anlehnung an bereits existierende hoch inferente Verfahren (vgl. Kunter, 2005; Rakoczy & Pauli, 2006). Zu Beginn der Operationalisierung wur-

de zunächst jedes Merkmal der Klassenführung und des Unterrichtsklimas mithilfe der sogenannten *Grundidee* inhaltlich beschrieben. Diese Grundidee wurde anschließend in möglichst verhaltensnah formulierte und im Grundschulunterricht beobachtbare Indikatoren und teilweise Negativindikatoren operationalisiert. Diese Indikatoren halfen den zwei Ratern, sich die Unterrichtssituationen besser vorzustellen. Die Aufgabe der Rater bestand darin, einzuschätzen, inwiefern der beobachtete Unterricht mit den formulierten Indikatoren übereinstimmt.

*Antwortformat:* Für die Einschätzung lag den Ratern eine vierstufige Antwortskala vor. Nach Rakoczy und Pauli (2006) kann die Einschätzung eines Merkmals auf einer vierstufigen Skala zu Unsicherheiten bei den Ratern führen, da dem zu treffenden Urteil mehrere Indikatoren zugrunde liegen. Der Rater muss sowohl auf die Häufigkeit eines Merkmals als auch auf dessen Intensität achten. Um diese Unsicherheiten in den Einschätzungen so gering wie möglich zu halten, wurden in Anlehnung an die IPN-Videostudie Physik die vier Antwortmöglichkeiten mit Unterstützung relevanter Indikatoren schriftlich festgelegt (vgl. Kobarg & Seidel, 2003; Schulmeiß, Seidel & Meyer, 2003). Zusätzlich gab es für einzelne hoch inferente Merkmale ergänzende Anmerkungen zur Ratingprozedur (vgl. Kapitel 19).

In Anlehnung an Pietsch (2010) hatten die Rater zusätzlich auch die Möglichkeit, eine fünfte Kategorie einzuschätzen. Wenn die Rater ein Merkmal der Klassenführung oder des Unterrichtsklimas für nicht beobachtbar bzw. einschätzbar hielten, musste ein „0“ im Sinne einer *Nichtkategorisierung* (vgl. Wirtz & Caspar, 2002) vergeben werden. Dies betraf ausschließlich das Verhalten einiger Nebenlehrpersonen in den BIP-Klassen. Besonders einzelne Aspekte der Klassenführung ließen sich für eine Nebenlehrperson, die eine geringe Rolle im gesamten Unterrichtsgeschehen spielte, nicht einschätzen, da die Hauptlehrperson z. B. für das Zeitmanagement verantwortlich war. Das Vorgehen bei der Einschätzung der Merkmale für die Nebenlehrperson wird detailliert in Kapitel 12.1.4.1 beschrieben.

Anschließend ist der Aufbau des Ratingsystems anhand des hoch inferenten Merkmals „Übergangsmanagement (Managing Transition)“ beispielhaft mit der dazugehörigen Grundidee, den Indikatoren und Negativindikatoren sowie dem Antwortformat dargestellt (vgl. Tabelle 14). Das komplette hoch inferente Ratingsystem befindet sich im Anhang dieser Arbeit (Kapitel 19) bzw. wurde bereits an anderer Stelle publiziert (vgl. Gabriel & Lipowsky, 2013a; b).

Tabelle 14: Aufbau des hoch inferenten Ratingsystems am Beispiel des Übergangsmanagements (Managing Transition)

Merkmal	Übergangsmanagement (Managing Transition)
Grundidee	Ein besonderes Merkmal des Zeitmanagements ist die Organisation der einzelnen Übergänge zwischen den Unterrichtsphasen. Die gute Organisation von Übergängen (z. B. beim Umstellen von Tischen, beim Bilden von Stuhl- bzw. Sitzkreisen oder beim Austeilen von Unterrichtsmaterialien) sorgt für einen reibungslosen Unterrichtsablauf. Das Merkmal „Übergangsmanagement“ erfasst demnach die Fähigkeit der Lehrperson, die Übergänge zwischen verschiedenen Unterrichtsphasen so zu organisieren, dass sie kurz und reibungslos erfolgen und keine unnötigen Pausen entstehen. Dabei geht es darum, dass die Lehrperson mit den Kindern wiederkehrende Abläufe zwischen den Unterrichtssequenzen so organisiert hat, dass sie keiner Erklärung mehr bedürfen.
Indikatoren	<ul style="list-style-type: none"> <li>– Die Übergänge verlaufen reibungslos (z. B. Übergang vom Stuhlkreis zu anderen Beschäftigungen, der Wechsel von lehrerorientiertem Unterricht zu Freiheitsphasen oder beim Wechsel zwischen verschiedenen Unterrichtsgegenständen).</li> <li>– Die Übergänge verlaufen relativ lärmfrei.</li> <li>– Die Übergänge zwischen einzelnen Unterrichtsphasen erfolgen durch knappe, eindeutige Überleitungen.</li> <li>– Die Lehrperson setzt akustische oder visuelle Impulse / Signale (Triangel, Glöckchen, Zeichen mit den Fingern etc.) ein, um die Übergänge zu organisieren.</li> <li>– Die Anleitungen für die Übergänge sind so logisch von der Lehrperson formuliert, dass die Schüler ihnen folgen können.</li> <li>– Es scheint Routinen zu geben, die die Übergänge regeln: Die Schüler scheinen zu wissen, was von ihnen verlangt wird (z. B. „in den Kreis kommen“, „in den Gruppen zusammen sitzen“), sodass solche Wechsel oder Übergänge ohne großen Aufwand geschehen und der Unterricht davon kaum unterbrochen wird.</li> <li>– Materialien und Unterlagen, die im Laufe der nächsten Unterrichtsphase benötigt werden, liegen / stehen bereit, sodass das Verteilen an die Schüler relativ zügig abläuft. (Das Austeilen oder Einsammeln ist gut organisiert und wird rasch durchgeführt, sei es von der Lehrperson selbst oder von einem Schüler / mehreren Schülern.)</li> </ul>
Negativindikatoren	<ul style="list-style-type: none"> <li>– Die Lehrperson gibt keine eindeutigen Anweisungen zum Übergang in andere Unterrichtsphasen (z. B. vom Unterrichtsgespräch zum Sitzkreis vor der Tafel).</li> <li>– Es scheint keine Routinen für die Übergänge zu geben.</li> <li>– Die Lehrperson achtet in den Übergangsphasen nicht auf eine zügige Durchführung.</li> <li>– Die Übergänge verlaufen nicht reibungslos. Die Lehrperson muss die Schüler häufig zurechtweisen und um eine zügige Durchführung bitten.</li> <li>– Die Lehrperson achtet nicht auf eine konsequente Durchführung: Die Lehrperson macht nach einem Übergang mit dem Stoff weiter, obwohl noch nicht alle Schüler ihre Hefte aufgeschlagen haben oder im Sitzkreis angekommen sind.</li> </ul>

– Fortsetzung nächste Seite –

Merkmal	Übergangsmanagement (Managing Transition)
Antwort (Gesamteindruck)	<p>Eine „4“ wird vergeben, wenn die Übergänge optimal geplant sind und reibungslos und ohne Zeitverluste ablaufen. Es kann davon ausgegangen werden, dass die Lehrperson Routinen, die die Übergänge regulieren, bereits eingeführt hat. Bei einer „3“ sind einzelne eher geringe Mängel zu erkennen. Die Übergänge sind mitunter etwas ineffektiv.</p> <p>Die „2“ wird vergeben, wenn die Übergänge wenig effektiv sind. Die Lehrperson achtet nicht immer auf eine zügige Durchführung und muss die Schüler während des Übergangs häufig zurechtweisen.</p> <p>Um eine „1“ zu vergeben, müssen beim Übergang zwischen verschiedenen Aktivitäten Zeitverluste ersichtlich sein. Die Lehrperson scheint keine klaren Übergangsregeln zu haben und macht häufig mit dem Unterrichtsstoff weiter, obwohl einzelne Schüler noch nicht im Sitzkreis sind oder noch keine Materialien haben. Der Beobachter muss häufig den Eindruck haben, dass die Schüler nicht wissen, was sie als nächstes machen sollen.</p>

### 12.1.3 Die Analyseeinheit

Entsprechend dem hoch inferenten Ansatz wurde in der vorliegenden Arbeit die gesamte Unterrichtsstunde als Analyseeinheit gewählt. Infolgedessen wird die komplette Lektionsdauer (ca. 90 Minuten Unterricht) als Analyseeinheit festgelegt. Für die Sequenzierung der zu ratenden Merkmale der Klassenführung und des Unterrichtsklimas wurde im Vorfeld die Lektionsdauer, d. h. der sekundengenaue Beginn und das Ende der Lektion bestimmt (vgl. Lotz, 2013a). Somit wurde gewährleistet, dass sich die Einschätzungen der zwei Rater auf die gleiche Analyseeinheit (Lektionsbeginn bis Lektionsende) beziehen.

Da im Rahmen der PERLE-Videostudien insgesamt ungefähr 90 Minuten Unterricht aufgezeichnet wurden, erschien es sinnvoll, pro 45 Minuten ein Urteil abzugeben. Dieses Vorgehen kann unter anderem mit der Tatsache begründet werden, dass die Lehrpersonen einen typischen und alltagsnahen Unterricht realisieren sollten und somit einige Lehrpersonen nach regulär ca. 45 Minuten eine Frühstückspause einlegten, andere jedoch aufgrund der Vorgabe von 90 Minuten Unterricht nicht. Anhand der Basiskodierung können, je nach Videostudie, Pausen von bis zu 20 Minuten ermittelt werden.<sup>78</sup> In diesen Frühstückspausen konnte bei der ersten Sichtung des Materials teilweise beobachtet werden, dass die Lehrpersonen unter anderem den nächsten Teil der Stunde vorbereiteten oder organisatorische Angelegenheiten (sog. Non-Instructional Activities, vgl. Kapitel 3.2) stattfanden. Um eine einheitliche Auswertung zu gewährleisten, wurde im Vorfeld beschlossen, dass die

<sup>78</sup> Da es sich bei der aufgezeichneten Stunde um eine typische Unterrichtsstunde handeln sollte, wurden den Lehrpersonen im Vorfeld der Videostudien keine Vorgaben gemacht, ob sie durchgehend 90 Minuten Unterricht halten oder ob sie Pausen für die Kinder einplanen sollen.

Aktivitäten der Frühstückspause nicht in die Einschätzungen der Rater einfließen sollten.

Für das Vorgehen beim Rating bedeutet dies, dass die Rater, bevor sie die Merkmale der Klassenführung und des Unterrichtsklimas einschätzten, mithilfe der Basiskodierung die Pause identifizieren mussten (vgl. Lotz, 2013b) und an entsprechender Stelle nicht berücksichtigen durften. In der Basiskodierung wurden auch kurze Pausen (z. B. Trink- oder Entspannungspausen oder -phasen) kodiert.

Zur Verwendung der Ratings für die Berechnung der Inter-Rater-Reliabilität (vgl. Kapitel 13) wurde jeder Lehrperson für jede Videostudie ein Wert zugeordnet, indem der Mittelwert der beiden Ratings pro einzuschätzendes Merkmal und pro Rater (2 x 45 Minuten) gebildet wurde. Hat beispielsweise eine Lehrperson in Mathematik für die ersten 45 Minuten vom Rater 1 für ein Merkmal eine „1“ und für den zweiten Teil der Stunde eine „2“ bekommen, ergibt sich für die gesamten 90 Minuten bei Rater 1 ein Mittelwert von „1.5“. Für Rater 2 wurde gleichermaßen vorgegangen.

In Abhängigkeit von der Höhe der Übereinstimmung der Rater wurden für alle weiterführenden Analysen (z. B. Faktorenanalysen, vgl. Kapitel 15) für jede Lehrperson die Mittelwerte der zwei Rater pro Merkmal nochmals gemittelt. Ein Beispiel soll dieses Vorgehen erläutern: Rater 1 schätzt z. B. das Merkmal „Allgegenwärtigkeit der Lehrperson (Allg)“ sowohl für den ersten als auch für den zweiten Teil der Stunde mit „3“ ein und hat somit im Mittel für die 90 Minuten Unterricht eine „3“ für dieses Merkmal der Klassenführung vergeben. Rater 2 beurteilt den ersten Teil der Stunde ebenfalls mit „3“. Für den zweiten Teil der Stunde vergibt er eine „1“. Im Mittel erhält die Lehrperson vom Rater 2 demnach für die „Allgegenwärtigkeit der Lehrperson (Allg)“ eine „2“. Die Mittelwerte über die Stunde wurden anschließend erneut über beide Rater gemittelt. Es ergibt sich somit für die Lehrperson ein Wert von „2.5“  $[(3+2)/2]$ . Im folgenden Kapitel wird näher auf das Vorgehen während des Ratings eingegangen.

#### *12.1.4 Die Ratingprozedur*

Das Rating bzw. die inhaltliche Auswertung des Videomaterials erfolgte mithilfe des Softwareprogramms Videograph (vgl. Rimmele, 2002). Videograph hat den Vorteil, dass es das gleichzeitige Abspielen der Klassen- und Lehrerkamera ermöglicht. Da zur Beurteilung der Klassenführung und des Unterrichtsklimas sowohl das Verhalten der Lehrperson als auch das der Schüler relevant ist, wurden beide Kameraperspektiven für die Auswertung herangezogen (vgl. Abbildung 19). Die Einschätzung der Merkmale erfolgte auf Grundlage der Basiskodierungen (vgl. Lotz, Lipowsky & Faust, 2013), d. h. dass im Vorfeld des Ratings die Stunden in Bezug auf die Lektionsdauer, die Klassenteilung und die Sozialformen sowie die inhaltsbezogenen Aktivitäten

der jeweiligen Fächer kodiert wurden. Für das hoch inferente Rating liegen somit bereits das Transkript sowie Angaben zur Lektionsdauer (als Analyseinheit, vgl. Kapitel 12.1.3) vor. Die im Vorfeld im Rahmen der Basiskodierung bestimmten Sozialformen wurden zusätzlich für die Identifikation der Pause(n) benötigt.

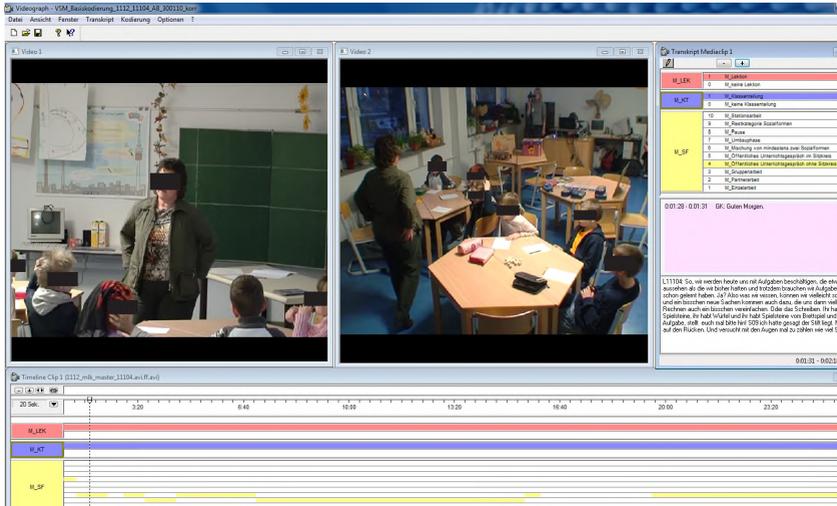


Abbildung 19: Auszug aus dem Softwareprogramm Videograph (vgl. Rimmele, 2002)

*Anmerkung:* Links Lehrerkamera, mittig Klassenkamera, rechts Transkript und Basiskodierung, unten Timeline

Für das Rating war nicht nur die Identifikation der Pausen von Bedeutung, sondern auch, ob es sich bei der einzuschätzenden Lektion um ein BIP-Video oder um ein Video aus einer staatlichen Grundschulklasse handelte. Auf welche Besonderheiten die Rater bei der Einschätzung von BIP-Klassen achten sollten und welche Rolle in den BIP-Klassen die Nebenlehrperson spielt, wird anschließend erläutert.

Die Unterrichtsorganisation in den BIP-Klassen stellt aufgrund der Anwesenheit von zwei Lehrpersonen in den Hauptfächern Deutsch und Mathematik sowie der Möglichkeit zur Nutzung eines zweiten Klassenraums (bzw. Hortzimmers) eine besondere Herausforderung für den Ratingprozess dar. Mit Ausnahme des Faches Kunst nutzten die Lehrpersonen in den Fächern Deutsch und Mathematik die Möglichkeit, die Schülergruppe durchgängig oder zeitweise in zwei Teilgruppen aufzuteilen und diese in unterschiedlichen Räumen zu unterrichten. In PERLE wurde der Unterricht in beiden Klassenzimmern gefilmt, sodass in diesen Fällen für eine Klasse zwei Videos vorliegen. Es existieren jedoch von Klasse zu Klasse Unterschiede in der Umsetzung dieser Klassenteilung bzw. in der Umsetzung des Teamteachings (vgl.

Lotz, 2013c). Einerseits gibt es Klassen, die die zwei Räume dafür nutzen, um voneinander unabhängige Lektionen mit jeweils einem Teil der Klasse durchzuführen (Klassenteilung, pro Klassenraum eine Lehrperson). In diesem Fall liegen für die Haupt- und Nebenlehrperson zwei getrennte Unterrichtsvideos vor und es konnte für beide Lehrpersonen (Haupt- oder Nebenlehrperson) sowohl die Klassenführung als auch das Unterrichtsklima von den zwei Ratern eingeschätzt werden.

Andererseits gibt es Klassen, in denen die Schüler und die Lehrpersonen die Räume je nach Bedarf wechselten. Für die Auswertung dieser Unterrichtsstunden musste eine Lösung gefunden werden, die es ermöglicht, das Verhalten beider Lehrpersonen beim Rating zu berücksichtigen (vgl. Lotz, Berner & Gabriel, 2013). In diesem Fall wurde wie folgt vorgegangen: Die zwei Rater beurteilten in einem ersten Ratingdurchgang zunächst die Hauptlehrperson nach den interessierenden Merkmalen der Klassenführung und des Unterrichtsklimas und gaben zwei Urteile ab. Welche Lehrperson die Hauptlehrperson im jeweiligen Fach war, wurde im Vorfeld per Nachfrage an den Schulen ermittelt. Im Anschluss schauten die Rater das Video erneut an und legten den Fokus der Beobachtung auf das Verhalten der Nebenlehrperson. Das bedeutet, dass das gleiche Video je nach Fokus auf die Haupt- oder Nebenlehrperson zweimal angesehen werden musste. Am Beispiel der Videostudie Mathematik (2. Schuljahr) soll das Vorgehen der Rater bei der Klassenteilung verdeutlicht werden (vgl. Abbildung 20). Folgende Situation ist gegeben: Der Unterricht beginnt im Klassenraum mit einem öffentlichen Unterrichtsgespräch. In dieser Zeit wird mit allen Schülern und in Anwesenheit beider Lehrpersonen die Multiplikation eingeführt. Je nachdem welche Lehrperson in ihrem Verhalten eingeschätzt werden soll (Haupt- oder Nebenlehrperson), wird diese von den Ratern während dieser Phase beobachtet.<sup>79</sup> Nach ungefähr 35 Minuten (Ende der Einführungsphase der Multiplikation) verlässt die Nebenlehrperson mit einer kleineren Gruppe von Schülern den Klassenraum und geht zum Üben der neuen Rechenart ins Hortzimmer. Ab diesem Zeitpunkt mussten die Rater je nach Fokus unterschiedlich vorgehen: Lag der Fokus auf der Beobachtung der Hauptlehrperson wurde das aktuelle Video weiter angesehen. Lag der Nebenlehrer im Fokus der Einschätzung, wurde entsprechend im Parallelvideo weiter beobachtet. Wenn die Nebenlehrperson nach einer gewissen Zeit wieder zurück in den Klassenraum kam, wurde sie ab diesem Zeitpunkt wieder im Hauptvideo beobachtet. Die gleiche Situation ist auch im Deutschunterricht des ersten Schuljahres zu beobachten, wenn eine Lehrperson mit einem Teil der Klasse für die Leseübung in den zweiten Klassenraum (bzw. Hortzimmer) wechselt.

---

79 In solchen Phasen war in den meisten Fällen jedoch die Hauptlehrperson verantwortlich, sodass die Nebenlehrperson in den Hintergrund rückte.

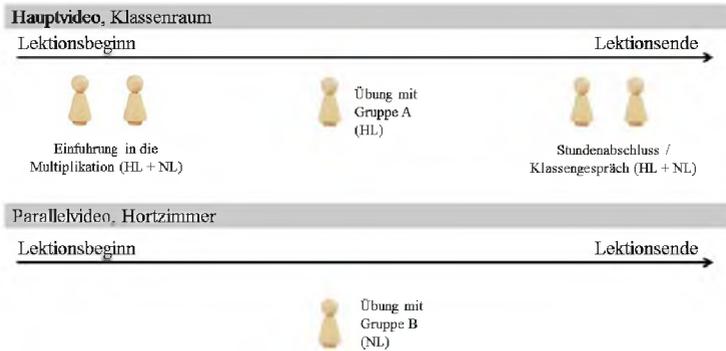


Abbildung 20: Vorgehen bei der Klassenteilung am Beispiel einer Mathematikstunde „Einführung in die Multiplikation“ (in Anlehnung an Lotz, Berner & Gabriel, 2013)

Anmerkung: HL = Hauptlehrperson, NL = Nebenlehrperson

#### 12.1.4.1 Die Rolle der Nebenlehrperson – Möglichkeit der Nichtkategorisierung

Da in jedem Klassenraum jeweils nur eine Lehrerkamera installiert wurde und diese jeweils nur eine Lehrperson fokussieren konnte, kam es im Unterrichtsgeschehen vor, dass in einem Großteil der Lektion die Hauptlehrperson gefilmt wurde, da diese die meisten Interaktionen mit den Schülern hatte. Aus diesem Grund war es in einzelnen Fällen schwierig, pro 45 Minuten eine Einschätzung für die Nebenlehrperson abzugeben. In Anlehnung an Pietsch (2010) mussten die Rater in solchen Fällen eine „0“ im Sinne einer *Nichtkategorisierung* vergeben (vgl. Kapitel 12.1.2). Die Nichtkategorisierung ist dabei nicht auf „Unstimmigkeiten zwischen den Ratern zurückzuführen, sondern darauf, dass für einen Beurteiler die Beurteilung nicht möglich war“ (Wirtz & Casper, 2002, S. 49). Bei der Berechnung der Inter-Rater-Reliabilität (vgl. Kapitel 13) wurde es als ein nichtübereinstimmendes Rating gewertet, wenn nur einer der beiden Rater keine Einschätzung vornahm. Sobald beide Rater für eine Nebenlehrperson einzelne Merkmale mit „0“ einschätzten, wurde dies als eine Übereinstimmung betrachtet.

Für alle Nebenlehrpersonen erfolgte zusätzlich die Einschätzung der ihnen am Unterrichtsgeschehen zugeteilten Rolle (auf einer Skala von „0“ bis „2“). Diejenigen Nebenlehrpersonen, die nur beim Austeilen der Arbeitsblätter halfen oder den Schülern als zusätzliche Ansprechpartner während der Übungs- oder Einzelarbeitsphasen zur Verfügung standen, jedoch sonst keine inhaltlichen Aufgaben und Lehrfunktionen in den videografierten Stunden übernahmen, bekamen eine „0“. Ihnen wurde somit eine sehr geringe Rolle am Unterrichtsgeschehen zugeschrieben. Für die Rater war es in diesen Fällen

besonders schwierig, einzelne Merkmale der Klassenführung oder des Unterrichtsklimas pro 45 Minuten einzuschätzen. Eine „1“ sollten die zwei Rater nur für diejenigen Nebenlehrpersonen vergeben, die „teilweise“ eine Rolle im Unterrichtsgeschehen spielten. Auch hier war die Einschätzung der hoch inferenten Merkmale oft schwierig. Sobald die Nebenlehrperson nach Einschätzung der beiden Rater jedoch eine gleichwertige Rolle im Unterrichtsgeschehen einnahm, sich die beiden Lehrpersonen inhaltlich abwechselten oder die Nebenlehrperson für eine gewisse Zeit mit einem Teil der Klasse in einen Nachbarraum ging (vgl. Abbildung 20), erhielten die Lehrpersonen eine „2“. Während dieser Klassenteilung betreute die Nebenlehrperson für eine gewisse Zeit eine kleinere Gruppe von Schülern in einem anderen Raum und war somit beispielsweise selbst für die Einhaltung von Regeln oder für die Lautstärke im Klassenraum verantwortlich. Dementsprechend können für diesen Zeitraum allgemeine Aussagen für einzelne Aspekte der Unterrichtsqualität (z. B für Klassenführung oder Unterrichtsklima) getroffen werden.

Da es sich auch bei der Einschätzung der Rolle der Nebenlehrperson um ein Urteil mit hoher Inferenz handelt, wurden anhand von 28 zu treffenden Einschätzungen für die BIP-Klassen überprüft, inwieweit die beiden Rater in der Einschätzung der Rolle der Nebenlehrperson übereinstimmten. Der relative G-Koeffizient beträgt .91. Es kann somit von einer zufriedenstellenden Qualität der Einschätzung der beiden Rater ausgegangen werden.

#### 12.1.4.2 Maßnahmen zu Vermeidung möglicher Beurteilungsfehler

In Kapitel 6.3.1 wurde auf mögliche Urteilsfehler eingegangen, die im Kontext hoch inferenter Ratings auftreten können. Welche Beurteilungsfehler während des Ratingprozesses denkbar sind und welche Maßnahmen zur Reduzierung systematischer Beurteilungsfehler im Rahmen der vorliegenden Studie getroffen wurden, wird anschließend erklärt.

*Halo-Effekte* sind dadurch gekennzeichnet, dass verschiedene Beurteiler tendenziell die Beurteilung verschiedener Facetten eines Merkmals von einem Globalurteil abhängig machen. Halo-Effekte können jedoch auch dann verstärkt auftreten, wenn das einzuschätzende Merkmal schwer beobachtbar ist oder schlecht definiert wurde. Um dem Auftreten solcher Effekte vorzubeugen, wurden die Rater im Vorfeld mittels des Ratingsystems geschult (vgl. Kapitel 12.1.1) und bekamen ausreichend Informationen über die Bedeutung der Merkmale der Klassenführung und des Unterrichtsklimas in der Grundschule.

*Milde-Streng-Fehler* entstehen dann, wenn mehrere Beurteiler ein Merkmal im Vergleich zum Beurteilermittel systematisch zu positiv oder zu negativ einschätzen. Insbesondere der Milde-Streng-Fehler kann als konstanter Messfehler bzw. robuste Determinante des Urteilerverhaltens immer wieder bestätigt werden (zuf. Eckes, 2004). Dieser Fehlertyp tritt vor allem bei Personenbeurteilungen auf. Hier spielen beispielsweise Sympathien für

Lehrpersonen eine Rolle. Lehrpersonen, die in den Unterrichtsvideos freundlich und aufgeschlossen erscheinen bzw. ein wirksames Auftreten besitzen, können von den Ratern generell positiver eingeschätzt werden als unsympathisch wirkende Lehrpersonen. Dadurch können beispielsweise eventuelle Schwächen in der Klassenführung leicht übersehen werden.

Um *zentrale Tendenzen* bzw. „*Tendenzen zur Mitte*“ zu vermeiden, wurde im Vorfeld eine vierstufige Antwortskala gewählt, die es den Ratern unmöglich machte, ihre Urteile in der Mitte der Skala zu konzentrieren. Außerdem wurde sowohl bei der Entwicklung des Ratingsystems als auch im Verlauf des Ratingprozesses darauf geachtet, dass die Beurteiler auch die Extrempole der Antwortskala hinreichend nutzen. Die Verankerung der Skalen durch Extrembeispiele kann als weitere geeignete Maßnahme angesehen werden.

Bei der Beurteilung von Personen können zusätzlich Urteilsverzerrungen in Abhängigkeit von der Position des Urteilers auf das zu beurteilende Merkmal entstehen. Dabei wird in der Literatur zwischen Ähnlichkeits- und Kontrastfehlern unterschieden. Während Ähnlichkeitsfehler entstehen, indem der Beurteiler dem zu bewertenden Merkmal fälschlicherweise ähnliche Eigenschaften zuschreibt, wie er selbst besitzt, entstehen Kontrasteffekte durch die Annahme, dass ein zu bewertendes Merkmal Eigenschaften besitzt, die der Beurteiler sich selbst abspricht. Beide Fehlertypen werden unter dem Begriff *Rater-Ratee-Interaktionen* zusammengefasst.

Da sich die Lehrerstichprobe zum Teil aus Lehrpersonen zusammensetzt, die sowohl an der Videostudie Deutsch als auch an der Videostudie Mathematik teilgenommen haben (vgl. Kapitel 11.1), kann es während eines Ratingprozesses zusätzlich zu sogenannten *Primacy-Recency-Effekten* kommen (vgl. Güttler, 2003). Diese können entstehen, wenn ein zu beurteilendes Merkmal, z. B. die Allgegenwärtigkeit der Lehrperson als Merkmal der Klassenführung im Fach Mathematik im Vergleich zum Fach Deutsch beurteilt wird. Um diesen Effekt zu minimieren, wurde die Reihenfolge der Videos systematisch variiert, sodass zwischen den beiden fachspezifischen Beurteilungen ausreichend weitere Unterrichtsstunden beurteilt werden mussten. Die einzelnen Bewertungen erfolgten somit unabhängig und nicht in Relation zum anderen Fach. Der Primacy-Recency-Effekt kann jedoch auch dann auftreten, wenn die ersten Eindrücke einer Unterrichtsstunde bzw. einer Lehrperson einen entscheidenden Einfluss auf die Wahrnehmung der Rater haben und so deren Beurteilungen maßgeblich beeinflussen.

## 12.2 Die Erhebung der Zielvariablen in Mathematik und Rechtschreiben

Nachdem der Aufbau des hoch inferenten Ratingsystems und das Vorgehen zur Ermittlung der Beobachterübereinstimmung erläutert wurde, gilt es nun, die Leistungstests und Schülerfragebögen, die für die Forschungsfragen der Studie 3 relevant sind, vorzustellen. Hierbei handelt es sich einerseits um die jeweiligen Leistungstests in den Fächern Deutsch und Mathematik und andererseits um die Fragebögen zur Erfassung der individuellen fachspezifischen Selbstkonzepte der Schüler.

### 12.2.1 Leistungen in Mathematik

Die mathematischen Leistungen beruhen auf Ergebnissen standardisierter Leistungstests (vgl. Karst & Lipowsky, 2013; Künsting, Post, Greb, Faust & Lipowsky, 2010; Lipowsky, Faust & Greb, 2009), die unterschiedliche mathematische Kompetenzen erfassen. Der Schwerpunkt der Leistungstests zu den jeweiligen Messzeitpunkten lag insbesondere auf dem Lösen von technischen Aufgaben aus dem Inhaltsbereich der Arithmetik (Zahlen und Operationen). Zu allen drei Messzeitpunkten wurde das Verfahren zum Einsatz des Leistungstests vereinheitlicht und standardisiert, d. h. dass die Tests von geschulten Testleitern durchgeführt wurden. Zum ersten Messzeitpunkt (T1, Anfang 1. Schuljahr) wurde der Test in Form eines 30-minütigen Einzelinterviews durchgeführt, die Testungen zu T2 (Ende 1. Schuljahr) und zu T3 (Ende 2. Schuljahr) fanden in Form von 40-minütigen Gruppentests im Klassenverband statt.

Die Aufgaben zur Erfassung der arithmetischen Grundkenntnisse und Fähigkeiten zu Beginn des ersten Schuljahres (T1) orientierten sich am TEDI-MATH-Test (vgl. Kaufmann, Nuerk, Graf, Krinzinger, Delazer & Willmes, 2009) und am LEst4-7-Test<sup>80</sup> (vgl. Moser, Berweger & Lüchinger-Hutter, 2004), während sich die Testversionen zum Ende des ersten und zum Ende des zweiten Schuljahres (T2 und T3) an den DEMAT 1 und 2 anlehnten (vgl. Krajewski, Küspert & Schneider, 2002; Krajewski, Liehm & Schneider, 2004). Aufgabenbeispiele sind: „Drei plus/und fünf, wie viel ergibt das insgesamt?“ (T1), „Rechne:  $15 - \dots = 8$ “ (T2) oder „Rechne:  $72 : 8 \dots$ “ (T3).

Um den Lernfortschritt der Schüler über die drei Messzeitpunkte mittels eines Wertes ausdrücken zu können, wurde die Skala *arithmetische Grundkenntnisse und Fähigkeiten* mittels einer Raschskalierung auf der Annahme

80 Inzwischen ist das Instrument weiterentwickelt und unter dem Titel „wortgewandt & zahlenstark“ (vgl. Moser & Berweger, 2007) veröffentlicht. Verfügbar unter: <http://www.ibe.uzh.ch/projekte/projektealt/entwicklungalt/LEst4-7/LesT-Flyer.pdf>, Zugriff: 31.01.2013).

eines eindimensionalen dichotomen Raschmodells (1PL) erzeugt (vgl. Rost, 2004). Die Grundlage für ein solches Vorgehen bilden die dichotom kodierten Antworten der Schüler (0 = falsche oder keine Lösung / 1 = richtige Lösung) aus den drei Mathematikleistungstests. Die Verknüpfung zwischen den Messzeitpunkten erfolgt über sogenannte Ankeritems.<sup>81</sup>

Die personenspezifische Veränderung wurde mittels virtueller Personen geschätzt (Hartig & Kühnbach, 2006). Als Personenfähigkeitsparameter wurde für jeden Schüler der WLE-Parameter (Weighted Likelihood Estimate) ausgelesen (vgl. Warm, 1989). Diese liefern die reliabelsten individuellen Schätzwerte (vgl. Hartig & Kühnbach, 2006). Für jeden Schüler wurde somit seine individuelle Fähigkeit in Abhängigkeit von seinem Antwortmuster und den Itemschwierigkeiten, korrigiert um einen Messfehleranteil, ermittelt. Das genaue Verfahren wird von Karst und Lipowsky (2013) detailliert beschrieben. Anschließend sind die deskriptiven Statistiken für die zwei relevanten Messzeitpunkte aufgeführt (vgl. Tabelle 15).

Tabelle 15: Deskriptive Statistiken der arithmetischen Grundkenntnisse und Fähigkeiten (Ende 1. und Ende 2. Schuljahr)

	Min	Max	MW	SD
Arithmetische Kompetenz Ende 1. SJ (Z-Score)	-1.59	2.10	.13	.61
Arithmetische Kompetenz Ende 2. SJ (Z-Score)	-.89	3.03	.87	.66

Anmerkung:  $N = 604$ , SJ = Schuljahr, Min = Minimum, Max = Maximum, MW = Mittelwert, SD = Standardabweichung

### 12.2.2 Leistungen im Rechtschreiben

Im Kontext der Studie 3 werden nicht nur die Leistungen im Fach Mathematik (2. Schuljahr) als abhängige Variablen herangezogen, sondern auch der Bereich Rechtschreiben (1. Schuljahr). Hier werden entsprechende Daten zum Anfang (T1) und zum Ende des ersten Schuljahres (T2) einbezogen. Zum Anfang des ersten Schuljahres (T1) liegen Daten zur phonologischen Bewusstheit und zu schriftsprachlichen Vorkenntnissen, sogenannte *schriftspracherwerbsspezifische Vorläuferfähigkeiten*, der Schüler vor. Um die schriftspracherwerbsspezifischen Vorläuferfähigkeiten in Form der phonologischen Bewusstheit erfassen zu können, wurden die Untertests *Anlaute*

81 Hierbei handelt es sich um Testaufgaben, welche mindestens zu zwei aufeinander folgenden Messzeitpunkten eingesetzt wurden. Zum einen waren die entsprechenden Aufgaben für den jeweiligen Messzeitpunkt curricular valide und zum anderen lagen sie eher über dem Leistungsniveau der Schüler, damit sie für den darauffolgenden Messzeitpunkt nicht zu leicht sind (vgl. Karst & Lipowsky, 2013).

*hören und Laute sprechen* des LEst 4–7 (Moser, Berweger & Lüchinger-Hutter, 2004) verwendet (vgl. Lotz, Schoreit & Kempfer, 2013). Um die spezifischen schriftsprachlichen Vorkenntnisse zu ermitteln, wurden die Untertests *Buchstaben lesen* und *Silben lesen* eingesetzt (vgl. Gresser, Pohl, Corvacho del Toro, Greb & Faust, 2009). Aus Gründen der leichteren Interpretierbarkeit der Daten sind in Tabelle 16 die Lösungshäufigkeiten in Prozent dargestellt. Für die weiteren Analysen wurden die z-standardisierten WLE-Personenparameter (querschnittlich) verwendet.

Mithilfe einer explorativen Faktorenanalyse wurde zunächst überprüft, dass die zuvor genannten Untertests einen gemeinsamen Faktor bilden. Daraufhin wurde die Reliabilität ermittelt. Es ergab sich ein Cronbachs Alpha von .87. Für jeden Schüler wurde anschließend ein individueller Faktorwert ausgelesen. Dieser Faktorwert kann als Indikator für die Vorläuferfähigkeiten im Schriftspracherwerb des jeweiligen Schülers betrachtet werden. In die nachfolgenden Analysen wurden diese Faktorwerte der Schüler als Voraussetzungen für die allgemeine Lesefähigkeit und für die Rechtschreibleistung einbezogen (vgl. Kapitel 16.1).

Tabelle 16: Deskriptive Statistiken der schriftspracherwerbsspezifischen Vorläuferfähigkeiten Anfang 1. Schuljahr

	<i>MW</i>	<i>SD</i>
Anlaute hören	71.14	27.87
Laute sprechen	23.51	31.29
Buchstaben lesen	53.14	30.73
Silben lesen	24.16	36.75

*Anmerkung:* Abgebildet sind die Lösungshäufigkeiten in Prozent;  $N = 711$ , *MW* = Mittelwert, *SD* = Standardabweichung

Die Rechtschreibleistungen am Ende des ersten Schuljahres (T2) wurden mithilfe eines standardisierten Verfahrens erhoben. Zur Erfassung der *Rechtschreibleistungen* der Schüler am Ende des ersten Schuljahres wurde der DERET 1–2+ (Deutscher Rechtschreibtest für das erste und zweite Schuljahr, vgl. Stock & Schneider, 2008) in adaptierter Form als Gruppentest (ca. 45 Minuten) eingesetzt. Der erste Teil des Tests besteht aus einem Fließtext (Version A). Dieser kurze Text wurde mithilfe einer standardisierten CD-Aufnahme abgespielt. Unter Anleitung von geschulten Datenerhebenden schrieben die Schüler z. B. folgenden Satz auf: „Morgen wird ein schöner Tag“. Aus Gründen der leichteren Interpretierbarkeit gibt die Tabelle 17 den prozentualen Anteil richtig geschriebener Wörter wieder. Für die weiteren Analysen wurden die z-standardisierten WLE-Personenparameter (querschnittlich) verwendet.

Tabelle 17: Deskriptive Statistiken der Rechtschreibleistungen Ende 1. Schuljahr

	<i>MW</i>	<i>SD</i>
Rechtschreibleistungen	55,78	22,54

*Anmerkung:* Abgebildet ist die Anzahl richtig geschriebener Wörter;  $N = 694$ , *MW* = Mittelwert, *SD* = Standardabweichung

### 12.2.3 Selbstkonzepte im Rechnen und Schreiben

Im Rahmen der PERLE-Studie wurde ein neu entwickeltes Instrument eingesetzt, mit dem sich die jeweiligen Selbstkonzepte für die Domänen Rechnen, Lesen und Schreiben sowie für die kreativen Tätigkeiten von Grundschulkindern über die gesamte Grundschulzeit erheben lassen (vgl. Poloczek, 2007; Poloczek, Greb & Lipowsky, 2008). Die Entwicklung des Instruments erfolgte in Anlehnung an das Modell von Marsh und Shavelson (vgl. Marsh & Shavelson, 1985; siehe auch Marsh, 1990 & Vispoel, 1995). Für die vorliegende Untersuchung werden die Selbstkonzeptdaten vom Anfang des ersten Schuljahres (T1, November 2006, circa acht Wochen nach Schulanfang), vom Ende des ersten Schuljahres (T2, Juni 2007) und vom Ende des zweiten Schuljahres (T3, Juli 2008) verwendet (vgl. Kapitel 8, Abbildung 11). Die Selbstkonzeptdaten aus dem dritten und vierten Schuljahr sind für die Fragestellungen der Studie 3 nicht relevant.

Zu den ersten beiden Messzeitpunkten (T1 und T2) wurde die Selbstkonzepterhebung mit kleinen Gruppen von ungefähr zwölf Schülern und unter Anleitung von zwei geschulten Datenerhebem durchgeführt. Am Ende des zweiten Schuljahres (T3) wurde das Selbstkonzept als Gruppenbefragung in der ganzen Klasse erhoben. Die Schüler erhielten zu allen drei Messzeitpunkten jeweils einen Fragebogen mit insgesamt 23 Items zu den unterschiedlichen Domänen: Sechs Items bezogen sich auf Rechnen, vier Items auf Schreiben und fünf Items auf Lesen; die restlichen acht Items auf kreative Tätigkeiten. Im Fragebogen liegen die einzelnen Items ungeordnet vor. Für die Domänen Rechnen, Lesen und Schreiben wurden die Items parallel formuliert (z. B. „Nun möchte ich wissen, wie gut Du beim Rechnen / beim Lesen / beim Schreiben bist“, für eine genauere Beschreibung des Instruments siehe Poloczek, 2007; Poloczek et al., 2008). Die Items und das Antwortformat für die Bereiche Rechnen und Schreiben sind in Tabelle 18 dargestellt. Das Selbstkonzept kreativer Tätigkeiten wurde unter anderem über folgendes Item erfasst: „Nun möchte ich wissen, wie gut du basteln kannst“.

Tabelle 18: Items und Antwortformate zur Erfassung der Selbstkonzepte im Rechnen und Schreiben

Item	Antwortformat
<i>Selbstkonzept Rechnen (6 Items)</i>	
Nun möchte ich wissen, wie gut du beim Rechnen bist.	sehr gut – gut – nicht so gut
Nun möchte ich wissen, ob dir das Rechnen leicht oder schwer fällt.	leicht – weder leicht noch schwer – schwer
Nun möchte ich wissen, wie viele Fehler du beim Rechnen machst.	keine Fehler – ein paar Fehler – viele Fehler
Nun möchte ich wissen, wie oft du bei Rechenaufgaben die richtige Antwort weißt.	immer – oft – manchmal
Nun möchte ich wissen, wie gut du Zahlen zusammenzählen kannst.	sehr gut – gut – nicht so gut
Nun möchte ich wissen, ob du beim Rechnen die richtige Antwort langsamer oder schneller als die anderen Kinder weißt.	schneller – genauso schnell – langsamer
<i>Selbstkonzept Schreiben (4 Items)</i>	
Nun möchte ich wissen, wie gut du beim Schreiben bist.	sehr gut – gut – nicht so gut
Nun möchte ich wissen, ob dir das Schreiben leicht oder schwer fällt.	leicht – weder leicht noch schwer – schwer
Nun möchte ich wissen, wie viele Fehler du beim Schreiben machst.	keine Fehler – ein paar Fehler – viele Fehler
Nun möchte ich wissen, wie oft du beim Schreiben weißt, wie es richtig geht.	immer – oft – manchmal

Die Schüler konnten in dem Fragebogen aus einem dreistufigen Antwortformat wählen (vgl. Abbildung 21). Die Selbstkonzeptdaten sind so kodiert, dass ein positives Selbstkonzept einen hohen Wert ergibt, d. h. die Schüler, die „sehr gut“ ankreuzten, erhielten eine 3, Schüler, die „gut“ ankreuzten, erhielten eine 2 und Schüler, die „nicht so gut“ ankreuzten, erhielten eine 1. Schüler mit einem hohen Selbstkonzept weisen demnach einen höheren Gesamtwert auf.

Die Ergebnisse konfirmatorischer Faktorenanalysen bestätigen unter Einbezug aller Selbstkonzeptitems eine vierfaktorielle Lösung mit den Faktoren Rechnen, Schreiben, Lesen und kreative Tätigkeiten (vgl. Poloczek, 2007). Es zeigte sich demnach, dass sich die vier erhobenen Domänen bereits für die ersten beiden Schuljahre differenziert erfassen lassen ( $RMSEA < .061$ ,  $TLI > .890$ ,  $CFI > .909$ ).

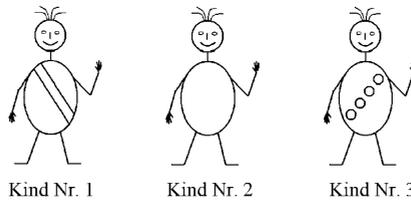


Abbildung 21: Drei Antwortmöglichkeiten im Rahmen der Selbstkonzepterhebung (vgl. Poloczec, 2007; Gabriel, Kastens, Poloczec, Schoret & Lipowsky, 2010; Gabriel, Mösko & Lipowsky, 2010)

Da für die vorliegende Studie nicht alle vier Selbstkonzeptskalen von Bedeutung sind – die Selbstkonzepte in Kunst und im Bereich Lesen werden in den weiteren Analysen nicht berücksichtigt – erfolgt im Folgenden der Überblick über die statistischen Kennwerte nur für die zwei Domänen Rechnen und Schreiben. Die zwei Skalen erwiesen sich zu allen drei Messzeitpunkten (T1, T2, T3) als ausreichend reliabel. In Tabelle 19 sind als Reliabilitätsmaße die internen Konsistenzen (Cronbachs Alpha) in den zwei Domänen zu den drei Messzeitpunkten dargestellt. Die interne Konsistenz einer Skala gibt an, inwieweit z. B. die mathematischen Selbstkonzeptitems tatsächlich das Gleiche messen. Ein Wert von  $\alpha > .70$  kann als zufriedenstellend angesehen werden. Wie Tabelle 19 zeigt, liegen alle internen Konsistenzen über diesem kritischen Wert ( $\alpha \geq .74$ ).

Neben den internen Konsistenzen sind in Unterrichtsklima und Selbstkonzept zusätzlich die Mittelwerte und Standardabweichungen für die drei Domänen Rechnen und Schreiben dargestellt. Bei Betrachtung der einzelnen Mittelwerte ist erkennbar, dass das mathematische Selbstkonzept der Schüler über die gesamte Stichprobe leicht rückläufig ist (T1 = 2.45, T3 = 2.35). Der gleiche rückläufige Trend zeigt sich für das Selbstkonzept im Schreiben (vgl. auch Poloczec, Greb & Lipowsky, 2009b; Gabriel et al., 2010; Kastens, Gabriel & Lipowsky, 2013).

Tabelle 19: Interne Konsistenzen (Cronbachs Alpha) sowie Mittelwerte und Standardabweichungen (in Klammern) für die Selbstkonzeptskalen Rechnen und Schreiben zu drei Messzeitpunkten

Selbstkonzept im ...	T1	T2	T3
Rechnen	$\alpha = .82$ (MW = 2.45, SD = .46)	$\alpha = .83$ (MW = 2.40, SD = .42)	$\alpha = .87$ (MW = 2.35, SD = .47)
Schreiben	$\alpha = .74$ (MW = 2.45, SD = .46)	$\alpha = .74$ (MW = 2.41, SD = .44)	$\alpha = .77$ (MW = 2.34, SD = .47)

Anmerkung:  $\alpha$  = Cronbachs Alpha, MW = Mittelwert, SD = Standardabweichung, T1 = Anfang 1. Schuljahr, T2 = Ende 1. Schuljahr, T3 = Ende 2. Schuljahr

Für die weiterführenden Analysen wurden die Faktorwerte aus der längsschnittlichen Skalierung ausgelesen (vgl. Kastens et al., 2013). Für alle Skalen ist die partielle Skaleninvarianz gewährleistet (Fit-Indices partielleskaleninvariantes-Modell:  $\chi^2 = 246.229$ ,  $df = 128$ ,  $p < .001$ ;  $RMSEA = .033$ ;  $CFI = .976$ ;  $TLI = .982$ ): Für alle Model-Fit-Indizes weicht das invariante Modell nicht mehr als 0.01 Punkte vom Baseline-Modell ab (Chen, 2007).

Für die Analysen in der Studie 3 (vgl. Kapitel 16) werden für den Bereich Rechnen lediglich die Daten vom Ende des ersten (T2) und vom Ende des zweiten Schuljahres (T3) herangezogen. Im Bereich Rechtschreiben wird auf Daten vom ersten Messzeitpunkt Anfang des ersten Schuljahres (T1) und Ende des ersten Schuljahres (T2) zurückgegriffen.

## IV Ergebnisse

Die Daten, die mithilfe des hoch inferenten Ratingsystems gewonnen wurden, werden in den folgenden Kapiteln mit Bezug auf die Forschungsfragen (vgl. Kapitel 7) ausgewertet. In einem ersten Schritt wird zunächst überprüft, ob die einzelnen Merkmale der Klassenführung und des Unterrichtsklimas in den beiden Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) reliabel durch die zwei Rater eingeschätzt werden konnten (Kapitel Unterrichtsklima und Selbstkonzept).

Aufbauend auf dieser Reliabilitätsprüfung erfolgt eine allgemeine Beschreibung der Ausprägung der reliabel erfassten Merkmale der Klassenführung und des Unterrichtsklimas (Kapitel 14) für beide Fächer. In einem nächsten Schritt soll die Struktur der Klassenführung und des Unterrichtsklimas im Anfangsunterricht, also die Zusammenhänge zwischen den einzelnen Aspekten der Klassenführung und des Unterrichtsklimas mithilfe von konfirmatorischen und explorativen Faktorenanalysen überprüft werden. Die Ergebnisse der fachunspezifischen und fachspezifischen Faktorenanalysen sind in Kapitel 15 dargestellt. Abschließend werden die Ergebnisse der Mehrebenenregressionsmodelle in *MPlus* vorgestellt, mit deren Hilfe die Bedeutung der Klassenführung und des Unterrichtsklimas für die Leistungen und Selbstkonzepte im Anfangsunterricht in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) aufgezeigt werden (vgl. Kapitel 16).

Für alle folgenden Berechnungen wurde das Signifikanzniveau auf 5 % festgelegt, was zur Folge hat, dass die Hypothesen erst dann als nicht bestätigt gelten, wenn die Irrtumswahrscheinlichkeit  $p$  kleiner oder gleich 5 % ist (vgl. Bortz, 2005). Dies muss im Vorfeld festgelegt werden „um eine gewisse Vergleichbarkeit und Qualität statistisch abgesicherter Entscheidungen zu gewährleisten“ (Bortz, 2005, S. 113).

### 13 Ergebnisse der Inter-Rater-Reliabilität

Um dem Gütekriterium der Reliabilität hoch inferenter Daten Rechnung zu tragen, wurde im Rahmen der Arbeit die Inter-Rater-Reliabilität berechnet. Zur Überprüfung der Beobachter-übereinstimmung wurde der relative Generalisierbarkeitskoeffizient (G-Koeffizient) als Übereinstimmungsmaß herangezogen (vgl. Kapitel 6.3.2). Mithilfe dieses Reliabilitätsmaßes sollen die ersten beiden Forschungsfragen beantwortet werden: (1) Lassen sich die

fachunspezifisch operationalisierten Merkmale der Klassenführung und des Unterrichtsklimas unabhängig vom Fach reliabel mithilfe von hoch inferenten Schätzverfahren erfassen? (2) Zeigen sich Unterschiede in der Inter-Rater-Reliabilität in Abhängigkeit vom Unterrichtsfach?

Die Berechnung des relativen G-Koeffizienten erfolgte nach dem Ansatz der Generalisierbarkeitstheorie (vgl. Kapitel 6.3.2) und in Anlehnung an die Arbeiten von Clausen et al. (2003) sowie in Bezug auf Renkl und Helmke (1993) mithilfe des Programms GT (vgl. Ysewijn, 1997). Um Aussagen über das Ausmaß der merkmalsunabhängigen Varianz bzw. über die systematische Fehlervarianz machen zu können, wurde für jedes der 18 hoch inferenten Merkmale der Klassenführung und des Unterrichtsklimas eine 2-Facetten-G-Studie berechnet, in der die Facetten Video und Rater berücksichtigt wurden. Entsprechend wurden die Rater sowie der beobachtete Unterricht (= Video) in ein varianzanalytisches Modell<sup>82</sup> aufgenommen. Da beide Rater alle Unterrichtssequenzen (bzw. -stunden) hinsichtlich der Höhe der Ausprägungen der einzelnen Merkmale eingeschätzt haben, spricht man von einem gekreuzten Design (Crossed Design<sup>83</sup>). Mithilfe dieser Berechnung sollen die Forschungsfragen 3.1 und 3.2 beantwortet und somit untersucht werden, wie groß der Varianzanteil ist, der auf Unterschiede in den Beurteilungen durch die zwei Rater zurückzuführen ist (= systematische Fehlervarianz) bzw. ob die Erfassung der Merkmale durch die Subjektivität der Rater systematisch verfälscht ist. Zeigen sich Unterschiede in der Höhe der systematischen Fehlervarianz in Abhängigkeit vom Unterrichtsfach?

Die Berechnung der G-Koeffizienten und der Varianzkomponenten erfolgte nach dem Rating von 20 %, 25 %, 50 %, 75 % und 100 % aller Unterrichtsvideos.<sup>84</sup> Anschließend werden die Ergebnisse der G-Studien einerseits fachunspezifisch (Kapitel Tabelle 16: Deskriptive Statistiken der schriftspracherwerbsspezifischen Vorläuferfähigkeiten Anfang 1. Schuljahr) und andererseits fachspezifisch (getrennt für Deutsch und Mathematik, Kapitel 13.2) dargestellt.

---

82 Zur Schätzung der Varianzkomponenten wurde auf die ANOVA-Methode (= varianzanalytisches Verfahren) zurückgegriffen (vgl. Stumpp & Großmann, 2009). Durch die Verwendung der ANOVA-Methode wurden zu mehreren Messzeitpunkten die Quadratsummen in SPSS Version 20 berechnet und im Anschluss in das GT-Programm (vgl. Ysewijn, 1997) eingespeist.

83 Im Unterschied dazu spricht man von einem genesteten Design (Nested Design), wenn jeder Rater nur einen Teil der Unterrichtssequenzen einschätzt. In diesem Fall sind die Unterrichtssequenzen in den Ratern genestet.

84 Obwohl die Anzahl der Videos nach Hugener und Kollegen (2006) bei den ersten beiden Übereinstimmungsberechnungen (nach 20% bzw. 25 % der Videos) für eine Berechnung des G-Koeffizienten zu gering sind, wurde bereits zu diesen Messzeitpunkten ein erster Blick auf die Übereinstimmung der Rater und die Verteilung der Varianzkomponenten geworfen.

### 13.1 Fachunspezifische Berechnung der Inter-Rater-Reliabilität

In Tabelle 20 sind die Ergebnisse der Generalisierbarkeitsstudie nach dem Rating von  $N = 97$  Deutsch- und Mathematikvideos dargestellt. Die vier Schulungsvideos wurden für die Berechnung nicht herangezogen.

In der Tabelle geben die Werte der Spalte „VK Video“ an, welcher Anteil der Varianz auf tatsächliche Unterschiede zwischen den Unterrichtsvideos und damit auf Unterschiede im beobachteten Unterricht zurückzuführen ist (= „wahre“ Varianz). Die Spalte „VK Rater“ macht Aussagen über den relativen Anteil der Varianz, die durch charakteristische Unterschiede in den Einschätzungen der Stunden durch die zwei Rater zustande gekommen sind, d. h. sie gibt die merkmalsunabhängige Varianz an (= systematische Fehlervarianz). In einer zusätzlichen Spalte ist die unsystematische Variation angegeben. Diese ergibt sich durch die Interaktion der beiden systematischen Variationen (Video und Rater) sowie einem Fehlerterm (VK V x R + Residuum). Der relative G-Koeffizient gibt in einer zusätzlichen Spalte an, inwieweit sich eine gemessene Varianz auf tatsächliche Unterschiede zwischen den Unterrichtsvideos zurückführen lässt und nicht auf Unterschiede in der Beurteilungspraxis der beiden Rater. Auf Basis dieses relativen G-Koeffizienten kann entschieden werden, ob die Einschätzungen der einzelnen Merkmale der Klassenführung und des Unterrichtsklimas ausreichende Generalisierbarkeit (Zuverlässigkeit) aufweisen. In der vorliegenden Arbeit wird ein G-Koeffizient  $\geq .65$  als Kriterium für eine zufriedenstellende Qualität der hoch inferenten Daten festgelegt (vgl. Kapitel 6.3.2). Dieser muss für die Verwendung der Daten in weiterführenden Analysen gewährleistet sein. Merkmale, die dem Kriterium G-Koeffizient  $\geq .65$  nicht standhalten, sind in Tabelle 20 grau unterlegt.

Wie aus der Tabelle 20 ersichtlich wird, erreichen die Merkmale „positiver Umgang der Lehrperson mit Fehlern (Fehl)“ und „positives Fehlerklima zwischen den Schülern (Fekl)“ in den Ratings der Deutsch- und Mathematikstunden das gesetzte Kriterium von  $g \geq .65$  nicht. Die geringen G-Koeffizienten bei diesen zwei Merkmalen des Unterrichtsklimas können einerseits auf eventuelle konzeptionelle Unschärfen dieser Konstrukte zurückgeführt werden, andererseits weisen die geringen Varianzen zwischen den Videos (VK Video = 28 % bzw. 31 %) darauf hin, dass sich die gefilmten Lektionen kaum unterschieden bzw. die Merkmale für die Beurteiler weniger deutlich sichtbar waren und somit ein höheres Maß an Interpretation erforderlich wurde. Damit kann auch für die Grundschule bestätigt werden, dass eine öffentliche Thematisierung von Fehlern bzw. eine negative Reaktion auf Fehler eher selten in Unterrichtsvideos zu beobachten ist (vgl. Kapitel 4.2). Die restlichen Merkmale der Klassenführung und des Unterrichtsklimas konnten über alle Deutsch- und Mathematikvideos hinweg von den beiden Ratern reliabel eingeschätzt werden. Allerdings erreichen die Merkmale „Vorbereitungen des

Klassenraums (Vorb)“ und „Wettbewerb und Leistungsdruck (Wett)“ nur einen relativen G-Koeffizienten von .65 bzw. .68. Die Höhe der übrigen relativen G-Koeffizienten variiert zwischen .81 und .92.

Tabelle 20: Varianzkomponenten und relative G-Koeffizienten für die einzelnen Merkmale der Klassenführung und des Unterrichtsklimas ( $N_{VSS+VSM} = 97$ )

	g (relativ)	Varianzkomponenten (VK)*		
		zwischen Videos (VK Video)	zwischen Ratern (VK Rater)	Residualanteil (VK V x R + Res)
<b>Klassenführung</b>				
Allgegenwärtigkeit der Lehrperson (Allg)	.89	79 %	0 %	20 %
Gruppenfokus (Grup)	.88	79 %	0 %	21 %
Zeitmanagement / effektive Zeitznutzung (Zeit)	.92	86 %	0 %	14 %
Übergangsmangement (Managing Transition) (Ueb)	.90	82 %	0 %	18 %
Regelklarheit bzw. -verwendung (Regl)	.87	77 %	0 %	23 %
Störungsfreiheit (Disz)	.81	67 %	1 %	32 %
effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)	.86	74 %	0 %	25 %
Vorbereitung des Klassenraums (Vorb)	.65	42 %	0 %	58 %
<b>Unterrichtsklima</b>				
humorvolle Lernatmosphäre (Humo)	.81	65 %	0 %	34 %
Anerkennung der Schüler durch die Lehrperson (Anes)	.86	75 %	0 %	24 %
gegenseitige Anerkennung der Schüler (Aneg)	.82	68 %	1 %	30 %
Fürsorglichkeit, Herzlichkeit und Wärme (Fuer)	.88	78 %	1 %	22 %
Einsatz von Lob (Lob)	.90	77 %	5 %	18 %
Ermunterung durch die Lehrperson (Emut)	.88	78 %	0 %	22 %
sachlich-konstruktive Rückmeldung (Rück)	.83	70 %	2 %	28 %
positiver Umgang der Lehrperson mit Fehlern (Fehl)	.46	28 %	7 %	68 %
positives Fehlerklima zwischen den Schülern (Fekl)	.48	31 %	2 %	67 %
Wettbewerb und Leistungsdruck (Wett)	.68	51 %	0 %	49 %

Anmerkung: g = relativer G-Koeffizient

Grau hinterlegte G-Koeffizienten konnten, nach dem Rating aller Deutsch- und Mathematikvideos, dem Kriterium  $\geq .65$  nicht standhalten.

\* Da das Programm GT nur ganze Prozentzahlen angibt, kommt es vermutlich zu Rundungsfehlern, die eine Abweichung in der Gesamtprozentzahl erklären können. Aus diesem Grund ergeben die Varianzkomponenten nicht bei jedem Merkmal in der Summe 100 %, sondern weichen um  $\pm 1$  % ab.

Bei einer näheren Betrachtung der einzelnen relativen G-Koeffizienten wird deutlich, dass diese zwischen den Merkmalen der Klassenführung und des Unterrichtsklimas variieren. Die Übereinstimmungen der Rater sind für die Merkmale der Klassenführung etwas höher ( $.65 \leq g \leq .92$ ,  $MW = .85$ ) als für die Merkmale des Unterrichtsklimas ( $.46 \leq g \leq .90$ ,  $MW = .72$ ). Die höhere mittlere Übereinstimmung bei den Merkmalen der Klassenführung liegt vermutlich daran, dass die Beurteilung der Klassenführungsmerkmale weniger Anforderungen an die Rater stellt als die Einschätzung des Unterrichtsklimas. Die Merkmale der Klassenführung lassen sich demnach besser übereinstimmend einschätzen. Dies zeigte sich für die Sekundarstufe auch bereits in der TIMSS-Videostudie (vgl. Clausen et al., 2003).

Ein zusätzlicher Blick auf die relative Verteilung der Varianzkomponenten macht deutlich, dass die größte Varianz bei allen Merkmalen, die das Kriterium  $g \geq .65$  erfüllen, auf tatsächliche Unterschiede zwischen dem beobachteten Unterricht zurückzuführen ist (VK Video,  $Min = 42\%$ ,  $Max = 86\%$ ). Dabei weist das Merkmal „Vorbereitung des Klassenraums (Vorb)“ die geringsten Unterschiede zwischen den Videos auf (VK Video =  $42\%$ ), gefolgt von dem Merkmal „Wettbewerb und Leistungsdruck (Wett)“, was unter anderem ein Grund für den niedrigeren relativen G-Koeffizienten ist. Die hohen Varianzanteile der Komponente Video sind besonders wünschenswert (vgl. Kapitel 7). Je höher der Varianzanteil der Komponente Video, desto stärker spiegeln sich in den Ratings der Beurteiler Unterschiede zwischen dem Unterricht in den beobachteten Klassen wider. Die Variation in der Bestimmung der beiden Unterrichtsqualitätsmerkmale ist demnach abhängig von tatsächlichen qualitativen Unterschieden im beobachteten Unterrichtsgeschehen.

Die eher geringe relative Varianz auf der Komponente Rater (vgl. Tabelle 20, Spalte VK Rater) bedeutet, dass der relative Anteil der Varianz, die durch Unterschiede in den Einschätzungen aller Unterrichtsvideos durch die zwei Rater zustande gekommen ist (= Beurteilungsfehler), so gering ist, dass sie nach Pietsch (2008) eine fast vernachlässigbare Fehlerquelle in der Bestimmung der beiden Merkmale von Unterrichtsqualität darstellt. Die Variation liegt auch deutlich unterhalb der durch Hoyt und Kerns (1999) – im Rahmen einer Metaanalyse – festgestellten Fehlervarianz von 37 Prozent, aber auch deutlich unter der berichteten Fehlervarianz von Hoyt (2000), der für geschulte Beurteiler eine Fehlervarianz von 14 Prozent berichtete (vgl. Kapitel 7). Die Varianz, die durch die Interaktion zwischen den Videos und den Ratern ( $V \times R + \text{Residuum}$ ) aufgeklärt wird (die unsystematische Variation), schwankt zwischen 14 Prozent und 49 Prozent, wobei das Merkmal „Vorbereitung des Klassenraums (Vorb)“ die größte Varianz aufweist ( $58\%$ ).

Zusammenfassend konnte bei der fachunspezifischen Berechnung der 2-Facetten-G-Studien gezeigt werden, dass 16 von 18 Merkmalen der Klassenführung und des Unterrichtsklimas einen G-Koeffizienten zwischen  $.65$  und  $.92$  erreichen und somit auch für den Anfangsunterricht verlässliche Aussagen über die zwei Basisdimensionen der Unterrichtsqualität auf Grundlage

hoch inferenter Ratings generiert werden können. Allerdings wurde für diese Berechnung das gesamte Material (Deutsch- und Mathematikvideos) herangezogen. Inwieweit Unterschiede bei der Berechnung solcher G-Studien in Abhängigkeit vom Unterrichtsfach existieren, wird anschließend explorativ untersucht.

## 13.2 Fachspezifische Berechnung der Inter-Rater-Reliabilität

Dieses Kapitel dient der Beantwortung der Forschungsfrage Nr. 2. Es soll der Frage nachgegangen werden, ob sich Unterschiede in der Inter-Rater-Reliabilität in Abhängigkeit vom Unterrichtsfach zeigen. Tabelle 21 stellt die relativen G-Koeffizienten für die Merkmale der Klassenführung und des Unterrichtsklimas getrennt für die zwei Videostudien dar. Für die fachbezogene Berechnung des G-Koeffizienten betrug die Anzahl der Fälle für die Videostudie Deutsch 48 und für die Videostudie Mathematik 49 Unterrichtsvideos. Die Schulungsvideos wurden erneut nicht für die Berechnung herangezogen.

Für die Merkmale des Unterrichtsklimas zeigen sich für das Fach Mathematik (2. Schuljahr) relative G-Koeffizienten zwischen .39 und .91. Auffällig sind hier vor allem die Merkmale „positiver Umgang der Lehrperson mit Fehlern (Fehl)“, „positives Fehlerklima zwischen den Schülern (Fekl)“ und „Wettbewerb und Leistungsdruck (Wett)“. Die niedrigen relativen G-Koeffizienten für diese Merkmale können vermutlich darauf zurückgeführt werden, dass diese im Fach Mathematik (2. Schuljahr) für die Rater weniger eindeutig sichtbar waren als im Fach Deutsch (1. Schuljahr). Anscheinend bringen diese Merkmale für das Fach Mathematik (2. Schuljahr) ein höheres Ausmaß an Interpretation mit sich. Im Fach Deutsch (1. Schuljahr) weist lediglich das Merkmal „positiver Umgang der Lehrperson mit Fehlern (Fehl)“ einen ähnlich niedrigen relativen G-Koeffizienten auf. Demnach konnte das Merkmal „positiver Umgang der Lehrperson mit Fehlern (Fehl)“ in beiden Videostudien nicht reliabel durch die zwei Rater eingeschätzt werden.

Deutliche Unterschiede in der Höhe der relativen G-Koeffizienten zwischen den beiden Videostudien können für die Merkmale „Zeitmanagement / effiziente Zeitnutzung (Zeit)“, „Übergangsmanagement (Managing Transition) (Ueb)“, „Regelklarheit bzw. -verwendung (Regl)“ und „Vorbereitung des Klassenraums (Vorb)“ bestätigt werden (vgl. Tabelle 21, fett gedruckt). Das Merkmal „Vorbereitung des Klassenraums (Vorb)“ offenbarte in den Ratings für die Videostudie Mathematik (VSM) nahezu keine Varianz zwischen den Videos (vgl. Tabelle 22). Aus diesem Grund konnte der relative G-Koeffizient für die Videos im Fach Mathematik nicht berechnet werden, während das Merkmal für das Fach Deutsch (1. Schuljahr) reliabel durch die zwei Rater eingeschätzt werden konnte ( $g = .67$ ).

Tabelle 21: Relative G-Koeffizienten für die einzelnen Merkmale der Klassenführung und des Unterrichtsklimas (getrennt nach Videostudie)

	Videostudie	
	VSS (N = 48) g (relativ)	VSM (N = 49) g (relativ)
<b>Klassenführung</b>		
Allgegenwärtigkeit der Lehrperson (Allg)	.90	.88
Gruppenfokus (Grup)	.82	.92
Zeitmanagement / effektive Zeitnutzung (Zeit)	<b>.66</b>	<b>.98</b>
Übergangsmanagement (Managing Transition) (Ueb)	<b>.79</b>	<b>.94</b>
Regelklarheit bzw. -verwendung (Regl)	<b>.68</b>	<b>.95</b>
Störungsfreiheit (Disz)	.86	.77
effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)	.85	.86
Vorbereitung des Klassenraums (Vorb)	.67	–*
<b>Unterrichtsklima</b>		
humorvolle Lernatmosphäre (Humo)	.80	.81
Anerkennung der Schüler durch die Lehrperson (Anes)	.90	.81
gegenseitige Anerkennung der Schüler (Aneg)	.85	.81
Fürsorglichkeit, Herzlichkeit und Wärme (Fuer)	.84	.91
Einsatz von Lob (Lob)	.90	.89
Ermutung durch die Lehrperson (Emut)	.93	.78
sachlich-konstruktive Rückmeldung (Rück)	.83	.81
positiver Umgang der Lehrperson mit Fehlern (Fehl)	.44	.53
positives Fehlerklima zwischen den Schülern (Fekl)	<b>.79</b>	<b>.39</b>
Wettbewerb und Leistungsdruck (Wett)	<b>.80</b>	<b>.61</b>

Anmerkung: g = relativer G-Koeffizient, VSS = Videostudie Deutsch (1. Schuljahr), VSM = Videostudie Mathematik (2. Schuljahr)

Grau hinterlegte G-Koeffizienten konnten dem Kriterium  $g \geq .65$  nicht standhalten. Fett gedruckte G-Koeffizienten sollen größere Abweichungen zwischen den Fächern hervorheben.

\* Das Merkmal „Vorbereitung des Klassenraums (Vorb)“ wies in den Ratings für die Videostudie Mathematik (VSM) zu wenig Varianz auf, um den G-Koeffizienten zu berechnen (vgl. auch Tabelle 22).

Für die Merkmale, die hohe relative G-Koeffizienten aufweisen, lässt sich durchgehend für beide Fächer ein günstiges Verhältnis der „wahren“ Variation (VK Video) zu den systematischen Raterunterschieden (VK Rater) und den unsystematischen Fehlervarianzen (VK Video x Rater + Res) feststellen (vgl. Unterrichtsklima und Selbstkonzept). Unterschiede zwischen den beiden Videostudien in der Höhe der relativen G-Koeffizienten lassen sich auf ein ungünstiges Verhältnis der eben erwähnten Varianzkomponenten in einer der beiden Videostudien zurückführen. Beispielsweise zeigt sich dieses ungünstige Verhältnis in der Videostudie Deutsch (1. Schuljahr) für die Merkmale „Zeitmanagement / effektive Zeitnutzung (Zeit)“, „Übergangsmanagement

(Managing Transition) (Ueb)“ und „Regelklarheit bzw. -verwendung (Regl)“. Bei der Einschätzung des Merkmals „Vorbereitung des Klassenraums (Vorb)“ konnte keine nennenswerte Variation zwischen den Mathematikvideos festgestellt werden, während für die Videostudie Deutsch 53 Prozent der Varianz auf die tatsächlichen Unterschiede zwischen den Unterrichtsvideos zurückzuführen sind. Sobald ein Merkmal ein solches Missverhältnis der Varianzkomponenten aufweist, resultiert automatisch eine schlechtere Modellgüte, die sich wiederum in Form eines niedrigen relativen G-Koeffizienten äußert.

Für die Merkmale des Unterrichtsklimas ergibt sich bei der fachspezifischen Betrachtung der Varianzkomponenten ein ähnliches Bild. Ein Teil der hoch inferenten Merkmale des Unterrichtsklimas zeigen in beiden Fächern ein eher ungünstiges Verhältnis zwischen der „wahren“ Varianz und der systematischen und unsystematischen Fehlervarianz.

Hervorzuheben ist an dieser Stelle, dass in beiden Videostudien der relative Anteil der Varianz, die durch Unterschiede in den Einschätzungen der Stunden durch die zwei Rater zustande gekommen ist (VK Rater), für einen Großteil der hoch inferenten Merkmale sehr gering ausfällt. Die einzige Ausnahme stellt das Merkmal „positiver Umgang der Lehrperson mit Fehlern (Fehl)“ in der Videostudie Mathematik (2. Schuljahr) dar. Die 18 Prozent, die bei der Einschätzung dieses Merkmals durch systematische Raterunterschiede erklärt werden, bedeuten, dass hier die Messergebnisse durch mögliche Beurteilereffekte verzerrt sein können, womit wiederum der sehr niedrige relative G-Koeffizient erklärt werden kann.

Insgesamt lässt sich an dieser Stelle festhalten, dass die Beurteilungen der verschiedenen Merkmale mit wenigen Ausnahmen eine gute bis sehr gute Messqualität aufweisen. Für das Fach Mathematik (2. Schuljahr) fallen im Vergleich zum Fach Deutsch (1. Schuljahr) vor allem die Merkmale der Klassenführung durch eine besonders gute Messqualität auf (Ausnahme: „Vorbereitung des Klassenraums (Vorb)“). Die fachspezifischen Ergebnisse der Verteilung der Varianzkomponenten lassen sich dahingehend interpretieren, dass für die hier zugrunde liegende Stichprobe von Unterrichtsvideos die Variabilität in den Messwerten in beiden Videostudien weitestgehend unabhängig von den Ratern, aber nicht frei von anderen Fehlerquellen ist. Dies wird durch die hohen relativen Anteile der Varianz deutlich, die auf die Interaktionen entfallen. Anschließend wurden für diejenigen Merkmale, die reliabel eingeschätzt werden konnten, die Werte über die zwei Rater gemittelt und als Gesamtscore für weitere Analysen verwendet, deren Ergebnisse nun dargestellt werden.

Tabelle 22: Varianzkomponenten für die einzelnen Merkmale der Klassenführung und des Unterrichtsklimas (getrennt nach Videostudie)

	Varianzkomponenten (VK)*					
	VSS (N = 48)			VSM (N = 49)		
	VK V	VK R	VK V x R + Res	VK V	VK R	VK V x R + Res
<b>Klassenführung</b>						
Allgegenwärtigkeit der Lehrperson (Allg)	79 %	2 %	19 %	78 %	0 %	22 %
Gruppenfokus (Grup)	69 %	1 %	30 %	86 %	0 %	14 %
Zeitmanagement / effektive Zeitnutzung (Zeit)	49 %	0 %	51 %	96 %	0 %	4 %
Übergangsmangement (Managing Transition) (Ueb)	65 %	0 %	35 %	89 %	0 %	11 %
Regelklarheit bzw. -verwendung (Regl)	51 %	1 %	47 %	90 %	0 %	10 %
Störungsfreiheit (Disz)	74 %	2 %	24 %	62 %	0 %	38 %
effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)	73 %	2 %	25 %	75 %	0 %	25 %
Vorbereitung des Klassenraums (Vorb)	53 %	0 %	47 %	0 %	0 %	100 %
<b>Unterrichtsklima</b>						
humorvolle Lernatmosphäre (Humo)	67 %	0 %	33 %	69 %	0 %	31 %
Anerkennung der Schüler durch die Lehrperson (Anes)	82 %	1 %	17 %	69 %	0 %	31 %
gegenseitige Anerkennung der Schüler (Aneg)	72 %	3 %	26 %	67 %	0 %	33 %
Fürsorglichkeit, Herzlichkeit und Wärme (Fuer)	72 %	1 %	27 %	83 %	0 %	17 %
Einsatz von Lob (Lob)	80 %	2 %	18 %	74 %	8 %	18 %
Ermütigung durch die Lehrperson (Emut)	86 %	1 %	13 %	63 %	0 %	36 %
sachlich-konstruktive Rückmeldung (Rück)	69 %	4 %	27 %	68 %	1 %	31 %
positiver Umgang der Lehrperson mit Fehlern (Fehl)	29 %	0 %	71 %	30 %	18 %	53 %
positives Fehlerklima zwischen den Schülern (Fekl)	65 %	0 %	35 %	22 %	6 %	72 %
<b>Wettbewerb und Leistungsdruck (Wett)</b>						

Anmerkung: VSS = Videostudie Deutsch (1. Schuljahr), VSM = Videostudie Mathematik (2. Schuljahr), VK V = Varianzkomponente Video, VK R = Varianzkomponente Rater, VK V x R + Res = Varianzkomponente Video x Rater + Residuum

\* Da das Programm GT nur ganze Prozentzahlen angibt, kommt es vermutlich zu Rundungsfehlern, die eine Abweichung in der Gesamtprozentzahl erklären können. Aus diesem Grund ergeben die Varianzkomponenten nicht bei jedem Merkmal in der Summe 100 %, sondern weichen um  $\pm 1$  % ab.

## 14 Ergebnisse der Studie 1 – Ausprägung der Merkmale

Mithilfe des relativen G-Koeffizienten als ein Maß für die Reliabilität der Einschätzungen der zwei Rater konnte im Vorfeld bereits gezeigt werden, dass sowohl im Fach Deutsch (1. Schuljahr) als auch im Fach Mathematik (2. Schuljahr) das Merkmal „positiver Umgang der Lehrperson mit Fehlern (Fehl)“ nicht übereinstimmend (bzw. reliabel) durch die zwei Rater beurteilt werden konnte. Im Fach Mathematik hielten zusätzlich die Merkmale „Vorbereitung des Klassenraums (Vorb)“, „positives Fehlerklima zwischen den Schülern (Fekl)“ und „Wettbewerb und Leistungsdruck (Wett)“ dem gesetzten Kriterium von  $g \geq .65$  nicht stand, womit auch diese Merkmale für einen Vergleich der beiden Videostudien nicht herangezogen werden können. Für die Analysen in der Studie 1 verringert sich die Anzahl der Merkmale somit auf 14, sieben Merkmale der Klassenführung und sieben Merkmale des Unterrichtsklimas. Im Folgenden wird näher darauf eingegangen, wie diese 14 verbliebenen Merkmale der Klassenführung und des Unterrichtsklimas im Fach Deutsch (1. Schuljahr) und im Fach Mathematik (2. Schuljahr) ausgeprägt sind. Mithilfe der hier dargestellten Ergebnisse wird die Forschungsfrage Nr. 4 beantwortet: Welche Ausprägungen zeigen die einzelnen Merkmale der Klassenführung und des Unterrichtsklimas in den videografierten Unterrichtsstunden im ersten (Fach Deutsch) und zweiten (Fach Mathematik) Schuljahr?

Zur Beantwortung der Forschungsfrage Nr. 4 sind in der Abbildung 22 die Mittelwerte<sup>85</sup> der hoch inferenten Ratings, getrennt für die beiden Videostudien Deutsch (VSS, 1. Schuljahr) und Mathematik (VSM, 2. Schuljahr), dargestellt. Datengrundlage bilden 50 Unterrichtsvideos aus der Videostudie Deutsch (VSS) und 51 aus der Videostudie Mathematik (VSM). Das Antwortformat war eine vierstufige Skala von 1 bis 4, ergänzt um die Möglichkeit, das entsprechende Merkmal bei der Beobachtung der Nebenlehrperson als nicht einschätzbar zu bewerten.<sup>86</sup> Je nach Merkmal variiert somit die Anzahl an gültigen Fällen (vgl. Kapitel 21).

Bis auf das Merkmal „humorvolle Lernatmosphäre (Humo)“, das je nach Videostudie einen Mittelwert von 1.63 bzw. 1.66 aufweist (vgl. Abbildung 22), und das Merkmal „Einsatz von Lob (Lob)“ im Fach Deutsch ( $MW = 2,23$ ) liegen die Mittelwerte der restlichen Merkmale des Unterrichtsklimas und der

85 Für die Berechnung der Mittelwerte wurde jeder Lehrperson für jede Videostudie ein Wert zugeordnet, indem der Mittelwert der vier Ratings (2 Rater x 2 Stunden) pro Merkmal gebildet wurde (vgl. Kapitel 12.1.3). Diese werden auch für alle weiteren Analysen verwendet.

86 Aufgrund dieses Vorgehens kam es für einige Nebenlehrpersonen zur Vergabe der „0“ im Sinne einer „Nichtkategorisierung“ (vgl. Kapitel 12.1.4). Die „0“ wurde anschließend als ‚Missing‘ definiert.

Klassenführung über dem Wert 2,5 (= theoretische Mitte). Die Hypothese, dass die Rater die Merkmale der Klassenführung und des Unterrichtsklimas für beide Fächer deutlich positiv einschätzen (Hypothese 1), kann demnach bestätigt werden. Die Merkmale der Klassenführung und des Unterrichtsklimas (mit Ausnahme von „humorvolle Lernatmosphäre (Humo)“) wurden von den zwei Ratern im Mittel deutlich positiv eingeschätzt. Die Häufigkeitsverteilungen und deskriptiven Statistiken für jedes einzelne hoch inferente Merkmal sind im Kapitel 21 tabellarisch dargestellt.

Wie die Abbildung 22 zeigt, sind die Merkmale der Klassenführung und des Unterrichtsklimas unterschiedlich hoch ausgeprägt. Während alle Merkmale der Klassenführung, unabhängig vom Fach, Mittelwerte über 2.99 erreichen und die Lehrpersonen durchweg deutlich positive Werte von den Ratern erhielten, zeigte sich für die Merkmale des Unterrichtsklimas ein eher uneinheitliches Bild. Neben den sehr geringen Ausprägungen für das Merkmal „humorvolle Lernatmosphäre (Humo)“ schätzen die beiden Rater im Mittel die Merkmale „Fürsorglichkeit, Herzlichkeit und Wärme (Fuer)“, „Einsatz von Lob (Lob)“, „Ermutigung durch die Lehrperson (Emut)“ und „sachlich-konstruktive Rückmeldung (Rück)“ nur zwischen 2.22 und 2.83 ein. Die Merkmale „Anerkennung der Schüler durch die Lehrperson (Anes)“ und „gegenseitige Anerkennung der Schüler (Aneg)“ sind mit Werten zwischen 3 und 4 ähnlich hoch ausgeprägt wie die Klassenführungsmerkmale.

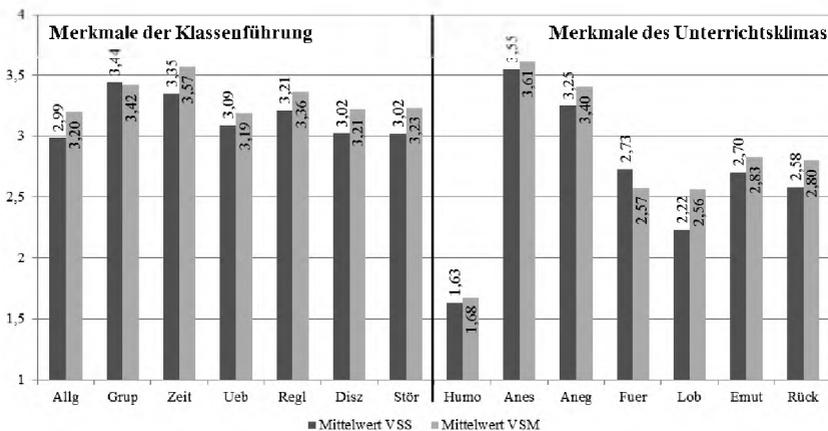


Abbildung 22: Mittelwerte getrennt für VSS (1. Schuljahr) und VSM (2. Schuljahr)

Anmerkung: VSS = Videostudie Deutsch (1. Schuljahr), VSM = Videostudie Mathematik (2. Schuljahr),  $N_{VSS} = 50$ ,  $N_{VSM} = 51$

Aus der zuvor dargestellten Deskription kann jedoch nicht geschlossen werden, ob die Unterschiede in den mittleren Ausprägungen der einzelnen Merkmale der Klassenführung und des Unterrichtsklimas zwischen den Fä-

chern bedeutsam sind. Aus diesem Grund werden im Anschluss die Unterschiede der Ausprägungen zwischen den Fächern auf ihre Signifikanz geprüft. Mithilfe dieses Vorgehens soll die Forschungsfrage Nr. 5 beantwortet werden. Um der Frage nachzugehen, ob sich im Hinblick auf die Ausprägungen der einzelnen Merkmale der Klassenführung und des Unterrichtsklimas signifikante Unterschiede zwischen den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) identifizieren lassen, wurden nur diejenigen Lehrpersonen herangezogen, für die in beiden Videostudien Werte vorlagen (vgl. Kapitel 11.1). Die Stichprobe reduzierte sich somit auf 30 Lehrpersonen bzw. 60 Videos (VSS + VSM). Signifikante Unterschiede in den Ausprägungen der Merkmale zwischen den zwei Videostudien lassen sich mit diesem Vorgehen auf das jeweilige Fach bzw. Schuljahr zurückführen.

Zur Analyse der Unterschiede zwischen der Videostudie Deutsch (VSS, 1. Schuljahr) und der Videostudie Mathematik (VSM, 2. Schuljahr) kommen univariate Varianzanalysen (*Analysis of Variance, ANOVA*) in SPSS zum Einsatz. Der Zweck der Varianzanalyse liegt darin, die Signifikanz von Mittelwertdifferenzen zu testen. In der Varianzanalyse bezeichnet man die unabhängigen Variablen als Faktoren. Im vorliegenden Fall nimmt der Faktor zwei Faktorstufen an: Videostudie Deutsch (1. Schuljahr) und Videostudie Mathematik (2. Schuljahr). Als abhängige Variable gehen die entsprechenden hoch inferenten Merkmale der Klassenführung und des Unterrichtsklimas in die Berechnung mit ein. Zusätzlich wird für jedes Merkmal der Eta-Koeffizient berichtet.<sup>87</sup> Die Ergebnisse der Signifikanzprüfung werden für die einzelnen Merkmale der Klassenführung (vgl. Tabelle 23) und des Unterrichtsklimas (vgl. Tabelle 24) dargestellt.

Die Ergebnisse der univariaten Varianzanalysen für die Merkmale der Klassenführungen (vgl. Tabelle 23) zeigen, dass sich die Ausprägungen der Klassenführungsmerkmale der Lehrpersonen nicht signifikant zwischen den Videostudien Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) unterscheiden. Die PERLE-Lehrpersonen zeigen demnach in beiden Videostudien – nach Kontrolle der Klasse – ähnliche Klassenführungs Kompetenzen, was für die Stabilität der Merkmale über die Fächer hinweg spricht.

In Bezug auf die Ausprägungen der Merkmale des Unterrichtsklimas kann lediglich für das Merkmal „Einsatz von Lob (Lob)“ tendenziell ein signifikanter Unterschied zwischen den beiden Videostudien bestätigt werden ( $F(1; 53) = 2.793$ ,  $p = .101$ ; vgl. Tabelle 24, grau hervorgehoben), allerdings wird dieser nur knapp auf einem 10 %-Niveau signifikant. Die Lehrpersonen der PERLE-Stichprobe setzen demnach im Vergleich zum Fach Deutsch

87 Ein Eta-Koeffizient von beispielsweise  $\eta^2 = .090$  bedeutet, dass für die entsprechende Dimension rund neun Prozent der Varianz von der einbezogenen abhängigen Variable simultan durch die Videostudie (VSS, 1. Schuljahr vs. VSM, 2. Schuljahr) als fester Faktor erklärt werden. Bortz und Döring (2006) sprechen von einer kleinen Effektstärke, wenn  $\eta^2 = .01$ . Bei  $\eta^2 = .06$  sprechen sie von einer mittleren und bei  $\eta^2 = .14$  von einer großen Effektstärke.

(1. Schuljahr) in Mathematik (2. Schuljahr) tendenziell im Mittel mehr Lob ein ( $M_{VSM} = 2.42$ ,  $M_{VSS} = 2.10$ ). Für das Merkmal „Einsatz von Lob (Lob)“ ergibt sich ein Eta-Koeffizient von  $\eta^2 = .051$ , d. h. dass lediglich 5.1 % der Varianz durch das Fach erklärt werden können. In Bezug auf die restlichen Merkmale des Unterrichtsklimas kann festgestellt werden, dass sich die absolute Höhe der Ausprägungen zwischen den beiden Videostudien nicht signifikant unterscheidet.

Neben der deskriptiven Analyse der hoch inferenten Daten aus den Videostudien Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) wird im Folgenden die Konstruktvalidität der Merkmale des hoch inferenten Ratingsystems überprüft, d. h. es wird ermittelt, ob diejenigen Merkmale, die dieselbe Basisdimension (Klassenführung oder Unterrichtsklima) messen sollen, tatsächlich auf einen gemeinsamen Faktor zurückführbar sind. Als ein wichtiges Verfahren zur Analyse der Konstruktvalidität gilt die Faktorenanalyse. Die Methode sowie die Ergebnisse dieser faktoriellen Überprüfung werden anschließend im Rahmen der Studie 2 dargestellt.

Tabelle 23: Unterschiede in den Ausprägungen der Klassenführungsmerkmale in der VSS (1. Schuljahr) und VSM (2. Schuljahr)

Merkmale der Klassenführung	N*	Video-studie	MW (SD)	$F_{(1,53)}$	p	$\eta^2$																																																								
Allgegenwärtigkeit der Lehrperson (Allg)	30	VSS	2.92 (.66)	.246	.662	.005																																																								
	30	VSM	3.11 (.57)				Gruppenfokus (Grup)	28	VSS	3.40 (.62)	.067	.796	.001	28	VSM	3.39 (.37)	Zeitmanagement / effektive Zeitnutzung (Zeit)	30	VSS	3.31 (.52)	1.496	.227	.028	26	VSM	3.49 (.60)	Übergangsmangement (Managing Transition) (Ueb)	30	VSS	3.05 (.59)	.199	.657	.004	27	VSM	3.00 (.57)	Regelklarheit bzw. -verwendung (Regl)	29	VSS	3.18 (.73)	.329	.569	.006	26	VSM	3.27 (.50)	Störungsfreiheit (Disz)	30	VSS	2.99 (.62)	1.353	.250	.025	28	VSM	3.25 (.59)	effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)	30	VSS	3.01 (.60)	1.358	.249
Gruppenfokus (Grup)	28	VSS	3.40 (.62)	.067	.796	.001																																																								
	28	VSM	3.39 (.37)				Zeitmanagement / effektive Zeitnutzung (Zeit)	30	VSS	3.31 (.52)	1.496	.227	.028	26	VSM	3.49 (.60)	Übergangsmangement (Managing Transition) (Ueb)	30	VSS	3.05 (.59)	.199	.657	.004	27	VSM	3.00 (.57)	Regelklarheit bzw. -verwendung (Regl)	29	VSS	3.18 (.73)	.329	.569	.006	26	VSM	3.27 (.50)	Störungsfreiheit (Disz)	30	VSS	2.99 (.62)	1.353	.250	.025	28	VSM	3.25 (.59)	effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)	30	VSS	3.01 (.60)	1.358	.249	.025	26	VSM	3.22 (.54)						
Zeitmanagement / effektive Zeitnutzung (Zeit)	30	VSS	3.31 (.52)	1.496	.227	.028																																																								
	26	VSM	3.49 (.60)				Übergangsmangement (Managing Transition) (Ueb)	30	VSS	3.05 (.59)	.199	.657	.004	27	VSM	3.00 (.57)	Regelklarheit bzw. -verwendung (Regl)	29	VSS	3.18 (.73)	.329	.569	.006	26	VSM	3.27 (.50)	Störungsfreiheit (Disz)	30	VSS	2.99 (.62)	1.353	.250	.025	28	VSM	3.25 (.59)	effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)	30	VSS	3.01 (.60)	1.358	.249	.025	26	VSM	3.22 (.54)																
Übergangsmangement (Managing Transition) (Ueb)	30	VSS	3.05 (.59)	.199	.657	.004																																																								
	27	VSM	3.00 (.57)				Regelklarheit bzw. -verwendung (Regl)	29	VSS	3.18 (.73)	.329	.569	.006	26	VSM	3.27 (.50)	Störungsfreiheit (Disz)	30	VSS	2.99 (.62)	1.353	.250	.025	28	VSM	3.25 (.59)	effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)	30	VSS	3.01 (.60)	1.358	.249	.025	26	VSM	3.22 (.54)																										
Regelklarheit bzw. -verwendung (Regl)	29	VSS	3.18 (.73)	.329	.569	.006																																																								
	26	VSM	3.27 (.50)				Störungsfreiheit (Disz)	30	VSS	2.99 (.62)	1.353	.250	.025	28	VSM	3.25 (.59)	effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)	30	VSS	3.01 (.60)	1.358	.249	.025	26	VSM	3.22 (.54)																																				
Störungsfreiheit (Disz)	30	VSS	2.99 (.62)	1.353	.250	.025																																																								
	28	VSM	3.25 (.59)				effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)	30	VSS	3.01 (.60)	1.358	.249	.025	26	VSM	3.22 (.54)																																														
effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)	30	VSS	3.01 (.60)	1.358	.249	.025																																																								
	26	VSM	3.22 (.54)																																																											

*Anmerkung:* Die Analysen beruhen auf einem N von 30 Lehrpersonen, die an beiden Videostudien teilgenommen haben. VSS = Videostudie Deutsch; VSM = Videostudie Mathematik, MW (SD) = Mittelwert (Standardabweichung), F = F-Wert (Freiheitsgrade), p = Irrtumswahrscheinlichkeit,  $\eta^2$  = partielles Eta-Quadrat

\* Die Höhe der Stichprobe variiert je nach Merkmal, da teilweise für vier Lehrpersonen für die entsprechenden Merkmale keine Werte vorliegen. In diesen Fällen handelt es sich um Nebenlehrpersonen. Wie diese „Nichtkategorisierung“ zustande kommt, ist in Kapitel 12.1.4 ausführlich beschrieben.

Tabelle 24: Unterschiede in den Ausprägungen der Unterrichtsklimamerkmale in der VSS (1. Schuljahr) und VSM (2. Schuljahr)

Merkmale des Unterrichtsklimas	N	Video- studie	MW (SD)	$F_{(1,53)}$	$p$	$\eta^2$
humorvolle Lernatmosphäre (Humo)	30	VSS	1.70 (.57)	.348	.558	.007
	30	VSM	1.78 (.68)			
Anerkennung der Schüler durch die Lehrperson (Anes)	30	VSS	3.48 (.67)	.010	.922	.000
	30	VSM	3.56 (.59)			
gegenseitige Anerkennung der Schüler (Aneq)	30	VSS	3.30 (.38)	.100	.753	.002
	30	VSM	3.38 (.57)			
Fürsorglichkeit, Herzlichkeit und Wärme (Fuer)	30	VSS	2.72 (.54)	.543	.464	.010
	30	VSM	2.60 (.52)			
Einsatz von Lob (Lob)	30	VSS	2.10 (.76)	2.793	.101	.051
	30	VSM	2.42 (.75)			
Ermutigung durch die Lehrperson (Emut)	30	VSS	2.70 (.54)	.994	.324	.019
	30	VSM	2.83 (.36)			
sachlich-konstruktive Rückmeldung (Rück)	30	VSS	2.54 (.46)	2.013	.162	.037
	30	VSM	2.76 (.45)			

*Anmerkung:* Die Analysen beruhen auf einem  $N$  von 30 Lehrpersonen, die an beiden Videostudien teilgenommen haben. VSS = Videostudie Deutsch, VSM = Videostudie Mathematik, MW (SD) = Mittelwert (Standardabweichung),  $F$  = F-Wert (Freiheitsgrade),  $p$  = Irrtumswahrscheinlichkeit,  $\eta^2$  = partielles Eta-Quadrat; grau hinterlegt: signifikante Unterschiede zwischen Deutsch (VSS, 1. Schuljahr) und Mathematik (VSM, 2. Schuljahr)

## 15 Ergebnisse der Studie 2 – Überprüfung der Konstruktvalidität

Die Zuordnung der Merkmale aus dem hoch inferenten Rating zu den theoretisch angenommenen Basisdimensionen der Unterrichtsqualität – Klassenführung und Unterrichtsklima – wird im Anschluss faktoranalytisch geprüft. Die Überprüfung der theoretischen Annahmen und der Datenstruktur erfolgt in zwei Schritten. Zunächst wird mithilfe sogenannter konfirmatorischer Faktoranalysen (CFA, *hypothesenprüfendes Verfahren*) überprüft, inwieweit die theoretische Zuordnung der einzelnen Merkmale des hoch inferenten Rating-systems zu den zwei Merkmalen Klassenführung und Unterrichtsklima bestätigt werden kann oder ob eventuell Alternativmodelle eher zu den empirischen Daten passen. In diesem Schritt geht es hauptsächlich darum, verschiedene Modelle anhand ausgewählter Gütekriterien zu vergleichen (vgl. Kapitel 15.1). Die Überprüfung der faktoriellen Struktur der Daten erfolgt in diesem ersten

Schritt unter Berücksichtigung aller Ratings aus der Deutsch- und Mathematikvideostudie ( $N_{VSS+VSM} = 101$ ). Die Analyse wird mithilfe des Softwareprogramms *MPlus* durchgeführt (vgl. Muthén & Muthén, 1998–2007). Mithilfe dieses Vorgehens soll die Forschungsfrage Nr. 6 beantwortet werden: Lässt sich die theoretisch angenommene 2-Faktoren-Struktur bestätigen?

Um Aussagen über die Generalisierbarkeit der Struktur über die Fächer Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) hinweg machen zu können, sollte sich eine fachspezifische CFA anschließen. Da die hier zugrunde liegende Stichprobe jedoch zu klein ist, um einen Mehrgruppenvergleich zwischen den beiden Fächern zu berechnen, erfolgt in einem zweiten Schritt eine explorative Faktorenanalyse (EFA, *hypothesengenerierendes Verfahren*), getrennt nach den zwei Fächern Deutsch (1. Schuljahr,  $N_{VSS} = 50$ ) und Mathematik (2. Schuljahr,  $N_{VSM} = 51$ ) (vgl. Kapitel 15.2). Die Analyse erfolgt mithilfe des Softwareprogramms SPSS.

Zur Überprüfung der faktoriellen Struktur der hoch inferenten Merkmale wurden sowohl für die CFA als auch für die fachbezogenen EFA nicht alle Merkmale der Klassenführung und des Unterrichtsklimas einbezogen. Neben dem Merkmal „positiver Umgang der Lehrperson mit Fehlern (Fehl)“ werden die folgenden drei Merkmale nicht berücksichtigt:

- Vorbereitung des Klassenraums (Vorb)
- Positives Fehlerklima zwischen den Schülern (Fekl)
- Wettbewerb und Leistungsdruck (Wett)

Diese Entscheidung wurde nicht willkürlich getroffen. Bei der Berechnung der fachspezifischen Beobachterübereinstimmung wiesen diese drei Merkmale in der Videostudie Mathematik einen zu geringen G-Koeffizienten auf und konnten somit nicht reliabel eingeschätzt werden (vgl. Kapitel 13.2). Durch den Ausschluss dieser Merkmale, auch in der Videostudie Deutsch, steht für die faktoranalytische Prüfung die gleiche Anzahl an Merkmalen pro Fach zur Verfügung. Die Anzahl der Merkmale verringert sich somit auf 14.

In Anlehnung an den Aufbau des hoch inferenten Ratingsystems (vgl. Kapitel 12.1.2) ist in Abbildung 23 die theoretisch zu erwartende Faktorenstruktur schematisch dargestellt. Die rechteckigen Kästchen stellen die 14 verbleibenden hoch inferenten Merkmale dar. Diese sind manifest. Die ovalen Formen bilden die zugrunde liegenden latenten Faktoren der Klassenführung und des Unterrichtsklimas ab. Die Pfeile zwischen den Faktoren Klassenführung bzw. Unterrichtsklima und den entsprechenden Merkmalen stellen die im Modell zugelassenen Korrelationen dar, wobei gerichtete Beziehungen mit gerichteten Pfeilen und ungerichtete Beziehungen mit Doppelpfeilen dargestellt werden. In der Abbildung 23 laden alle theoretisch angenommenen Merkmale der Klassenführung auf dem ersten Faktor und alle theoretisch hergeleiteten Merkmale des Unterrichtsklimas auf einem zweiten Faktor. Die nicht durch diese 2-faktorielle Struktur erklärte Varianz der einzel-

nen Merkmale wird durch Fehlerterme (err) dargestellt. Fehlerterme beziehen sich demnach auf den Varianzanteil der Merkmale, der nicht durch die latenten Konstrukte erklärt werden kann.

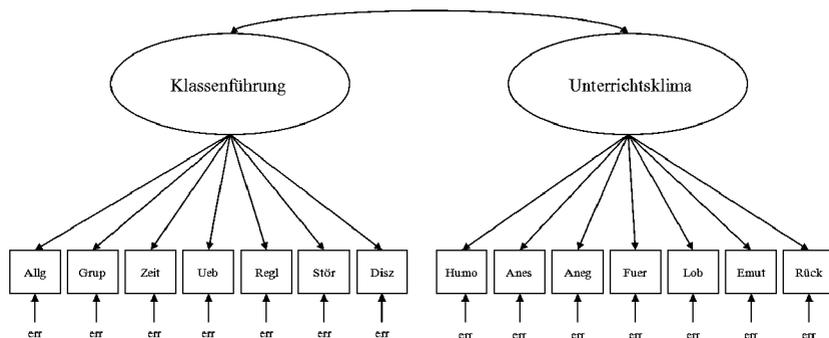


Abbildung 23: Schematische Darstellung der erwarteten Faktorenstruktur (2-Faktoren-Modell)

Um einen ersten Überblick über die zu erwartenden Zusammenhänge zwischen den Merkmalen der Klassenführung und des Unterrichtsklimas zu bekommen, sind in der folgenden Tabelle 25 die Korrelationen der 14 verbliebenen Merkmale der Klassenführung und des Unterrichtsklimas auf Basis der Daten aus beiden Videostudien ( $N_{VSS+VSM} = 101$ ) dokumentiert. Das Muster der Interkorrelationen der einzelnen Merkmale der Klassenführung und des Unterrichtsklimas weist auf eine inhaltliche Ausdifferenzierung der Konstrukte hin. Die Merkmale der Klassenführung korrelieren untereinander teilweise sehr hoch ( $.352 \leq r \leq .944$ ,  $p < .001$ ). Die Zusammenhänge zwischen den Merkmalen des Unterrichtsklimas variieren dagegen sehr stark. Zusätzlich scheint das Merkmal „gegenseitige Anerkennung der Schüler (Aneg)“ kein zentrales Merkmal des Unterrichtsklimas zu sein, da sich mehr signifikante Zusammenhänge mit Merkmalen der Klassenführung bestätigen lassen. Des Weiteren deutet sich an, dass die Merkmale „Ermutigung durch die Lehrperson (Emut)“ und „sachlich-konstruktive Rückmeldung (Rück)“ nur sehr gering bis gar nicht mit weiteren Merkmalen des Unterrichtsklimas zusammenhängen. Dies könnte bereits ein Hinweis für einen zusätzlichen dritten Faktor sein.

Auf Grundlage der dargestellten Korrelationsmatrix (vgl. Tabelle 25) wird demnach davon ausgegangen, dass die oben theoretisch angenommene 2-Faktoren-Struktur (vgl. Abbildung 23) weniger gut zu den zugrunde liegenden Daten passt. Inwieweit das Merkmal „gegenseitige Anerkennung der Schüler (Aneg)“ zum Faktor Klassenführung gezählt werden kann und welches Modell die besten Fit-Indizes aufweist, wird anhand der folgenden Ausführungen überprüft.

Tabelle 25: Korrelationsmatrix für die 14 verbliebenen Merkmale

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Allg (1)	–													
Grup (2)	.438***	–												
Zeit (3)	.408***	.352***	–											
Ueb (4)	.558***	.490***	.683***	–										
Regl (5)	.566***	.486***	.603***	.734***	–									
Disz (6)	.591***	.420***	.586***	.659***	.771***	–								
Stor (7)	.607***	.377***	.533***	.638***	.765***	.929***	–							
	<b>Klassenführung</b>													
Humo (8)	-.004 ns	.000 ns	-.055 ns	.105 ns	.124 ns	.039 ns	.115 ns	–						
Anes (9)	.279**	.108 ns	.270*	.288**	.300**	.250*	.236*	.250*	–					
Aneg (10)	.381**	.182+	.251*	.410***	.545***	.543***	.611***	.061 ns	.126 ns	–				
Fuer (11)	.070 ns	-.013 ns	.165 ns	.187+	.135 ns	.076 ns	.090 ns	.382***	.639***	.044 ns	–			
Lob (12)	.354***	.230*	.422***	.467***	.406***	.418***	.463***	.124 ns	.227*	.314***	.314***	–		
Emut (13)	.249+	.091 ns	.348***	.308**	.326***	.405***	.333***	-.183+	.052 ns	.119 ns	.037 ns	.339***	–	
Ruck (14)	.321***	.201*	.406***	.387***	.477***	.454***	.383***	-.067 ns	.453***	.173+	.298**	.273*	.611***	–
	<b>Unterrichtsklima</b>													

Anmerkung: N = 101 (VSS + VSM), \*\*\*p < .001, \*\*p < .01, \*p < .05, +p < .10; grau hinterlegt: nicht signifikante (ns) Korrelationen

## 15.1 Konfirmatorische Faktorenanalysen (CFA)

Mithilfe der konfirmatorischen Faktorenanalyse (Confirmatory Factor Analysis, CFA) können unterschiedliche konkurrierende Modelle auf Grundlage desselben Datensatzes gegeneinander getestet werden (vgl. Jöreskog, 1993). Im vorliegenden Fall bedeutet dies, dass ein Generalfaktor-Modell gegen das theoretisch angenommene 2-Faktoren-Modell getestet werden kann. Wie sich bereits im Rahmen der bivariaten Korrelationen andeutet (vgl. Kapitel 15, Tabelle 25), ist auch ein dritter Feedback-Faktor zu erwarten. Aus diesem Grund wird zusätzlich ein 3-Faktoren-Modell expliziert. Der Modellvergleich erfolgt in Kapitel 15.1.3. Im Folgenden wird die gewählte Schätzmethode dargestellt.

### 15.1.1 Wahl der Schätzmethode

Neben der theoretischen Modellspezifikation muss für die CFA im Vorfeld die Methode der Parameterschätzung festgelegt werden. Im Rahmen des Modellvergleichs wurde als Schätzmethode Maximum Likelihood (ML) gewählt. Diese ist als Methode einerseits ausgesprochen robust gegenüber Verletzungen der Normalverteilungsannahme, andererseits ist die ML-Methode auch bei geringeren Stichprobengrößen ( $N < 100$ ) anwendbar (vgl. Bühner, 2006). Nach Empfehlungen von Bühner (2006) sollte bei einer geringen Stichprobengröße und einer Schiefe von 2 sowie einem Exzess von 7 die ML-Methode angewendet werden. Welche Merkmale normalverteilt sind, wurde im Vorfeld anhand des Kolmogorov-Smirnov-Tests (bei einer Stichprobe) überprüft, ein Test auf multivariate Normalverteilung konnte nicht durchgeführt werden. Der Test wurde auf Basis aller Deutsch- und Mathematikvideos durchgeführt. Als normalverteilt können folgende Merkmale erachtet werden: „Allgegenwärtigkeit der Lehrperson (Allg)“, „Störungsfreiheit (Disz)“, „effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)“ und „gegenseitige Anerkennung der Schüler (Aneg)“ (vgl. Kapitel 21, Tabelle 43).

### 15.1.2 Theoretische Herleitung der Modelle

Faktorenanalysen erlauben es zu überprüfen, wie gut die zuvor theoretisch angenommene Struktur (vgl. Abbildung 23, vgl. Kapitel 15.1) die empirisch bestätigten Zusammenhänge repräsentieren kann. Um das Modell zu identifizieren, das im Vergleich zu weiteren postulierten Modellen die beste Modellgüte aufweist, wird anschließend zunächst ein Generalfaktor-Modell gegen das erwartete 2-Faktoren-Modell getestet. Zur Beantwortung der Frage, ob das 2-Faktoren-Modell die Merkmalszusammenhänge der zugrunde liegenden Daten besser wiedergeben kann als ein Generalfaktor-Modell (For-

schungsfrage Nr. 6), können zur Beurteilung der Anpassungsgüte (Fit) verschiedene Anpassungstests herangezogen werden (zusf. Bühner, 2006; vgl. auch Moosbrugger & Schermelleh-Engel, 2006), d. h. es kann überprüft werden, „ob eine hinreichende Übereinstimmung (Modellfit) zwischen den empirischen Daten und dem theoretischen Modell besteht oder ob das Modell verworfen werden muss“ (Moosbrugger & Schermelleh-Engel, 2006, S. 316).

Zur Überprüfung der Güte eines Modells wird üblicherweise der Chi-Quadrat-Differenzentest herangezogen. Dieser errechnet sich aus der Abweichung der im Modell implizierten Kovarianzmatrix von der empirischen Kovarianzmatrix und misst somit, wie stark die Variablenzusammenhänge des spezifizierten Modells vom zugrunde liegenden Datensatz abweichen. Ist dieser Chi-Wert nicht signifikant, wird das Modell als passend angesehen (vgl. Schermelleh-Engel, Moosbrugger & Müller, 2003). Der aus dem Modelltest resultierende  $\chi^2$ -Wert sollte möglichst klein sein, für einen guten Modellfit kleiner als zweimal die Anzahl der Freiheitsgrade ( $\chi^2 \leq 2df$ ) (vgl. Moosbrugger & Schermelleh-Engel, 2006). Auch bei einem Verhältnis des  $\chi^2$ -Werts zu den Freiheitsgraden ( $df$ ) von  $(\chi^2 / df) \leq 2.5$  kann von einem gut angepassten Modell ausgegangen werden (vgl. Homburg & Baumgartner, 1995). Der Test ist jedoch abhängig von der Stichprobengröße (vgl. Bühner, 2006). Bei einer kleinen Stichprobengröße kann der  $\chi^2$ -Test eventuell zu einer Annahme eines spezifizierten Modells führen, obwohl große Abweichungen zum perfekten Modell existieren. Aus diesem Grund wurden im vorliegenden Fall zu zusätzliche Indizes, wie das *BIC* (Bayesian Information Criterion) und das *AIC* (Akaike Information Criterion, vgl. Akaike, 1987) zur Beurteilung der Modellgüte herangezogen. Ein nicht signifikanter  $\chi^2$ -Anpassungstest sowie niedrige Werte für das *BIC* und das *AIC* sprechen dafür, dass ein Modell am besten auf die empirischen Daten passt. Zusätzlich werden der *TLI* (Tucker-Lewis Index, vgl. Tucker & Lewis, 1973), der *CFI* (Comparative Fit Index, vgl. Bentler, 1990) und der *RMSEA* (Root-Mean-Square Error of Approximation, vgl. Browne & Cudeck 1993) betrachtet (vgl. Tabelle 26). Um eine akzeptable Modellpassung annehmen zu können, sollten der *CFI* und der *TLI*  $> .90$  sein. Bühner (2006) nennt für den *RMSEA* als Richtwert für eine kleine Stichprobe  $N < 250$  den Cut-off-Wert  $\leq .08$ . Vorteile des *CFI* sind vor allem die geringe Sensitivität gegenüber einer geringen Stichprobengröße und die Robustheit gegenüber der Verletzung der Verteilungsannahme. Tabelle 26 enthält zusammenfassend sowohl Wertebereiche als auch Cut-off-Werte der zuvor dargelegten Fit-Indizes. Allerdings ist zu beachten, dass über diese Cut-off-Werte eine rege Diskussion entbrannt ist und diese stark umstritten sind. Da solche Cut-off-Kriterien nicht generalisierbar (vgl. Bühner, 2006), sondern vielmehr situationsabhängig sind, sollten sie ausschließlich als Daumenregeln verwendet werden, wie folgendes Zitat nahelegt: „Decision rules such as *RMSEA*  $< .05$  [...] may be useful in some solutions but they often lead to inappropriate decisions in other solutions [...] and should be considered only as rules of thumb“ (Hu & Bentler, 1999, S. 4).

Tabelle 26: Wertebereiche und Cut-off-Werte der verwendeten Fit-Indizes nach Bühner (2006)

Index	Wertebereiche	Cut-off-Werte	
		gut	akzeptabel
$\chi^2 / df$	Maximalwert > 1	$\leq 2.5$	$\leq 2.0$
CFI	Maximalwert = 1	$\approx .95$	.90
TLI	Maximalwert > 1	$\approx .95$	.90
AIC / BIC	je niedriger, desto besser das Modell		
RMSEA	Maximalwert = 1	$\leq .05$	$\leq .08$

*Modell (1) – Generalfaktor Lernumwelt*

In einem ersten Modell wird lediglich ein Generalfaktor für die Lernumwelt angenommen. Das bedeutet, dass eine Differenzierung zwischen den Merkmalen der Klassenführung und des Unterrichtsklimas nicht spezifiziert wird. Dieses Modell nimmt demnach an, dass die Ausprägungen der jeweiligen Merkmale der Klassenführung und des Unterrichtsklimas in der Grundschule über einen gemeinsamen latenten Faktor „Lernumwelt“ erklärt werden können. Somit werden sowohl die Merkmale der Klassenführung als auch die Merkmale des Unterrichtsklimas über einen gemeinsamen Faktor Lernumwelt abgebildet (vgl. Abbildung 24).

Da jedoch auf Basis der Korrelationsmatrix (vgl. Kapitel 15, Tabelle 25) anzunehmen ist, dass die 14 Merkmale nicht nur einen Generalfaktor, sondern bis zu drei differenzierbare Konstrukte bilden können (Klassenführung, Unterrichtsklima und Feedback), sollte dieses Modell die Daten weniger gut erklären als ein Modell, in dem zwei bzw. drei latente Faktoren angenommen werden.

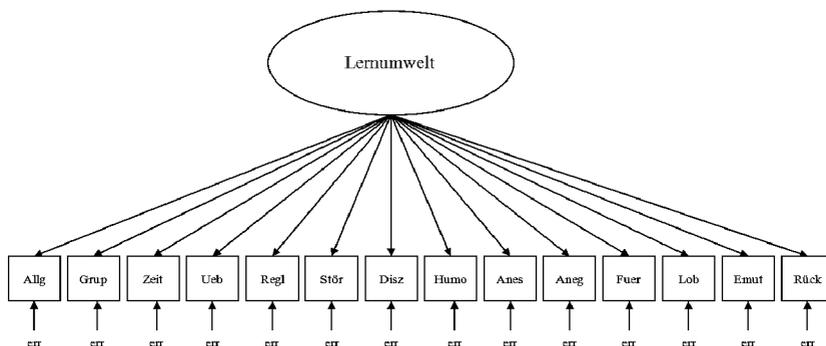


Abbildung 24: Das Generalfaktor-Modell (Modell 1)

Anmerkung: In dem Modell 1 wird ein Generalfaktor Lernumwelt angenommen.

*Modell (2) – 2-Faktoren-Modell*

In einem zweiten Modell (vgl. Abbildung 25) wird anstelle des Generalfaktor-Modells in Anlehnung an die in Kapitel 3.2 und 4.2 theoretisch dargestellten Merkmale angenommen, dass die 14 verbliebenen Merkmale der Klassenführung und des Unterrichtsklimas zwei korrelierte latente Faktoren (Faktor 1: Klassenführung und Faktor 2: Unterrichtsklima) abbilden.

Mithilfe der Korrelationsmatrix (vgl. Kapitel 15, Tabelle 25) konnte jedoch im Vorfeld bereits gezeigt werden, dass das Merkmal „gegenseitige Anerkennung der Schüler (Aneg)“ nicht sehr hoch mit den Merkmalen des Unterrichtsklimas korreliert, obwohl die gegenseitige Wertschätzung als Aspekt der sozialen Eingebundenheit (vgl. Kapitel 4.2) bzw. des ‚Emotional Supports‘ (vgl. CLASS, Pianta et al., 2008) angesehen werden kann. Vielmehr zeigten sich in der vorliegenden Studie hohe Zusammenhänge mit den Merkmalen der Klassenführung wie der ‚Regelklarheit bzw. -verwendung (Regl)‘ ( $r = .55, p < .001$ ), der ‚Störungsfreiheit (Disz)‘ ( $r = .54, p < .001$ ) oder dem ‚effektiven Umgang mit auftretenden Unterrichtsstörungen (Stör)‘ ( $r = .61, p < .001$ ). Demnach scheint das Merkmal „gegenseitige Anerkennung der Schüler (Aneg)“ in der Grundschule eher ein Aspekt der Klassenführung als ein Aspekt des Unterrichtsklimas zu sein. Eine mögliche Erklärung für dieses Ergebnis liefert u. a. Neuenschwander (2005). Nach Neuenschwander (2005) werden vor allem soziale Interaktionen und soziale Beziehungen maßgeblich durch Regeln und Normen reguliert. Demnach kann angenommen werden, dass die gegenseitige Anerkennung der Schüler bzw. der Respekt unter den Schülern Einfluss auf die Umgangsformen und die soziale Kommunikation nehmen können. Wenn es einer Lehrperson gelingt, frühzeitig Regeln für das soziale Miteinander im Unterricht zu setzen, kann sie Einfluss auf das soziale Verhalten der Schüler nehmen, womit auch eine inhaltliche Begründung für die gefundenen Zusammenhänge zwischen dem Merkmal „gegenseitige Anerkennung der Schüler (Aneg)“ und den Klassenführungsmerkmalen gegeben ist. Auch Wannack et al. (2009) haben für die Schweizer Schuleingangsstufe ein pädagogisches Konzept entwickelt, in dem sie bereits im Vorfeld Aspekte der gegenseitigen Anerkennung zum ‚Classroom Management‘ zählen. In Anlehnung an Evertson et al. (2006) ließen sie im Rahmen einer videobasierenden Unterrichtsbeobachtung im Kindergarten und in der Unterstufe Regeln für allgemeine Verhaltensweisen einschätzen, die einerseits die Integrität der Schüler sichern und andererseits darauf eingehen, wie die Schüler miteinander kommunizieren (vgl. Wannack et al., 2011). Auch im Rahmen der PERLE-Videostudien wurde das Merkmal „gegenseitige Anerkennung der Schüler (Aneg)“ zum Teil als Regel für das soziale Miteinander operationalisiert. Beispielsweise beinhaltet die Dimension Indikatoren wie „Die Schüler lachen einander nicht aus“ oder „Die Schüler lassen die anderen ausreden“ (vgl. Kapitel 19.2). In Klassen mit einer hohen Einschätzung auf diesem Merkmal kann somit implizit davon ausgegangen werden, dass Regeln bezüglich des sozialen Miteinanders eingeführt wurden.

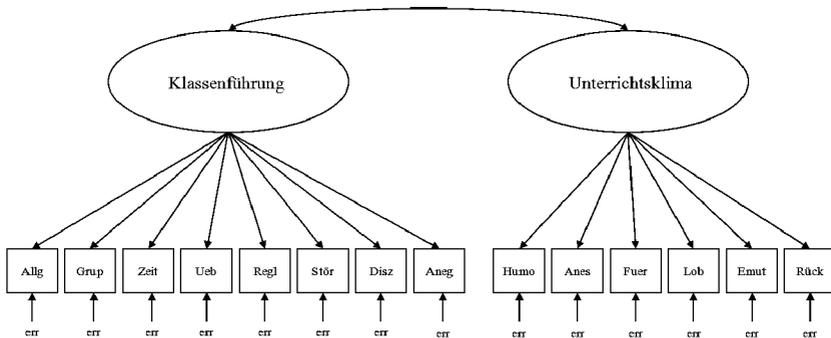


Abbildung 25: Das 2-Faktoren-Modell (Modell 2a)

*Anmerkung:* Das Modell 2a spezifiziert zwei separate latente Faktoren (Klassenführung und Unterrichtsklima), wobei das Merkmal „gegenseitige Anerkennung der Schüler (Aneg)“ als Aspekt der Klassenführung betrachtet wird. Es wird angenommen, dass die zwei latenten Faktoren untereinander korrelieren (siehe Doppelpfeil).

Als Folge dieses Vorgehens wird angenommen, dass das Modell – wie es ursprünglich angenommen wurde (vgl. Abbildung 23) – schlechtere Fit-Indizes aufweist, als ein Modell, in dem das Merkmal „gegenseitige Anerkennung der Schüler (Aneg)“ auf dem Faktor Klassenführung lädt (= Modell 2a, vgl. Abbildung 25). Dies soll anschließend überprüft werden. Das Modell 2a sollte auch im Vergleich zum Modell 1 eine bessere Anpassung an die Daten aufweisen.

### Modell (3) – 3-Faktoren-Modell

Da die Ergebnisse der Korrelationsmatrix (vgl. Kapitel 15, Tabelle 25) zusätzlich auf einen dritten Faktor hinweisen bzw. vor allem die Merkmale „Ermutigung durch die Lehrperson (Emut)“, „sachlich-konstruktive Rückmeldung (Rück)“ aber auch „Einsatz von Lob (Lob)“<sup>88</sup> untereinander stärker korrelieren als mit den restlichen Merkmalen des Unterrichtsklimas, wird alternativ ein 3-Faktoren-Modell erstellt (vgl. Abbildung 26).

Neben den Faktoren Klassenführung und Unterrichtsklima wird das Modell 3 somit um den latenten Faktor Feedback erweitert und sollte im Vergleich zum Modell 2a eine ähnlich gute, eventuell eine bessere Anpassung an die Daten aufweisen. Außerdem entspricht dieses 3-Faktoren-Modell annähernd der Konzeptualisierung der Unterrichtsqualität von Pianta et al. (2008) (vgl. auch Hamre et al., 2007), die mithilfe des CLASS-Beobachtungssystems zeigen konnten, dass ein 3-Faktoren-Modell mit den Faktoren ‚Emotional Support‘, ‚Classroom Organization‘ and ‚Instructional Support‘ besser auf

88 Die Dimension „Einsatz von Lob (Lob)“ korreliert außerdem hoch mit einzelnen Aspekten der Klassenführung (vgl. Tabelle 25). Vermutlich ergibt sich hieraus eine eher problematische Zuordnung zu einem der Faktoren. Dies gilt es zu überprüfen.

die Daten von 3.000 Schülern passt als ein 2- oder 1-Faktoren-Modell (3-Faktoren-Modell wies bessere Fit-Indizes auf, vgl. Kapitel 2.1).

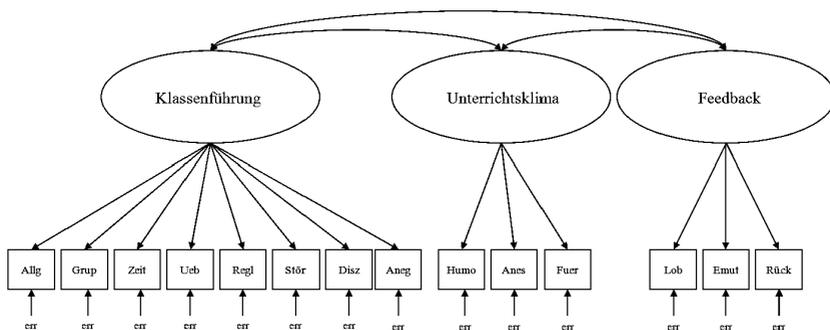


Abbildung 26: Das 3-Faktoren-Modell (Modell 3)

*Anmerkung:* Das Modell 3 nimmt neben den latenten Faktoren Klassenführung und Unterrichtsklima einen dritten Faktor Feedback an. Es wird angenommen, dass die drei latenten Faktoren untereinander korrelieren (siehe Doppelpfeile).

### 15.1.3 Ergebnis des Modellvergleichs

Die in Abbildung 24 bis Abbildung 26 dargestellten Modelle lassen sich unter Berücksichtigung der verwendeten Modellgütekriterien (vgl. Kapitel 15.1.2) miteinander vergleichen. Ziel dieses Vergleichs ist es herauszufinden, welches der Modelle für die zugrunde liegenden hoch inferenten Daten besser passt. Das Ergebnis des Modellvergleichs ist im Anschluss zusammenfassend dargestellt (vgl. Tabelle 27). Alle Modelle wurden auf Basis von  $N = 101$  Unterrichtsvideos (sowohl Mathematik- als auch Deutschvideos) berechnet (vgl. Kapitel 22). Neben den Cut-off-Kriterien wird für alle Modelle mit mehr als einem latenten Faktor zusätzlich die Korrelation der latenten Faktoren (= latentes  $r$ ) angegeben. Um von differenzierbaren latenten Faktoren ausgehen zu können, sollte die latente Korrelation nicht größer als .90 sein.

Modell 1 (= Generalfaktor-Modell, vgl. Abbildung 24), das für alle Merkmale der Klassenführung und des Unterrichtsklimas einen Generalfaktor Lernumwelt annimmt, erweist sich im Vergleich zu den restlichen spezifizierten Modellen als das schlechteste ( $CFI$  und  $TLI$  sind am weitesten von 1 und  $RMSEA$  ist am weitesten von 0 entfernt, vgl. Tabelle 27). Nach den bestehenden Gütekriterien muss das Generalfaktor-Modell demnach verworfen werden. Das Modell 2 (= 2-Faktoren-Modell, vgl. Abbildung 23), in dem das Merkmal „gegenseitige Anerkennung der Schüler (Aneg)“ auf dem Faktor „Unterrichtsklima“ lädt, weist – wie erwartet – eine schlechtere Modellanpassung auf und muss demnach verworfen werden. Das Modell 2a (vgl. Ab-

bildung 25), in dem die Ausprägungen der hoch inferenten Merkmale jeweils durch einen latenten Faktor für die Klassenführung und einen für das Unterrichtsklima erklärt werden, konnte ohne zusätzlich zugelassene Residualkorrelationen<sup>89</sup> zwischen den drei Merkmalen „Einsatz von Lob (Lob)“, „Ermutigung durch die Lehrperson (Emut)“ und „sachlich-konstruktive Rückmeldung (Rück)“ nicht berechnet werden und muss an dieser Stelle verworfen werden. Aus diesem Grund wurden in einem weiteren Modell 2b diese Korrelationen festgesetzt (vgl. Abbildung 27). Die Zulassung der Korrelationen der Fehlerterme dieser drei Merkmale bedeutet, dass in dem Modell berücksichtigt wird, dass sie etwas Gemeinsames messen, das in dem Modell nicht vorhanden ist.

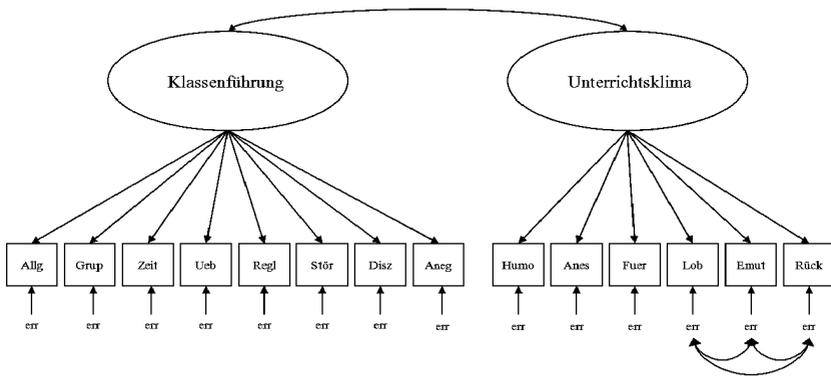


Abbildung 27: Modell 2b mit zugelassenen Residualkorrelationen zwischen den Merkmalen des Feedbacks

*Anmerkung:* Im Vergleich zu Modell 2a wird im Modell 2b angenommen, dass die Residuen der Merkmale „Einsatz von Lob (Lob)“, „Ermutigung durch die Lehrperson (Emut)“ und „sachlich-konstruktive Rückmeldung (Rück)“ miteinander korrelieren.

Durch die Zulassung der Residualkorrelationen im Modell 2b konnte im Vergleich zum Generalfaktor-Modell (Modell 1, vgl. Abbildung 24) eine bessere, wenn auch nicht akzeptable Modellgüte erzielt werden (vgl. Tabelle 27). Unabhängig von den zusätzlichen Residualkorrelationen kann an dieser Stelle die Hypothese 2 bestätigt werden. Trotz einer nicht akzeptablen Modellgüte,

<sup>89</sup> Als Residualkorrelationen werden Korrelationen zwischen den Fehlertermen manifester Variablen bezeichnet. Eine Nichtberücksichtigung dieser Residualkorrelationen zwischen manifesten Variablen würde zu einer Fehlschätzung der Pfadkoeffizienten führen (vgl. Marsh & Hau, 1996). Grundsätzlich sind Fehlerterme voneinander und von den Faktoren unabhängig. Die Zulassung der Korrelation einzelner Fehlerterme bedeutet, dass sie etwas Gemeinsames messen, was im Modell nicht vorhanden ist. Fehlerterme beziehen sich demnach auf den Varianzanteil der Dimensionen, der nicht durch die latenten Konstrukte erklärt werden kann. Im vorliegenden Fall deuten die drei zusätzlichen Residualkorrelationen auf einen dritten latenten Faktor (Feedback) hin.

passt ein zweidimensionales Modell (mit zusätzlichen Residualkorrelationen) besser auf die zugrunde liegenden hoch inferenten Daten als das eindimensionale Modell. Die Korrelation zwischen den latenten Faktoren Klassenführung und Unterrichtsklima betrug in diesem Modell  $r = .325$  (vgl. Tabelle 27) und bestätigt, dass es sich um zwei getrennte Konstrukte handelt.

Tabelle 27: Der Modellvergleich (Modell 1, 2, 2b und 3) und die Gütekriterien der konfirmatorischen Faktorenanalysen ( $N = 101$ ).

	Modell 1 (Generalfaktor- Modell)	Modell 2 (urspr. 2-Fakto- ren-Modell)	Modell 2b* (2-Faktoren- Modell+Reskorr)	Modell 3 (3-Faktoren- Modell)
$\chi^2$	268.862	255.903	169.895	176.424
<i>df</i>	77	76	74	73
$\chi^2/df$	3.492	3.367	2.327	2.384
<i>CFI</i>	.751	.767	.874	.867
<i>TLI</i>	.706	.721	.843	.837
<i>AIC</i>	1799.159	1788.200	1708.191	1712.721
<i>BIC</i>	1908.994	1900.650	1828.487	1830.401
<i>RMSEA</i>	.157	.153	.115	.117
<i>latentes r</i>	–	.702	.325	.206/ .391/ .585

*Anmerkung:* Reskorr = Residualkorrelationen. Im Modell 1 wurde ein globaler latenter Faktor Lernumwelt angenommen, während Modell 2 einen latenten Faktor Klassenführung und einen latenten Faktor Unterrichtsklima beinhaltet. Das Modell 2 entspricht dem ursprünglich erwarteten 2-Faktoren-Modell, in dem das Merkmal „gegenseitige Anerkennung der Schüler (Aneg)“ auf dem Unterrichtsklima lädt. Das Modell 2a konnte nicht berechnet werden. Aus diesem Grund ist es auch nicht in der Tabelle aufgeführt. Im Modell 2b wurde das Merkmal „gegenseitige Anerkennung der Schüler (Aneg)“ als Merkmal der Klassenführung spezifiziert. Zusätzlich enthält das Modell 2b Residualkorrelationen für die Merkmale „Einsatz von Lob (Lob)“, „Ermutigung durch die Lehrperson (Emut)“ und „sachlich-konstruktive Rückmeldung (Rück)“. Im Modell 3 wurde ein 3-Faktor-Modell spezifiziert, das neben den Faktoren Klassenführung und Unterrichtsklima zusätzlich einen Faktor Feedbackverhalten der Lehrperson annimmt.

\* Dunkelgrau hervorgehoben: Modell mit der besten Modellgüte

Da die Modellierung von Residualkorrelationen generell bedeutet, dass die manifesten Merkmale etwas Gemeinsames messen, was im Modell nicht vorhanden ist, wird in einem weiteren Modell zusätzlich getestet, ob anstatt der korrelierten Fehler ein weiterer latenter Faktor – im vorliegenden Fall ein Feedback-Faktor – in das Modell mit aufgenommen werden muss und somit die Ausprägungen der hoch inferenten Merkmale durch ein 3-Faktoren-Modell erklärt werden können (vgl. Abbildung 26). Im Vergleich zum Modell 2b

erreicht das Modell 3 jedoch keine bessere Modellgüte. Die Korrelationen der latenten Faktoren betragen in dem Modell 3  $r = .206$  zwischen der Klassenführung und dem Unterrichtsklima,  $r = .585$  zwischen der Klassenführung und dem Feedback und  $r = .391$  zwischen dem Unterrichtsklima und dem Feedback.

In einem nächsten Schritt wird auf Basis explorativer Faktorenanalysen überprüft, ob sich Unterschiede in der Struktur in Abhängigkeit vom Unterrichtsfach bestätigen lassen. Die Ergebnisse werden anschließend vorgestellt.

## 15.2 Fachspezifische explorative Faktorenanalyse (EFA)

Um zu überprüfen, ob für beide Fächer die gleiche Faktorenstruktur angenommen werden kann bzw. ob sich Unterschiede in der faktoriellen Struktur in Abhängigkeit vom Unterrichtsfach zeigen (Forschungsfrage Nr. 7), wurde zusätzlich eine fachbezogene explorative Faktorenanalyse durchgeführt. Für die Videostudie Deutsch (1. Schuljahr) bilden 50 und für die Videostudie Mathematik (2. Schuljahr) 51 Einschätzungen die Grundlage. Dabei gehen – wie bei der fachunspezifischen konfirmatorischen Faktorenanalyse – erneut die Mittelwerte der Ratings pro Rater und pro Merkmal als Indikatoren für die Konstrukte in zwei separate Faktorenanalysen ein. Als *Extraktionsmethode* der Faktoren wurde die Hauptkomponentenanalyse (Principal Components Analysis, PCA) und als *Rotationsmethode* die Oblimin-Rotation als eine oblique Rotationstechnik gewählt. Diese führen zu korrelierten Faktoren (vgl. Bühner, 2006). Die Zuordnung der einzelnen Merkmale zu latenten Faktoren erfolgt datenorientiert, d. h. aufgrund eines vorher bestimmten Abbruchkriteriums. Als *Abbruchkriterium* wurde das Kaiser-Kriterium (auch Kaiser-Guttman-Kriterium genannt; Guttman, 1954) gewählt, d. h. in allen folgenden Analysen werden Faktoren mit einem Eigenwert größer eins als bedeutsam erachtet, „da sie mehr Varianz erklären als eine einzelne standardisierte Variable aufweist“ (Moosbrugger & Schermelleh-Engel, 2006, S. 312). Die Analyse erfolgt mit dem Statistikprogramm SPSS.

In Anlehnung an die fachunabhängigen konfirmatorischen Faktorenanalysen (vgl. Kapitel 15.1) werden für die fachspezifischen explorative Faktorenanalysen die gleichen 14 Merkmale der Klassenführung und des Unterrichtsklimas verwendet. Für beide Videostudien werden anschließend tabellarisch sowohl die Faktorenladungen der Merkmale auf den rotierten Faktoren als auch die Eigenwerte und die internen Konsistenzen (Cronbachs Alphas) der extrahierten Faktoren dargestellt. Der Eigenwert eines Faktors entspricht der Summe der quadrierten Ladungen aller Merkmale über einen Faktor (vgl. Bühner, 2006). Der relative Anteil, den der extrahierte Faktor an der Gesamtvarianz aller Merkmale aufklärt, gibt dessen Bedeutung wieder. Cron-

bachs Alpha gilt als die am häufigsten verwendete Methode zur Bestimmung der internen Konsistenz. Die interne Konsistenz ist umso höher, je höher die durchschnittlichen Korrelationen aller Merkmale sind (vgl. Schermelleh-Engler & Werner, 2006). Im vorliegenden Fall geht es darum, die durchschnittliche Korrelation der einzelnen Merkmale eines Faktors zu berechnen und Aussagen über die Reliabilität einzelner Faktoren zu machen. Voraussetzung ist, dass alle Merkmale eines Faktors das gleiche Merkmal, z. B. die Klassenführung, messen. Das Cronbachs Alpha nimmt in den meisten Fällen einen Wert zwischen null und eins an. Nach Nunnally (1978) wird ein Wert von  $\leq .70$  für eine akzeptable Reliabilität empfohlen. Im Rahmen der PERLE-Studie gilt ein Cronbachs Alpha von  $< .60$  als nicht mehr zufriedenstellend, sodass Skalen mit diesem Wert für weitere Analysen nicht verwendet werden sollten (vgl. Greb, Lipowsky & Faust, 2009b).

Bevor anschließend die Ergebnisse der fachspezifischen Faktorenanalysen dargestellt werden, erfolgt zunächst erneut ein Überblick über die Zusammenhänge der einzelnen Merkmale der Klassenführung und des Unterrichtsklimas in Abhängigkeit vom Fach (vgl. Tabelle 28 und 29). In beiden Videostudien ergeben sich erneut mittlere bis hohe Korrelationen für die Merkmale der Klassenführung. Das Merkmal „gegenseitige Anerkennung der Schüler (Aneg)“ weist auch bei der fachspezifischen Betrachtung größere Zusammenhänge mit den Merkmalen der Klassenführung auf. Innerhalb der Merkmale der Klassenführung deuten sich jedoch Unterschiede in der Höhe der Zusammenhänge in Abhängigkeit vom jeweiligen Fach an. Inwieweit sich Unterschiede in der faktoriellen Struktur in Abhängigkeit vom Fach zeigen, wird anschließend im Rahmen der fachspezifischen EFA dargestellt.

Tabelle 30 zeigt, dass sich bei der explorativen Faktorenanalyse im Fach Deutsch (1. Schuljahr) die drei Faktoren Klassenführung ( $\alpha = .91$ ), Unterrichtsklima ( $\alpha = .76$ ) und Feedback ( $\alpha = .77$ ) sinnvoll voneinander abgrenzen lassen. Die in Tabelle 30 dargestellte Faktorenlösung erfolgte auf Basis der Mustermatrix der rotierten Faktorenladungen. Für die Merkmale „gegenseitige Anerkennung der Schüler (Aneg)“, „Zeitmanagement / effektive Zeitnutzung (Zeit)“ sowie „humorvolle Lernatmosphäre (Humo)“ zeigen sich nennenswerte Nebenladungen größer  $\pm .30$  auf dem dritten Feedback-Faktor. Die einzelnen Merkmale mit Nebenladungen größer  $\pm .30$  wurden nur in die Reliabilitätsanalyse der Faktoren einbezogen, auf denen sie die höchste Ladung zeigten. Die Varianzaufklärung der 3-Faktorenlösung ist mit 68.15 % zufriedenstellend. Aus Gründen der Lesbarkeit werden sowohl in Tabelle 30 als auch in Tabelle 31 (Ergebnisse der EFA für die Videostudie Mathematik) die Itemtrennschärfen nicht berichtet. Die Darstellung der Itemtrennschärfen erfolgt im Rahmen der Skalendokumentation im Anhang dieser Arbeit (vgl. Kapitel 23).





Tabelle 30: Faktorladungen der hoch inferenten Merkmale sowie interne Konsistenzen der extrahierten Faktoren in der Videostudie Deutsch ( $N = 50$ )

Merkmale	Klassen- führung	Unterrichts- klima	Feedback
effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)	.92		
Störungsfreiheit (Disz)	.88		
gegenseitige Anerkennung der Schüler (Aneg)	.86		(.38)
Regelklarheit bzw. -verwendung (Regl)	.83		
Übergangsmanagement (Managing Transition) (Ueb)	.70		
Allgegenwärtigkeit der Lehrperson (Allg)	.65		
Zeitmanagement / effektive Zeitznutzung (Zeit)	.56		(-.41)
Gruppenfokus (Grup)	.47		
<hr/>			
Fürsorglichkeit, Herzlichkeit und Wärme (Fuer)		.88	
Anerkennung der Schüler durch die Lehrperson (Anes)		.84	
humorvolle Lernatmosphäre (Humo)		.70	(.31)
Einsatz von Lob (Lob)*	(.40)	(.48)	
<hr/>			
sachlich-konstruktive Rückmeldung (Rück)			-.82
Ermutigung durch die Lehrperson (Emut)			-.76
Cronbachs Alpha	.91	.76	.77

*Anmerkung:* Es werden nur Faktorladungen größer  $\pm .30$  dargestellt. ( ) kennzeichnen die Nebenladungen der Merkmale. Die Merkmale wurden nur dann in die Reliabilitätsberechnung der einzelnen Faktoren einbezogen, wenn sie auf dem entsprechenden Faktor die höchste Ladung zeigten.

\* Das Merkmal „Einsatz von Lob (Lob)“ wurde aus der Skala Unterrichtsklima herausgenommen. Das Cronbachs Alpha steigt von .73 auf .76.

Analog zu dem beschriebenen Vorgehen für die Videostudie Deutsch (1. Schuljahr) wurde auch im Rahmen der Videostudie Mathematik (2. Schuljahr) die faktorielle Struktur der hoch inferenten Merkmale ermittelt. Tabelle 31 stellt die Faktorladungen und internen Konsistenzen der extrahierten Faktoren dar.

Tabelle 31: Faktorladungen der hoch inferenten Merkmale sowie interne Konsistenzen der extrahierten Faktoren in der Videostudie Mathematik ( $N = 51$ )

Merkmale	effektiver Umgang mit Störungen	effizientes Zeitmanagement	Unterrichtsklima	Feedback
effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)	.84			
Störungsfreiheit (Disz)	.80			
gegenseitige Anerkennung der Schüler (Aneg)	.73			
Regelklarheit bzw. -verwendung (Regl)	.55	(-.44)		
Zeitmanagement / effektive Zeitznutzung (Zeit)		-.80		
Übergangsmanagement (Managing Transition) (Ueb)		-.73		
Gruppenfokus (Grup)		-.74		
Allgegenwärtigkeit der Lehrperson (Allg)		-.70		
Fürsorglichkeit, Herzlichkeit und Wärme (Fuer)			.89	
Anerkennung der Schüler durch die Lehrperson (Anes)			.81	(-.30)
humorvolle Lernatmosphäre (Humo)	(.46)		.50	(.52)
Einsatz von Lob (Lob)*		(.33)		(-.40)
Ermutigung durch die Lehrperson (Emut)				-.91
sachlich-konstruktive Rückmeldung (Rück)			(.42)	-.73
Cronbachs Alpha	.89	.82	.62	.72

*Anmerkung:* Es werden nur Faktorladungen größer  $\pm .30$  dargestellt. ( ) kennzeichnen die Nebenladungen der Merkmale. Die Merkmale wurden nur dann in die Reliabilitätsberechnung der einzelnen Faktoren einbezogen, wenn sie auf dem entsprechenden Faktor die höchste Ladung zeigten.

\* Das Merkmal „Einsatz von Lob (Lob)“ wurde für die Vergleichbarkeit mit den Skalen der Videostudie Deutsch aus der Skala Feedback herausgenommen. Das Cronbachs Alpha steigt von .52 auf .72.

Im Vergleich zur Videostudie Deutsch (1. Schuljahr) lassen sich auf Grundlage der Daten der Videostudie Mathematik im zweiten Schuljahr ( $N = 51$ ) vier Faktoren sinnvoll voneinander abgrenzen. Während sich im ersten Schuljahr im Fach Deutsch nur ein Gesamtfaktor Klassenführung extrahieren ließ, un-

terteilt sich der Faktor Klassenführung im Fach Mathematik (2. Schuljahr) in zwei separate Faktoren: (1) „effizientes Zeitmanagement“ und (2) „effektiver Umgang mit Störungen“. Das Merkmal „Regelklarheit bzw. -verwendung (Regl)“ als Aspekt des Faktors „effektiver Umgang mit Störungen“ hat jedoch mit  $-44$  eine hohe Nebenladung auf dem Faktor „effizientes Zeitmanagement“. Theoretisch ergibt die Zuordnung zum Faktor „effektiver Umgang mit Störungen“ Sinn, denn die Einführung eines Regelsystems und die konsequente Einhaltung der Regeln schaffen günstige Voraussetzungen für einen effektiven Umgang mit Unterrichtsstörungen.

Neben den zwei Klassenführungsfaktoren bestätigen sich in Mathematik die Faktoren Unterrichtsklima und Feedback, wobei das Merkmal „humorvolle Lernatmosphäre (Humo)“ eine höhere Ladung auf dem Feedback hat als auf dem Unterrichtsklima. Aus inhaltlichen Gründen und im Sinne der Vergleichbarkeit mit dem Unterrichtsklima-Faktor im Fach Deutsch (1. Schuljahr) wird das Merkmal jedoch zum Unterrichtsklima gezählt. Für alle vier extrahierten Faktoren ergeben sich gute Cronbachs Alphas ( $\alpha > .62$ , vgl. Tabelle 31) und eine insgesamt zufriedenstellende Varianzaufklärung von 72.03 %.

*Zusammenfassend* zeigen die fachspezifischen explorativen Faktorenanalysen, dass sich die Merkmale des hoch inferenten Ratings in beiden Fächern zu unterschiedlich vielen Faktoren zusammenfassen lassen, was eine mögliche Erklärung für die vergleichsweise schlechten Fit-Indizes der konfirmatorisch geprüften Modelle sein könnte. Es ergeben sich vor allem Unterschiede bei der Zusammenfassung der Merkmale der Klassenführung. Während im Fach Deutsch (1. Schuljahr) alle Merkmale der Klassenführung und das Merkmal „gegenseitige Anerkennung der Schüler (Aneg)“ einen Gesamtfaktor abbilden, unterteilt sich dieser im Fach Mathematik (2. Schuljahr) in zwei separate Faktoren („effizientes Zeitmanagement“ und „effektiver Umgang mit Störungen“). Dabei zeigen sich auf den Merkmalen der Klassenführung Unterschiede in der Höhe der einzelnen Faktorenladungen. Die Faktoren Unterrichtsklima und Feedback scheinen in beiden Fächern vergleichbar zu sein, trotz Unterschieden in den Reliabilitäten. Für die weiteren Analysen in Studie 3 wurden entsprechend den Ergebnissen der fachspezifischen EFA Skalen gebildet. Die Merkmale „sachlich-konstruktive Rückmeldung (Rück)“, „Ermutigung durch die Lehrperson (Emut)“ sowie „Einsatz von Lob (Lob)“ wurden für die Analysen in der Studie 3 nicht herangezogen, stehen aber als einzelne Merkmale für weitere Analysen zur Verfügung.

### 15.3 Skalenbildung

Die fachspezifischen Faktor- und Reliabilitätsanalysen lassen, je nach Fach, Unterschiede in der Faktorenstruktur erkennen. Für die Daten aus der Vi-

deostudie Deutsch (1. Schuljahr) ließen sich die drei Faktoren Klassenführung, Unterrichtsklima und Feedback sinnvoll voneinander abgrenzen. In Mathematik (2. Schuljahr) zeigte sich hingegen eine Lösung mit vier Faktoren, in der die Klassenführung nochmals unterteilt werden konnte (vgl. Kapitel 15.2). Für weitere fachspezifische Analysen in den jeweiligen Fächern wurde entsprechend den Ergebnissen der fachspezifischen EFA für beide Fächer eine Skala „Unterrichtsklima“ gebildet, die folgende Merkmale enthält: „humorvolle Lernatmosphäre (Humo)“, „Fürsorglichkeit, Herzlichkeit und Wärme (Fuer)“ und „Anerkennung der Schüler durch die Lehrperson (Anes)“. Da sich für das Fach Mathematik in der EFA eine klare Unterteilung des Faktors Klassenführung gezeigt hat, wurden zwei separate Skalen („effizientes Zeitmanagement“ und „effektiver Umgang mit Störungen“) gebildet. Im Fach Deutsch ergab sich eine Gesamtskala „Klassenführung“.

Tabelle 32: Skalenkennwerte für die Videostudie Deutsch (1. Schuljahr) und die Videostudie Mathematik (2. Schuljahr)

Videostudie	Skala	$\alpha$	$N$	Min	Max	$MW$	$SD$
VSS (1. Schuljahr)	Klassenführung	.91	50	1.94	3.91	3.16	.47
	Unterrichtsklima	.76	50	1.17	3.33	2.64	.44
VSM (2. Schuljahr)*	effizientes Zeitmanagement	.82	48	2.06	4.00	3.34	.44
	effektiver Umgang mit Störungen	.89	47	1.81	4.00	3.29	.50
	Unterrichtsklima	.62	51	1.00	3.50	2.61	.43

Anmerkung: VSS = Videostudie Deutsch (1. Schuljahr), VSM = Videostudie Mathematik (2. Schuljahr),  $\alpha$  = Cronbachs Alpha,  $N$  = Anzahl Fälle, Min = Minimum, Max = Maximum,  $MW$  = Mittelwert,  $SD$  = Standardabweichung

\* Für die Videostudie Mathematik konnten aufgrund des 75 %-Kriteriums für drei bzw. vier Lehrpersonen keine Skala gebildet werden. Aus diesem Grund weicht die Größe der Stichprobe ab.

Das Merkmal „Einsatz von Lob (Lob)“ konnte im Fach Mathematik keinem der Faktoren zugeordnet werden. So zeigte sich beispielsweise bei der Berechnung der Reliabilität der Skala „Feedback“ (auf diesem Faktor hat das Merkmal die höchste Ladung), dass das Cronbachs Alpha ohne das Merkmal „Einsatz von Lob (Lob)“ von .52 auf .72 steigt (siehe Anhang, Kapitel 23.2). Aus diesem Grund wurde das Merkmal bei der Skalenbildung nicht berücksichtigt. Im Fach Deutsch (1. Schuljahr) lädt das Merkmal am höchsten auf dem Faktor „Unterrichtsklima“. Da sich jedoch die Höhe der Reliabilität der Skala ohne das Merkmal „Einsatz von Lob (Lob)“ kaum verändert (siehe Anhang, Kapitel 23.1) und für die Studie 3 vergleichbare Skalen wünschenswert sind, wurde das Merkmal auch im Fach Deutsch bei der Skalenbildung vorerst nicht weiter berücksichtigt.

Für die Bildung der Skalen wurde ein 75 %-Kriterium herangezogen. Dies besagt, dass nur für diejenigen Lehrpersonen ein Skalenmittelwert errechnet wird, für die mindestens 75 % der Merkmale einer Skala von den Ratern eingeschätzt werden konnten. Im vorliegenden Fall konnten für alle Lehrpersonen, die an der Videostudie Deutsch im ersten Schuljahr teilgenommen haben ( $N = 50$ ), Skalenmittelwerte gebildet werden. Für die Videostudie Mathematik (2. Schuljahr) erwiesen sich je nach Klassenführungsskala drei bis vier Nebenlehrpersonen bei der Bildung der Skalen als problematisch. Hier konnte aufgrund zahlreicher nicht einschätzbarer Merkmale keine Skala Klassenführung gebildet werden. Tabelle 32 gibt die Skalenkennwerte für die jeweiligen Skalen pro Fach wieder. Die Berechnung der Mehrebenenregressionsmodelle im Rahmen der Studie 3 (vgl. Kapitel 16) erfolgt auf Basis dieser gebildeten Skalen.

## 16 Ergebnisse der Studie 3 – Multikriteriale Effekte der Klassenführung und des Unterrichtsklimas

Mithilfe von mehreren Mehrebenenregressionsmodellen wird anschließend untersucht, welchen Einfluss die Basisdimensionen effektive Klassenführung und positives Unterrichtsklima im Anfangsunterricht auf multikriteriale Schülermerkmale haben. In Kapitel 16.1 werden die Effekte der Unterrichtsqualitätsmerkmale im Fach Deutsch (1. Schuljahr) fokussiert, während in Kapitel 16.2 die Effekte im Fach Mathematik (2. Schuljahr) untersucht werden. Im Vorfeld der Berechnung der Mehrebenenregressionsmodelle erfolgt in einem ersten Schritt die Überprüfung der Multikollinearität der Prädiktoren. Zuvor wird jedoch beschrieben, wie im Rahmen der Studie 3 mit den Nebenlehrpersonen in den entsprechenden BIP-Klassen verfahren wurde.

### *Umgang mit Nebenlehrpersonen in Studie 3*

Für einen Teil der BIP-Klassen liegen zwei Urteile für die Effektivität der Klassenführung und das Unterrichtsklima vor, die daraus resultieren, dass in diesen BIP-Klassen der Unterricht im Sinne des „Teamteachings“ von zwei Lehrpersonen gestaltet wurde. Im Vorfeld des Ratings wurden die zwei Rater dahingehend instruiert, dass sie in einem ersten Durchgang die Hauptlehrperson und in einem zweiten Durchgang die Nebenlehrperson fokussieren und entsprechend Urteile auf den einzelnen Merkmalen der Klassenführung und des Unterrichtsklimas abgeben (vgl. Kapitel 12.1.4.1). Ziel war es unter anderem zu überprüfen, ob sich die zwei Lehrpersonen in Bezug auf die Ausprägung der einzelnen Merkmale der Klassenführung und des Unterrichtsklimas unterscheiden. Für die vorliegende Teilstudie werden nur Daten derjenigen

Lehrpersonen herangezogen, die im Vorfeld als Hauptlehrpersonen identifiziert wurden. Es wird davon ausgegangen, dass die Unterrichtsplanung vor allem durch diese Lehrpersonen erfolgt ist. Demnach ist die Hauptlehrperson beispielsweise im Wesentlichen für die Zeitplanung, aber auch für die Unterrichtsorganisation allgemein zuständig. Dadurch reduziert sich die Stichprobe in Deutsch von 50 auf 37 Lehrpersonen bzw. Klassen, in Mathematik von 51 auf 36.

Anschließend wurde für die 37 Lehrpersonen in Deutsch und die 36 Lehrpersonen in Mathematik überprüft, inwieweit die im Vorfeld gebildeten Skalen der Klassenführung und des Unterrichtsklimas (vgl. Kapitel 15.3) miteinander korrelieren bzw. inwieweit Multikollinearität der Prädiktoren gegeben ist.

#### *Schritt 1: Überprüfung der Multikollinearität der Level-2-Prädiktoren*

Zur Berechnung der Mehrebenenregressionsmodelle – deren Ergebnisse anschließend dargestellt werden (vgl. Kapitel 16.1 und 16.2) – wurde das Programm *MPlus* (Version 5; Muthén & Muthén, 1998–2007) verwendet. *MPlus* greift auf die multiple Regressionsanalyse<sup>90</sup> zurück und erlaubt die gleichzeitige Betrachtung mehrerer Prädiktorvariablen (vgl. Geiser, 2010; Urban & Mayerl, 2011). Dies ist vor allem für die vorliegende Arbeit von Bedeutung, da im Vorfeld ein Zusammenhang zwischen den beiden Prädiktorvariablen (Klassenführung und Unterrichtsklima) angenommen wird (vgl. Kapitel 1.2.2), der an dieser Stelle berücksichtigt werden kann.

Wie bereits in Kapitel 7.4 angesprochen, sollte in Anlehnung an Urban und Mayerl (2011), die Korrelation zwischen den beiden unabhängigen Variablen auf Level 2 nicht größer  $\pm .80$  sein. Kuß (2012) nennt als eine Konsequenz von Multikollinearität der Prädiktorvariablen die Erzeugung von „geteilter“ Varianz, was zur Folge hätte, dass die spezifische Varianz der einzelnen Prädiktoren auf ein Niveau reduziert wird, das die Ermittlung der Effekte dieser Variablen problematisch macht. Ein Weg zur Identifizierung von Multikollinearität ist die Berechnung der bivariaten Korrelation der Unterrichtsvariablen. In den folgenden Tabellen werden die Ergebnisse der Prüfung der Multikollinearität für beide Videostudien dargestellt. Zur Überprüfung des Zusammenhangs wurden die zuvor gebildeten Skalen herangezogen (vgl. Kapitel 15.3). Die Überprüfung erfolgte in SPSS. Die Datengrundlage bilden für die Videostudie Deutsch 37 und für die Videostudie Mathematik 36 Klassen. Mithilfe der bivariaten Korrelationen können in einem ersten Schritt die Forschungsfragen Nr. 8.1 und 8.2 beantwortet werden (vgl. Kapitel 7.4): Zeigt sich der theoretisch angenommene Zusammenhang zwischen der Klassen-

---

90 „Rein verfahrensmäßig betrachtet, werden in einer multiplen Regression die empirischen Werte einer jeden unabhängigen Variable um diejenigen Anteile bereinigt, die durch lineare Effekte der anderen unabhängigen Variablen bestimmt werden, und wird sodann eine Regression der abhängigen Variable auf die nunmehr bereinigten (kontrollierten) unabhängigen Variablen durchgeführt“ (Urban & Mayerl, 2011, S. 86).

führung und dem Unterrichtsklima auch empirisch? Unterscheidet sich die Höhe des Zusammenhangs in Abhängigkeit vom Fach?

Im Vorfeld wurde angenommen, dass eine effektive Klassenführung sowohl im Fach Deutsch (1. Schuljahr) als auch im Fach Mathematik (2. Schuljahr) eng mit einem positiven Unterrichtsklima zusammenhängt. Jedoch kann die Hypothese 3 nur teilweise bestätigt werden, da sich lediglich für das Fach Deutsch (1. Schuljahr) ein signifikant positiver Zusammenhang des beobachteten Unterrichtsklimas mit der hoch inferent eingeschätzten Klassenführung nachweisen lässt (vgl. Tabelle 33;  $r = .296$ ). Allerdings wird die Korrelation, vermutlich aufgrund der kleinen Stichprobe, nur auf einem 10 %-Niveau signifikant. Auf Ebene der einzelnen Merkmale zeigen sich teilweise etwas höhere Zusammenhänge (vgl. Kapitel 15.2, Tabelle 28 und Tabelle 29).

Tabelle 33: Zusammenhang zwischen den zwei Unterrichtsmerkmalen in der Videostudie Deutsch (VSS, März 2007)

	VSS Klassenführung (März 2007)
VSS Unterrichtsklima (März 2007)	.296 <sup>+</sup>

Anmerkung:  $N = 36$ , <sup>+</sup> $p < .10$ , bivariate Korrelation

Im Gegensatz zum Fach Deutsch zeigt sich für das Fach Mathematik (2. Schuljahr) kein signifikanter Zusammenhang des Unterrichtsklimas mit den beiden Teilskalen der Klassenführung (effektives Zeitmanagement; effektiver Umgang mit Störungen). Klassen mit einer hohen beobachteten Ausprägung für die Effektivität der Klassenführung im Fach Mathematik (2. Schuljahr) erhalten von den Ratern nicht gleichzeitig höhere Werte für das Unterrichtsklima. Lehrpersonen, die in ihrer Klasse einen effektiven Umgang mit Störungen praktizieren, wurden im Mittel von den Ratern jedoch gleichzeitig besser hinsichtlich der Effizienz ihres Zeitmanagements eingeschätzt (vgl. Tabelle 34;  $r = .623$ ,  $p < .001$ ).

Tabelle 34: Zusammenhang zwischen den drei Unterrichtsmerkmalen in der Videostudie Mathematik (VSM, März 2008)

	VSM eff. Zeitmanagement (März 2008)	VSM eff. Umgang mit Störungen (März 2008)
VSM Unterrichtsklima (März 2008)	.074 <i>n.s.</i>	.163 <i>n.s.</i>
VSM eff. Zeitmanagement (März 2008)	–	.623***

Anmerkung:  $N = 37$ , \*\*\* $p < .001$ , *n.s.* = nicht signifikant, bivariate Korrelationen

Die Ergebnisse der bivariaten Korrelationen rechtfertigen für das Fach Deutsch (1. Schuljahr) die gleichzeitige Aufnahme beider Unterrichtsquali-

tätsmerkmale in die nachfolgenden Mehrebenenregressionsmodelle. Statistische Folgen der Multikollinearität werden durch die geringe Höhe der Korrelation von .296 ( $p < .10$ ) vermieden (vgl. Kapitel 7.4). Für das Fach Mathematik (2. Schuljahr) ergab sich hingegen eine hohe Korrelation der beiden Klassenführungsmerkmale. Aufgrund dieser hohen Korrelation müssen Interaktionseffekte der beiden Variablen berücksichtigt werden, d. h. die erklärte Varianz der abhängigen Variable (je nach Modell entweder Leistung oder Selbstkonzept) setzt sich aus dem jeweiligen Erklärungspotenzial der einzelnen Prädiktoren und einem gemeinsamen Erklärungspotenzial der beiden Prädiktoren zusammen. Damit sind nach Fromm (2008) „die Regressionskoeffizienten nicht als Maß für die Bedeutsamkeit eines Prädiktors interpretierbar: Der Regressionskoeffizient eines Prädiktors wird rechnerisch nicht mehr ausschließlich vom Zusammenhang des Prädiktors mit dem Kriterium bestimmt, sondern zusätzlich durch die Interkorrelation der Prädiktoren“ (S. 351). Obwohl die Korrelation der beiden Teilskalen der Klassenführung in Mathematik (2. Schuljahr) unter dem Richtwert von  $\pm .80$  von Urban und Mayerl (2011) liegt, gehen die beiden Teilaspekte separat in die Mehrebenenregressionsmodelle ein.

#### *Schritt 2: Mehrebenenregressionsmodelle für Deutsch und Mathematik*

In einem nächsten Schritt erfolgt die Spezifikation von sechs Mehrebenenregressionsmodellen in *MPlus* über die Option *type is twolevel* und unter Verwendung der Klassen-ID als Clustervariable (vgl. Kapitel 11.4). Tabelle 35 gibt einen Überblick über die in der vorliegenden Arbeit spezifizierten Mehrebenenregressionsmodelle, getrennt nach den beiden Videostudien Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) und unter Berücksichtigung der Ergebnisse der Multikollinearitätsprüfung.

Bevor die Mehrebenenregressionsmodelle in *MPlus* spezifiziert wurden, erfolgte in einem weiteren Schritt für jede der vier abhängigen Variablen die Berechnung eines sogenannten ‚Twolevel-basic-Modells‘ (vgl. Geiser, 2010). Ziel dieser Berechnung war es, sich einen ersten Überblick über die Clusterstruktur und den Grad der Abhängigkeit der Beobachtungen zu verschaffen (vgl. Geiser, 2010). Als Maß der Abhängigkeit wurde in diesem ersten Schritt in *MPlus* automatisch die Intraklassenkorrelation ( $ICC, p$ ) berechnet. Diese gibt den Anteil der Ebene-2-Varianz an der Gesamtvarianz in der abhängigen Variablen an (Gleichung 2;  $\sigma_{u0}$  = Level-2-Varianz;  $\sigma_r$  = Level-1-Varianz):

$$p = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_r^2} \quad (2)$$

Tabelle 35: Überblick über die Mehrebenenregressionsmodelle der Studie 3

	Deutsch 1. SJ		Mathematik 2. SJ			
	Modell 1a	Modell 1b	Modell 2a	Modell 2b	Modell 2c	Modell 2d
AV	Rechtschr. Leistungen / DERET (Ende 1. SJ)	Selbstkonzept Schreiben (Ende 1. SJ)	Leistungen Mathematik (Ende 2. SJ)	Leistungen Mathematik (Ende 2. SJ)	Selbstkonzept Rechnen (Ende 2. SJ)	Selbstkonzept Rechnen (Ende 2. SJ)
Prädiktor L1	Vorläuferfähigkeiten (Anfang 1. SJ)	Selbstkonzept Schreiben (Anfang 1. SJ)	Leistungen Mathematik (Ende 1. SJ)	Leistungen Mathematik (Ende 1. SJ)	Selbstkonzept Rechnen (Ende 1. SJ)	Selbstkonzept Rechnen (Ende 1. SJ)
Prädiktoren L2	(1) Klassenführung (2) Unterrichtsklima (VSS, März 2007)	(1) Klassenführung (2) Unterrichtsklima (VSS, März 2007)	(1) Eff. Zeitmanagement (2) Unterrichtsklima (VSM, März 2008)	(1) Eff. Umgang mit Störungen (2) Unterrichtsklima (VSM, März 2008)	(1) Eff. Zeitmanagement (2) Unterrichtsklima (VSM, März 2008)	(1) Eff. Umgang mit Störungen (2) Unterrichtsklima (VSM, März 2008)

Anmerkung: AV = abhängige Variable auf Level 1, L1 = Level 1; L2 = Level 2, VSS = Videostudie Deutsch, VSM = Videostudie Mathematik, SJ = Schuljahr

In Tabelle 36 sind die ICC der vier abhängigen Variablen der Studie 3 tabellarisch dargestellt. Da *MPlus* im Rahmen der Mehrebenenanalyse nur diejenigen Personen in die Analysen einbezieht, die für die verwendete Variable gültige Werte aufweisen (= listwise deletion, vgl. Kapitel 11.5), wird in der Spalte  $N_{\text{Schüler}}$  zusätzlich angegeben, auf welcher Datengrundlage die ICC berechnet wurde.<sup>91</sup> Die ICC kann Werte zwischen 0 und 1 annehmen. Je höher der Wert für die ICC, desto größer ist der Anteil an Varianz in der entsprechenden Variable (z. B. Rechtschreibleistung), der auf die Zugehörigkeit der Schüler zu einer Klasse zurückgeführt werden kann.

Tabelle 36: Intraklassenkorrelationen (ICC) für abhängige Variablen der Studie 3

Fach	Abhängige Variable (Schülerebene)	$N_{\text{Schüler}}$	ICC
Deutsch (Bereich Rechtschreiben)	Leistungen Rechtschreiben (Ende 1. SJ)	694	.109
	Selbstkonzept Schreiben (Ende 1. SJ)	704	.025
Mathematik	Leistungen Mathematik (Ende 2. SJ)	604	.069
	Selbstkonzept Rechnen (Ende 2. SJ)	680	.037

Anmerkung: Im Fach Deutsch werden 37 und im Fach Mathematik 36 Klassen berücksichtigt. ICC = Intraklassenkorrelation, SJ = Schuljahr

91 Die Anzahl der einbezogenen Schüler weicht aufgrund der hohen Anzahl fehlender Werte auf den abhängigen Variablen von der ursprünglichen Größe der Schülerstichprobe ab (vgl. Kapitel 11.2).

Die Höhe der ICC-Koeffizienten deuten auf eine mehr oder weniger große Abhängigkeit der Beobachtungen von der Zugehörigkeit zu einer Klasse (= Cluster) hin, je nachdem, welche abhängige Variable man betrachtet. Vor allem in Bezug auf die Rechtschreibleistungen lässt sich ein Varianzanteil von rund 11 % auf Unterschiede zwischen den Klassen zurückführen. Der Rest geht auf Unterschiede zwischen den Schülern selbst zurück. Der Anteil der Varianz, der auf Unterschiede zwischen den Klassen in den Selbstkonzepten zurückgeführt werden kann, ist hingegen eher gering. In Anlehnung an Geiser (2010), der bereits bei Intraklassenkorrelationen von .05 oder .01 Verzerrungen<sup>92</sup> der Ergebnisse von Signifikanztests postuliert (vgl. Kapitel 11.4), ist es trotz der eher geringen ICC in den beiden Bereichen sinnvoll, die Mehrebenenstruktur für die Analysen zu berücksichtigen.

Nachdem bestätigt werden konnte, dass ein Teil der Varianz der abhängigen Variablen sowohl in Deutsch (Ende 1. Schuljahr) als auch in Mathematik (Ende 2. Schuljahr) durch Unterschiede zwischen den Klassen erklärt werden kann, besteht der nächste Schritt darin, das „Basic-Modell“ zu erweitern, indem die Prädiktoren auf Level 1 und auf Level 2 mit in das Modell aufgenommen werden. Auf Level 2 gehen die entsprechenden Skalen der Klassenführung und des Unterrichtsklimas in das Modell ein. In *MPlus* werden die Klassenführung und das Unterrichtsklima gleichzeitig als „Between-Variable“ definiert und gehen als manifeste Variablen in die Modelle ein. Auf Individualebene (Level 1) werden, je nach Modell, die vorherigen Leistungen bzw. Selbstkonzepte kontrolliert. Die Kontrollvariablen auf Level 1 werden in dem Modell als „Within-Variable“ definiert und gehen in *MPlus* als *grand-mean*-zentrierte<sup>93</sup> Variablen in die Modelle ein.

Die Ergebnisse dieser fachspezifischen Mehrebenenregressionsanalysen werden anschließend dargestellt. Mithilfe dieser Analysen sollen die Forschungsfragen Nr. 9 bis 12 beantwortet werden. Es werden für alle sechs Mehrebenenregressionsmodelle keine Statistiken zur Beurteilung der Modellpassung angegeben (Model-Fit-Indizes), da nur einfache Twolevel-Modelle mit ausschließlich manifesten Variablen gerechnet wurden. Ein solches einfaches Modell schöpft die in den Daten vorhandenen Informationen voll aus, daher werden solche Modelle auch als *saturierte Modelle* bezeichnet. Saturierte Modelle weisen immer eine perfekte Anpassung der Daten auf (vgl. Geiser, 2010). Abgebildet sind in den Modellen jeweils die standardisierten Regressionskoeffizienten sowie die Korrelation zwischen den Prädiktorvariablen auf

---

92 Ohne Berücksichtigung der Mehrebenenstruktur der Daten würde es zu einer Unterschätzung des Standardfehlers und damit zu einer Unterschätzung des Alpha-Risikos führen (vgl. Raudenbush & Bryk, 2002).

93 Eine Grand-Mean-Zentrierung der Prädiktorvariable auf Level 1 bedeutet, dass die Prädiktorvariable am Gesamtmittelwert zentriert ist. Die „Regressionssteigungen beruhen in diesem Fall auf Zusammenhängen zwischen einem Kriterium und den Prädiktoren, die in Form von Abweichungen vom Gesamtmittelwert repräsentiert sind“ (vgl. Nezlek, Schröder-Abé & Schütz, 2006, S. 220).

Level 2. Der standardisierte Regressionskoeffizient gibt an, welchen Effekt die jeweilige Prädiktorvariable X (Level 1 oder Level 2) auf die abhängige Variable (Level 1) hat, unter der Voraussetzung, dass alle anderen unabhängigen Variablen bei der Einflussnahme der Variable X auf die abhängige Variable konstant bleiben, d. h. keinen Einfluss ausüben (vgl. Urban & Mayerl, 2011). Dieses Verfahren der „Bereinigung“ der Variable X um die Einflüsse der anderen unabhängigen Variablen wird auch als Auspartialisierung von X-Anteilen beschrieben. Aus diesem Grund werden diese Regressionskoeffizienten auch als „partielle Regressionskoeffizienten“ bezeichnet (vgl. Urban & Mayerl, 2011). Es wird kein Maß für den Anteil der aufgeklärten Varianz durch die Prädiktorvariablen bzw. für die Güte der Regressionsschätzung angegeben, da aufgrund der eher kleinen Stichprobe von 37 bzw. 36 Klassen, die Varianzanteile in mehrebenenanalytischen Auswertungsverfahren teilweise instabil ausfallen, was Simulationsstudien bestätigen (Maas & Hox, 2005; Van den Noortgate, Opdenakker & Onghena, 2005). Lüdtke et al. (2007) sowie Nezelek et al. (2006) verweisen auf Studien, die von mindestens 50 Level-2-Einheiten sprechen, wenn es um die Aufklärung von Varianzanteilen geht. Aus diesem Grund werden anschließend nur die standardisierten Regressionskoeffizienten dargestellt.

Da sich im Verlauf der Analysen für die Studie 3 kaum Abweichungen im Ergebnistrend zwischen einem schrittweisen Vorgehen<sup>94</sup> und der Berechnung eines Gesamtmodells bestätigen ließen, erfolgt die Ergebnisdarstellung für das jeweilige Mehrebenenregressionsmodell in Form eines Gesamtmodells.

## 16.1 Die Bedeutung der Unterrichtsmerkmale im Fach Deutsch (1. SJ)

Die Ergebnisse für das Fach Deutsch (1. Schuljahr) werden mithilfe der zwei spezifizierten Mehrebenenregressionsmodelle 1a und 1b dargestellt (vgl. Kapitel 16, Tabelle 35).

### *Abhängige Variable Leistungen Rechtschreiben (Ende 1. SJ) – Modell 1a*

Im Mehrebenenregressionsmodell 1a (vgl. Abbildung 28) wird als abhängige Variable die individuelle Rechtschreibleistung vom Ende des ersten Schuljahres herangezogen. Auf Level 1 werden die schriftspracherwerbsspezifischen Vorläuferfähigkeiten der Schüler (Anfang 1. Schuljahr, vgl. Kapitel 12.2.2) kontrolliert. Diese sind am Gesamtmittelwert zentriert. Als Prädiktorvariablen auf Level 2 werden die Skalen Klassenführung und Unterrichtsklima aus der Videostudie Deutsch (VSS, März 2007) herangezogen (vgl. Kapitel 15.3).

<sup>94</sup> In einem schrittweisen Vorgehen werden die Prädiktoren auf Level 1 und 2 zeitlich nacheinander in das Modell aufgenommen, um Veränderungen zu überprüfen.

Wie aus Abbildung 28 (Modell 1a) hervorgeht, beträgt der standardisierte Level-1-Regressionskoeffizient für die Regression von den Vorläuferfähigkeiten (Anfang 1. Schuljahr; Anlaute hören, Laute sprechen, Silben lesen, Buchstaben lesen) auf die Rechtschreibleistungen (Ende 1. Schuljahr) .523 ( $p < .001$ ). Demnach haben vor allem individuelle Vorläuferfähigkeiten einen Effekt auf die Leistungen der Schüler im Rechtschreiben am Ende des ersten Schuljahres. Die standardisierten Regressionskoeffizienten der Klassenführung und des Unterrichtsklimas auf die Rechtschreibleistungen werden hingegen nicht signifikant. Dieser Ergebnistrend bestätigt sich auch dann, wenn jeweils nur eines der beiden Unterrichtsmerkmale auf Level 2 modelliert wird (o. Abb.; Klassenführung:  $\beta = .024$ , *n.s.*; Unterrichtsklima:  $\beta = .125$ , *n.s.*).

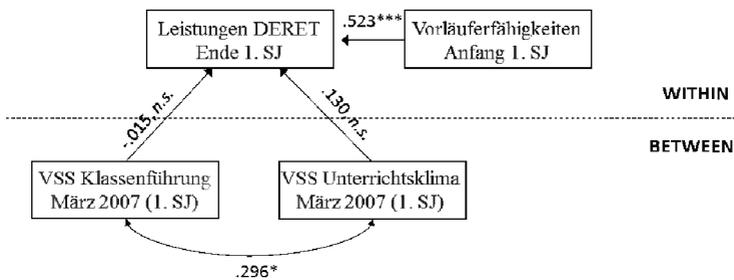


Abbildung 28: Mehrebenenregressionsmodell 1a: Standardisierte Regressionskoeffizienten (AV = Rechtschreibleistungen / DERET Ende 1. SJ)

Anmerkung: \*\*\* $p < .001$ , \* $p < .05$ , *n.s.* = nicht signifikant, VSS = Videostudie Deutsch, SJ = Schuljahr, within = innerhalb der Klassen, between = zwischen den Klassen

Die schriftspracherwerbsspezifischen Vorläuferfähigkeiten vom Anfang des ersten Schuljahres sind am Gesamtmittelwert zentriert (grand mean centered).

### Abhängige Variable Selbstkonzept Schreiben (Ende 1. SJ) – Modell 1b

In Abbildung 29 ist das Mehrebenenregressionsmodell 1b mit den standardisierten Regressionskoeffizienten dargestellt. Als abhängige Variable dient das Selbstkonzept im Schreiben vom Ende des ersten Schuljahres. Auf Level 1 wird zusätzlich das Selbstkonzept im Schreiben vom Anfang des ersten Schuljahres einbezogen (vgl. Kapitel 12.2.3), das erneut am Gesamtmittelwert zentriert wurde. Auf Level 2 gehen erneut die Skalen Klassenführung und Unterrichtsklima aus der Videostudie Deutsch (1. Schuljahr) in das Modell ein.

Der standardisierte Level-1-Regressionskoeffizient für die Regression vom Selbstkonzept Schreiben (Anfang 1. Schuljahr) auf das Selbstkonzept Schreiben (Ende 1. Schuljahr) beträgt .343 ( $p < .001$ ). Zudem zeigt sich ein negativer standardisierter Regressionskoeffizient der Klassenführung auf das mittlere Selbstkonzept im Schreiben am Ende des ersten Schuljahres

( $\beta = -.488, p < .05$ ), nach Kontrolle der Selbstkonzepte vom Anfang des ersten Schuljahres und der gleichzeitigen Modellierung des Unterrichtsklimas auf Level 2. Dieser negative Effekt bleibt auch dann konstant, wenn auf Level 2 das Unterrichtsklima nicht einbezogen wird ( $\beta = -.401, p < .05$ ) (o. Abb.).

Nach Kontrolle der vorherigen Selbstkonzepte im Schreiben (vom Anfang 1. Schuljahr) und der Klassenführung wird der standardisierte Regressionskoeffizient des Unterrichtsklimas auf das Selbstkonzept im Schreiben am Ende des ersten Schuljahres erwartungskonform positiv ( $\beta = .309$ ), jedoch nicht signifikant. Dieser Ergebnistrend bleibt auch dann bestehen, wenn auf Level 2 die Klassenführung nicht mit in das Modell aufgenommen wird ( $\beta = .189, n.s.$ ) (o. Abb.).

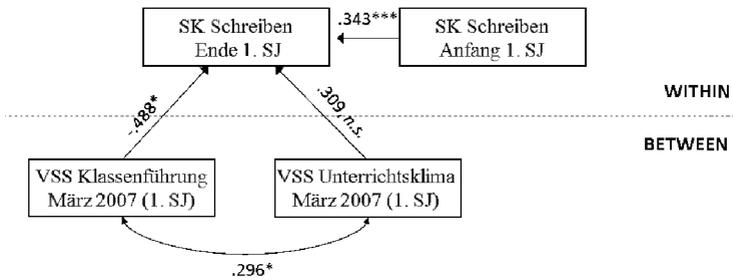


Abbildung 29: Mehrebenenregressionsmodell 1b: Standardisierte Regressionskoeffizienten (AV = Selbstkonzept Schreiben Ende 1. SJ)

Anmerkung: \*\*\* $p < .001$ , \* $p < .05$ , *n.s.* = nicht signifikant, VSS = Videostudie Deutsch, SJ = Schuljahr, SK = Selbstkonzept, within = innerhalb der Klassen, between = zwischen den Klassen

Die Selbstkonzepte im Schreiben vom Anfang des ersten Schuljahres sind am Gesamtmittelwert zentriert (grand mean centered).

## 16.2 Die Bedeutung der Unterrichtsmerkmale im Fach Mathematik (2. SJ)

Da die Videostudie Mathematik im zweiten Schuljahr stattfand, wurden als abhängige Variablen je nach Modell die Leistung in Mathematik bzw. das Selbstkonzept im Rechnen vom Ende des zweiten Schuljahres herangezogen (vgl. Tabelle 35). Wie sich im Vorfeld zeigte, werden aufgrund der Multikollinearität nicht beide Klassenführungsskalen gleichzeitig in die Modelle aufgenommen. Dies hat zur Folge, dass anschließend vier separate Modelle berechnet werden (vgl. Kapitel 16, Tabelle 35).

### Abhängige Variable Leistungen in Mathematik (Ende 2. SJ) – Modelle 2a und 2b

Im Modell 2a (vgl. Abbildung 30) wird als abhängige Variable die Leistungen der Schüler in Mathematik vom Ende des zweiten Schuljahres herangezogen. Auf Level 1 wird um die vorherigen Leistungen der Schüler in Mathematik vom Ende des ersten Schuljahres kontrolliert (vgl. Kapitel 12.2.1, zentriert am Gesamtmittelwert). Als Prädiktorvariablen auf Level 2 gehen das effiziente Zeitmanagement (1. Teilkomponente der Klassenführung) und das Unterrichtsklima aus der Videostudie Mathematik (2. Schuljahr, März 2008) mit ein.

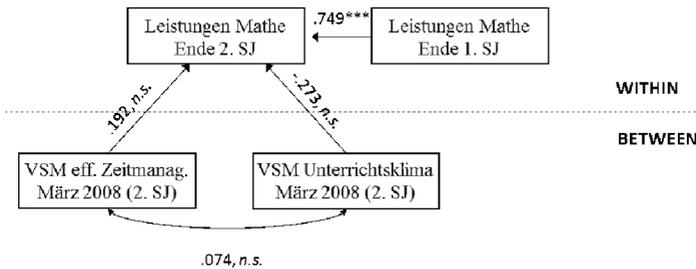


Abbildung 30: Mehrebenenregressionsmodell 2a (AV = Leistungen Mathematik Ende 2. SJ)

Anmerkung: \*\*\* $p < .001$ , n.s. = nicht signifikant, VSM = Videostudie Mathematik, SJ = Schuljahr, within = innerhalb der Klassen, between = zwischen den Klassen, eff. Zeitmanag. = effizientes Zeitmanagement

Die Leistungen in Mathematik vom Ende des ersten Schuljahres sind am Gesamtmittelwert zentriert (grand mean centered).

Im Modell 2a beträgt der standardisierte Level-1-Regressionskoeffizient für die Regression von der vorherigen Leistung in Mathematik (Ende 1. Schuljahr) auf die Leistungen Ende zweites Schuljahr .749 ( $p < .001$ ). Demnach haben auch im Fach Mathematik vor allem frühere individuelle Leistungen einen Einfluss auf die Leistungen der Schüler am Ende des zweiten Schuljahres. Für das Unterrichtsklima zeigt sich nach Kontrolle der vorherigen Leistungen und des effizienten Zeitmanagements ein negativer standardisierter Regressionskoeffizient auf die mathematischen Leistungen der Schüler am Ende des zweiten Schuljahres ( $\beta = -.273$ ), dieser wird jedoch nicht signifikant. Auch das effiziente Zeitmanagement hat keinen signifikanten Effekt auf die Leistungen der Schüler in Mathematik ( $\beta = .192$ , n.s.), geht jedoch in die erwartete Richtung. Die zusätzliche Berechnung von Modellen ohne gleichzeitige Aufnahme beider Prädiktoren ergab ein ähnliches Bild. Die Richtung der Effekte blieb konstant (o. Abb.; effizientes Zeitmanagement:  $\beta = .175$ , n.s.; Unterrichtsklima:  $\beta = -.262$ ,  $p < .10$ ).

Im Modell 2b (vgl. Abbildung 31) wird auf Level 2 neben dem Unterrichtsklima der effektive Umgang mit Störungen (2. Teilkomponente der Klassenführung) als zusätzlicher Prädiktor herangezogen. Nach Kontrolle der vorherigen Leistungen und bei gleichzeitiger Modellierung des Prädiktors effektiver Umgang mit Störungen auf Level 2 ist der Level-2-Regressionskoeffizient für die Regression des Unterrichtsklimas auf die Leistungen in Mathematik Ende zweites Schuljahr erneut negativ und lediglich auf einem 10 %-Niveau signifikant ( $\beta = -.286, p < .10$ ). Der Level 2-Regressionskoeffizient der Regression vom Merkmal effektiver Umgang mit Störungen auf die Leistungen der Schüler in Mathematik am Ende des zweiten Schuljahres liegt bei .163, wird jedoch auch nicht signifikant. Erneut bleibt die Richtung des Effekts auch bei der Berechnung einzelner Modelle, d. h. mit nur einem der hier genannten Prädiktoren auf Level 2, konstant (o. Abb.; effektiver Umgang mit Störungen:  $\beta = .117, n.s.$ ; Unterrichtsklima: s. o.).

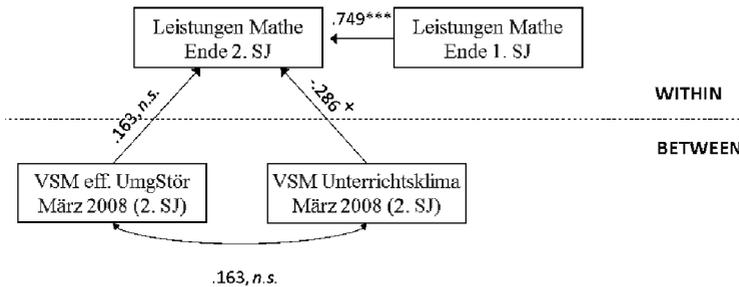


Abbildung 31: Mehrebenenregressionsmodell 2b (AV = Leistungen Mathematik Ende 2. SJ)

Anmerkung: \*\*\* $p < .001$ , + $p < .10$ , n.s. = nicht signifikant, VSM = Videostudie Mathematik, SJ = Schuljahr, within = innerhalb der Klassen, between = zwischen den Klassen, eff. UmgStör = effektiver Umgang mit Störungen

Die Leistungen in Mathematik vom Ende des ersten Schuljahres sind am Gesamtmittelwert zentriert (grand mean centered).

*Abhängige Variable Selbstkonzept Rechnen (Ende 2. SJ) – Modelle 2c und 2d*

Im Modell 2c wird anstelle der Leistungen in Mathematik das Selbstkonzept im Rechnen vom Ende des zweiten Schuljahres als abhängige Variable auf Level 1 herangezogen. Auf Level 1 wird um die vorherigen Selbstkonzepte der Schüler im Rechnen kontrolliert. Der Level-1-Prädiktor geht erneut als „grand-mean-zentrierte“ Variable in das Modell ein. Auf Level 2 werden, wie bereits im Modell 2a, die Prädiktoren effizientes Zeitmanagement und Unterrichtsklima gleichzeitig in das Modell aufgenommen. In Abbildung 32 sind die Ergebnisse dieses Mehrebenenregressionsmodells dargestellt. Der standardisierte Level-1-Regressionskoeffizient beträgt im Modell 2c für die Re-

gression von den vorherigen Selbstkonzepten (Ende 1. Schuljahr) auf die abhängige Variable  $.555$  ( $p < .001$ ). Demnach wird das Selbstkonzept im Rechnen vom Ende des zweiten Schuljahres zu einem großen Teil durch frühere individuelle Selbstkonzepte vom Ende des ersten Schuljahres vorhergesagt.

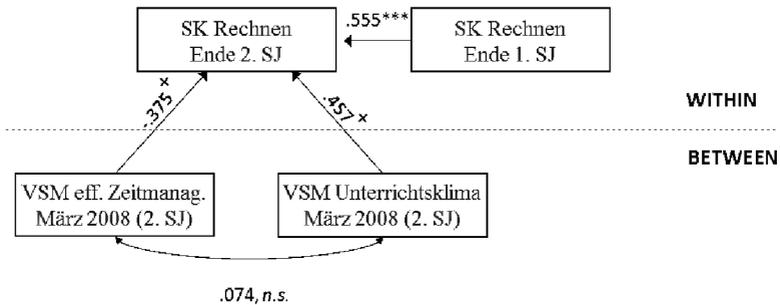


Abbildung 32: Mehrebenenregressionsmodell 2c (AV = Selbstkonzept Rechnen Ende 2. SJ)

Anmerkung: \*\*\* $p < .001$ , \* $p < .10$ , *n.s.* = nicht signifikant, VSM = Videostudie Mathematik, SJ = Schuljahr, SK = Selbstkonzept, within = innerhalb der Klassen, between = zwischen den Klassen, eff. Zeitmanag. = effizientes Zeitmanagement

Die Selbstkonzepte im Rechnen vom Ende des ersten Schuljahres sind am Gesamtmittelwert zentriert (grand mean centered).

Für die Regression des Unterrichtsklimas auf das mittlere Selbstkonzept im Rechnen am Ende des zweiten Schuljahres ergibt sich ein positiver standardisierter Regressionskoeffizient ( $\beta = .457$ ), der jedoch lediglich auf einem 10 %-Niveau signifikant wird. Für die Klassenführungskomponente „effizientes Zeitmanagement“ zeigt sich ein negativer Regressionskoeffizient auf das Selbstkonzept im Rechnen am Ende des zweiten Schuljahres ( $\beta = -.375$ ). Auch dieser Effekt wird lediglich auf einem 10 %-Niveau signifikant. Dieser Ergebnistrend bleibt auch dann konstant, wenn beide Prädiktoren einzeln auf Level 2 aufgenommen werden (o. Abb.). Der Effekt des Unterrichtsklimas wird in diesem Fall allerdings signifikant ( $\beta = .432$ ,  $p < .05$ ), während der Effekt des effektiven Zeitmanagements erneut lediglich auf einem 10 %-Niveau signifikant wird ( $\beta = -.345$ ).

Im Modell 2d wird im Vergleich zu Modell 2c lediglich die Prädiktorvariable effizientes Zeitmanagement gegen die zweite Klassenführungskomponente effektiver Umgang mit Störungen ausgetauscht (vgl. Abbildung 33). Die abhängige Variable sowie die Kontrollvariable auf Level 1 und das Unterrichtsklima als Level-2-Prädiktor bleiben im Modell erhalten. Nach Kontrolle des Unterrichtsklimas und des vorherigen Selbstkonzepts liegt der Level-2-Regressionskoeffizient des effektiven Umgangs mit Störungen auf das Selbstkonzept Rechnen vom Ende des zweiten Schuljahrs bei  $-.415$  ( $p < .05$ ,

vgl. Abbildung 33). Es ergibt sich erneut ein positiver Effekt des Unterrichtsklimas auf das Selbstkonzept im Rechnen am Ende des zweiten Schuljahres ( $\beta = .497, p < .05$ ), nach Kontrolle der weiteren Merkmale.

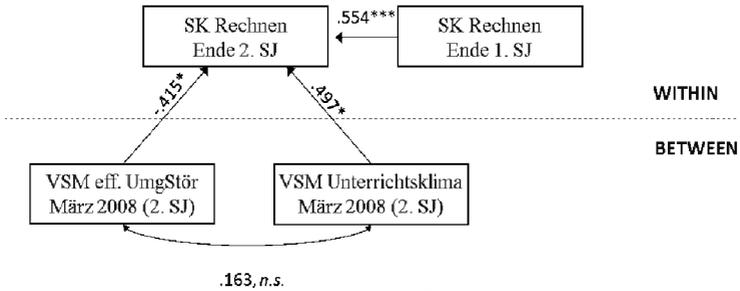


Abbildung 33: Mehrebenenregressionsmodell 2d (AV = Selbstkonzept Rechnen Ende 2. SJ)

Anmerkung:  $***p < .001$ ,  $*p < .05$ , *n.s.* = nicht signifikant, VSM = Videostudie Mathematik, SJ = Schuljahr, SK = Selbstkonzept, within = innerhalb der Klassen, between = zwischen den Klassen, eff. UmgStör. = effizienter Umgang mit Störungen

Die Selbstkonzepte im Rechnen vom Ende des ersten Schuljahres sind am Gesamtmittelwert zentriert (grand mean centered).

## V Zusammenfassung und Diskussion

Die vorliegende Arbeit entstand im Rahmen der PERLE-Studie. Die Bedeutung und Umsetzung einer effektiven Klassenführung und eines positiven Unterrichtsklimas wurde in 101 gefilmten Lektionen aus dem Deutschunterricht (1. Schuljahr) und dem Mathematikunterricht (2. Schuljahr) empirisch untersucht. Zunächst wurden die zwei Basisdimensionen guten Unterrichts als allgemeine Unterrichtsqualitätsmerkmale hoch inferent operationalisiert (vgl. Kapitel 12). Ziel dieser Arbeit war es, ein hoch inferentes Ratingsystem zu entwickeln, das für die Fächer Deutsch und Mathematik (sowie Kunst) eingesetzt werden kann. Nach Überprüfung der fachunspezifischen sowie fachspezifischen Inter-Rater-Reliabilität (Kapitel 13) wurde der Deutsch- und Mathematikunterricht im Hinblick auf die Ausprägung der einzelnen Merkmale der Klassenführung und des Unterrichtsklimas inhaltlich beschrieben (Studie 1, Kapitel 14). Da beide Basisdimensionen durch mehrere hoch inferente Merkmale operationalisiert wurden, erfolgte auf Basis der hoch inferent erfassten Daten zudem eine Überprüfung der Faktorenstruktur (Studie 2, Kapitel 15). Im Sinne der prognostischen Validität ermöglichten es die Ergebnisse der Mehrebenenregressionsanalysen (Studie 3, Kapitel 16) zudem, die Bedeutung der beiden Basisdimensionen effektive Klassenführung und positives Unterrichtsklima zu untersuchen. Dabei wurde untersucht, inwieweit die beiden Basisdimensionen Effekte auf multikriteriale (kognitive und nicht-kognitive) Schülermerkmale in den zwei Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) aufweisen.

Dieses Kapitel dokumentiert die Diskussion der Ergebnisse. Die Diskussion orientiert sich an den in Kapitel 7 dargestellten Forschungsfragen. Zuerst werden die Ergebnisse der fachunspezifischen und fachspezifischen Reliabilitätsprüfung zusammengefasst und diskutiert (Kapitel 17) und im Anschluss erfolgt die Diskussion der Ergebnisse der drei Teilstudien (Kapitel 18).

Die Interpretation sowie Diskussion der Ergebnisse sollte auf Grundlage der hier untersuchten Stichprobe und der damit einhergehenden Einschränkungen erfolgen. Zwar gab ein Großteil der Lehrpersonen an, dass die Repräsentativität ihres aufgezeichneten Unterrichts gewährleistet ist (vgl. Kapitel 11.3), jedoch muss an dieser Stelle die Stichprobensammensetzung allgemein diskutiert werden, denn die Stichprobe setzt sich sowohl aus Lehrpersonen an staatlichen Grundschulen als auch an BIP-Kreativitätsgrundschulen zusammen. Da die Teilnahme der BIP-Kreativitätsgrundschulen verpflichtend war – die PERLE-Studie diente hauptsächlich der Evaluation dieses Schultyps – ist die Repräsentativität der Stichprobe nicht gewährleistet. Zudem handelt es sich im Falle der BIP-Schulen um Privatschulen. Die El-

tern müssen für den Besuch ihrer Kinder Schulgeld zahlen. Dies hat Folgen für die Höhe des sozioökonomischen Status der Stichprobe, der, wie bereits in Kapitel 11.2 dargestellt, vergleichsweise hoch ist. Im Vergleich dazu war die Teilnahme der staatlichen Schulen freiwillig. Aus diesem Grund kann vermutet werden, dass es sich auch hier um eine selektive Stichprobe von engagierten Lehrpersonen / Schulen handelt. Die hier dargestellten Ergebnisse beziehen sich demnach nur auf die vorliegende Stichprobe und ermöglichen keine darüber hinausgehenden allgemeinen Aussagen über die Qualität des Grundschulunterrichts in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr).

## 17 Inter-Rater-Reliabilität

Resümierend kann festgehalten werden, dass die in Kapitel 3.2 und 4.2 beschriebenen Merkmale der Klassenführung und des Unterrichtsklimas – mit Ausnahme einzelner Merkmale – in den aufgezeichneten Unterrichtslektionen im Anfangsunterricht (Fächer Deutsch und Mathematik) reliabel durch die zwei Rater beurteilt werden konnten und somit das entwickelte Rating-system grundsätzlich brauchbar ist, um Daten auf Basis hoch inferenter Beurteilungen für den Anfangsunterricht zu generieren. Jedoch resultierten in Abhängigkeit vom Fach unterschiedlich hohe relative G-Koeffizienten und damit verbunden Unterschiede in der Höhe der systematischen und unsystematischen Messfehleranteile. Demnach scheinen einzelne Merkmale in Abhängigkeit vom Fach für die Rater weniger deutlich sichtbar zu sein, was wiederum ein hohes Maß an Interpretation mit sich bringt (vgl. Clausen et al., 2003). Relative Schwächen ergaben sich allgemein vor allem für die Merkmale „Vorbereitung des Klassenraums (Vorb)“, „positiver Umgang der Lehrperson mit Fehlern (Fehl)“, „positives Fehlerklima zwischen den Schülern (Fekl)“ sowie „Wettbewerb und Leistungsdruck (Wett)“. Hier besteht Verbesserungspotenzial beispielsweise hinsichtlich konzeptueller Schärfung und/oder des Antwortformats, worauf im Folgenden genauer eingegangen wird.

### 17.1 Ergebnisse der fachunspezifischen Inter-Rater-Reliabilität

Betrachtet man die Ergebnisse der *fachunspezifischen* G-Studien, so erreichen lediglich zwei der insgesamt 18 erfassten Merkmale der Klassenführung und des Unterrichtsklimas keine zufriedenstellende Inter-Rater-Reliabilität

( $g_{relativ} \geq .65$ ). Hierbei handelt es sich um die Unterrichtsklimamerkmale „positiver Umgang der Lehrperson mit Fehlern (Fehl)“ und „positives Fehlerklima zwischen den Schülern (Fekl)“. Die Betrachtung der Varianzkomponenten für diese zwei Merkmale zeigt, dass in beiden Fällen nur eine geringe „wahre“ Variabilität besteht (28 % bzw. 31 %), demnach entfällt nur ein geringer Varianzanteil auf die Varianzkomponente „Video“. Die Generalisierbarkeit ist somit sehr gering. In Anlehnung an Ergebnisse aus der Sekundarstufe (vgl. Kapitel 4.2), kann die sehr geringe „wahre“ Variabilität vermutlich darauf zurückgeführt werden, dass auch für den Anfangsunterricht (Deutsch / Mathematik) eine öffentliche Thematisierung von Fehlern bzw. eine Reaktion auf Fehler eher selten im Unterricht zu beobachten ist. Aus methodischer Sicht kann das „Nicht-Auftreten“ von Fehlersituationen zu Schwierigkeiten in der reliablen Einschätzung durch die zwei Rater führen, da aufgrund der geringen Anzahl an beobachtbaren Situationen deutlich mehr interpretative Schlussfolgerungen von den Ratern notwendig sind. Zwar sind eine positive Fehlerkultur bzw. ein positiver Umgang der Lehrperson und der Schüler mit Fehlern aus theoretischer Sicht von besonderer Relevanz (vgl. Kapitel 4.2), sie lassen sich jedoch mithilfe des hoch inferenten Ratings nur schwer für den Anfangsunterricht erfassen. Damit können bisherige Befunde aus der Sekundarstufe bestätigt werden (z. B. Meyer et al., 2006). Dieser Diskussionspunkt wird ausführlich noch einmal in Kapitel 17.2 aufgegriffen, da sich unter anderem für diese zwei Merkmale des Fehlerklimas Unterschiede in der Höhe der relativen G-Koeffizienten zwischen den zwei untersuchten Fächern / Schuljahren zeigten.

Die geringe „wahre“ Variabilität führt automatisch zu einem ungünstigen Verhältnis von systematischen und unsystematischen Fehleranteilen (vgl. Clausen et al., 2003). Zwar stellen die zwei Rater an sich – mit einem Varianzanteil von 2 bzw. 7 % – eine eher zu vernachlässigende Fehlerquelle dar, jedoch weist der hohe unsystematische Messfehleranteil (Wechselwirkungsvarianz Video\*Rater) darauf hin, dass möglicherweise weitere Messfehler bei der Einschätzung dieser zwei Merkmale des Unterrichtsklimas eine Rolle spielen, die im Rahmen der vorliegenden Arbeit nicht weiter spezifiziert werden können. Eine mögliche Erklärung für den hohen Anteil an systematischer Varianz könnte darauf zurückgeführt werden, dass es schwierig ist, hoch inferent und reliabel einzuschätzen, in welchen Situationen die Lehrperson einen Fehler thematisieren bzw. nicht thematisieren sollte.

Es bleibt an dieser Stelle zu diskutieren, wie angemessen das entwickelte hoch inferente Ratingsystem in der Lage war, die Aspekte der Klassenführung und des Unterrichtsklimas in verschiedenen Fächern zu erfassen.

## 17.2 Ergebnisse der fachspezifischen Inter-Rater-Reliabilität

Bei der Berechnung der *fachspezifischen* Inter-Rater-Reliabilität zeigte sich, dass je nach Merkmal teilweise deutliche Unterschiede in der Höhe der relativen G-Koeffizienten bestehen. Anschließend werden die problematischen Merkmale diskutiert, beginnend mit dem Merkmal „positiver Umgang der Lehrperson mit Fehlern (Fehl)“, das in beiden Fächern das gesetzte Kriterium von  $g_{relativ} \geq .65$  für eine zufriedenstellende Übereinstimmung der zwei Rater nicht erreicht. Ausdruck der geringen Übereinstimmung ist in beiden Fächern der geringe Varianzanteil Video (29 % bzw. 30 %). 71 % der beobachteten Variabilität geht in beiden Fächern zulasten von systematischen und unsystematischen Fehleranteilen, wobei im Fach Mathematik (2. Schuljahr) der systematische Fehleranteil mit 18 % sehr hoch ist. Demnach unterscheiden sich die Rater vor allem im Fach Mathematik (2. Schuljahr) in ihrer allgemeinen Wahrnehmung bezüglich ihrer Skalenhandhabung, d. h. sie sind unterschiedlich streng bzw. mild.

Neben dem ungünstigen Verhältnis der „wahren“ Varianz zu den systematischen und unsystematischen Fehlervarianzen für das Merkmal „positiver Umgang der Lehrperson mit Fehlern (Fehl)“, hängt die nicht zufriedenstellende Inter-Rater-Reliabilität in beiden Fächern vermutlich mit der Operationalisierung des Merkmals und dem Antwortformat zusammen. Die Beschreibung der Grundidee war stark an bereits existierenden Rating- bzw. Beobachtungssystemen aus der Sekundarstufe orientiert. Dabei wurde auf eine systematische Identifizierung von Fehlern verzichtet. Die systematische Erfassung von Schülerfehlern scheint jedoch wichtig, um Aussagen über den Umgang der Lehrperson mit diesen hoch inferent einzuschätzen. Ohne eine beobachtbare Reaktion der Lehrperson auf einen Schülerfehler hatten die Rater, trotz intensiver Schulung, scheinbar Schwierigkeiten, Fehler oder Misskonzepte der Schüler als solche zu identifizieren. Dass vor allem das Fehlerklima differenzierter erfasst werden sollte, belegt eine aktuelle Studie von Steuer, Rosentritt-Brunn und Dresel (2013). In der Studie zeigte sich, dass sich aus Schüler-sicht unterschiedliche Teilfacetten des Fehlerklimas (z. B. Fehlertoleranz der Lehrperson, Unterstützung der Lehrperson nach Fehlern oder Fehleranalyse) identifizieren lassen. Auch für die videobasierte Erfassung des Fehlerklimas können die Ergebnisse dieser Studie Impulse für eine stärkere Ausdifferenzierung geben.

Die Rolle nicht ausreichend vorhandener fachdidaktischer Kenntnisse der Rater im entsprechenden Fach kann an dieser Stelle als Nachteil angesehen werden. Hier sollte ergänzend eine niedrig inferente Kodierung durchgeführt werden, im Rahmen derer Schülerfehler in beiden Fächern systematisch – im Idealfall von Fachdidaktikern – identifiziert werden. Erst dann könnte in einem weiteren Schritt hoch inferent eingeschätzt werden, ob die Lehrperson auf die einzelnen Schülerfehler eingeht und diese konstruktiv nutzt oder

ob sie Fehler übergeht bzw. nicht aufgreift und somit tendenziell ein negatives Fehlerklima schafft. Da mithilfe einer solchen niedrig inferenten Kodierung Daten zur Häufigkeit von Schülerfehlern im Unterricht generiert werden können, könnte im Rahmen der PERLE-Studie sowohl für das Fach Deutsch (1. Schuljahr) als auch für das Fach Mathematik (2. Schuljahr) untersucht werden, wie viele Fehlersituationen im Unterricht überhaupt vorkommen.

Es stellt sich die Frage, ob schwerwiegende Fehler oder Misskonzepte der Schüler im öffentlichen Anfangsunterricht überhaupt zum Vorschein kommen. Es kann vermutet werden, dass solche Fehler, die als Fenster zum Denken der Kinder angesehen werden können, vermutlich eher selten auftreten. Für das Fach Mathematik bestätigt Heinze (2005) beispielsweise für die Sekundarstufe eine eher geringe Anzahl an Fehlern pro Unterrichtsstunde. In seiner Studie konnten im Durchschnitt nur 4.7 Fehler pro Unterrichtsstunde identifiziert werden, wobei es auch Klassen gab, in denen durchschnittlich 0 bzw. 12 Fehler gemacht wurden. Heinze (2005) erklärt die geringe Anzahl an Fehlersituationen in Mathematik mit dem typischen Stil einer Mathematikstunde, die häufig in Form eines fragend-entwickelnden Unterrichts durchgeführt wird. In einem solchen Unterricht sind „die Fragen, die der Lehrer stellt, [...] überwiegend konvergent, das heißt, sie zielen auf eine ganz bestimmte Antwort ab“ (Klieme et al., 2001, S. 45). Für die Diskussion von Fehlern ist in einem fragend-entwickelten Unterrichtsgespräch kein Platz und somit sind auch nur geringe Fehlerzahlen zu beobachten (vgl. Heinze, 2004). Auch für den im Kontext der PERLE-Studie aufgezeichneten Mathematikunterricht im zweiten Schuljahr zeigt sich, dass die Einführung der Multiplikation eher nach einem fragend-entwickelnden Unterrichtsskript und somit eher lehrerzentriert erfolgt (vgl. Kapitel 9.3). In diesem Zusammenhang stellt sich die Frage, ob in einem Unterricht, der beispielsweise nicht die Einführung einer neuen Rechenart, sondern das Lösen von mathematisch anspruchsvolleren Aufgaben beinhaltet, Misskonzepte oder Fehler der Schüler eher zum Vorschein kommen. Vermutlich wären Unterschiede zwischen den Lehrpersonen im Umgang mit Fehlern in einem solchen Unterricht dann deutlicher sichtbar. Dies gilt es in weiteren Studien zu überprüfen.

Unberücksichtigt blieben bei der Operationalisierung außerdem nonverbale Reaktionen der Lehrperson auf Schülerfehler, die über das Drannehmen anderer Schüler zur Fehlerkorrektur hinausgehen. So kann vor allem für die lehrerzentrierten Phasen vermutet werden, dass die Reaktion der Lehrperson auf Schülerfehler viel häufiger nonverbal erfolgt, zu Beispiel um im Lehrstoff voranzukommen. Im Kontext weiterführender Studien sollten aus diesem Grund weitere Formen der nonverbalen Reaktion der Lehrperson auf Schülerfehler berücksichtigt werden. Beispielsweise markiert das Verschränken der Arme, das ungeduldige Wippen eines Fußes oder das Abwenden der Lehrperson von einem Schüler eher einen negativen Umgang mit Fehlern, da der Schüler kaum über eine falsche Aussage nachdenken kann oder die Lehrperson durch das Abwenden Desinteresse signalisiert. Berücksichtigt wurde

im vorliegenden Fall lediglich das Drannehmen anderer Schüler als eine mögliche nonverbale Reaktion, die als negativ im Umgang der Lehrperson mit Schülerfehlern angesehen wird.

An diesem Punkt sollte zudem die Frage gestellt werden, ob die geringe Generalisierbarkeit der Merkmale des Fehlerklimas darauf zurückgeführt werden kann, dass von den Ratern innerhalb dieser Merkmale mehrere inhaltlich unterschiedliche Aspekte zu einem Urteil integriert werden mussten. Dadurch ist der Grad der interpretativen Schlussfolgerungen für die Rater deutlich erhöht. So sollten die Rater nicht nur beurteilen, ob die Lehrperson allgemein eine positive Fehlerkultur schafft, sondern zusätzlich, ob auftretende Fehler von der Lehrperson konstruktiv genutzt werden, um mangelndes oder unzutreffendes Verständnis zu korrigieren. Es kann angenommen werden, dass die Verbesserung der Reliabilität dieser hoch inferenten Klimamerkmale unter anderem dann gelingen könnte, wenn nicht mehrere inhaltlich unterschiedliche Aspekte in einem Merkmal erfasst werden. Allerdings geht damit „auch ein spezifisches Merkmal hoch inferenter Ratingsysteme, dass nämlich mehrere Aspekte zu einem Globalurteil integriert werden und dadurch Tiefenstrukturen besser erfasst werden können, verloren“ (Lotz et al., 2013, S. 377).

Bei dem Merkmal „positives Fehlerklima zwischen den Schülern (Fekl)“ ging es darum, dass die Rater das Fehlerklima zwischen den Schülern einschätzen sollten. Dieses Merkmal konnte allerdings nur für das Fach Deutsch (1. Schuljahr) reliabel von den beiden Ratern eingeschätzt werden. Der niedrige relative G-Koeffizient für dieses Merkmal bei der fachunspezifischen Analyse über alle Unterrichtsvideos resultierte also nur daraus, dass für das Fach Mathematik (2. Schuljahr) kaum „wahre“ Variabilität in den Daten vorhanden ist. Für das Fach Deutsch (1. Schuljahr) ist der Anteil an „wahrer“ Varianz deutlich höher (vgl. Kapitel 17.2; Tabelle 22). Demnach eignen sich hoch inferente Beurteilungen zur Einschätzung des Fehlerklimas zwischen den Schülern je nach Fach unterschiedlich gut. Allerdings sollte dieses Ergebnis in weiteren Studien erneut bestätigt werden. Außerdem wird an dieser Stelle deutlich, dass es durchaus sinnvoll ist, die Generalisierbarkeit (bzw. Zuverlässigkeit) fachspezifisch zu überprüfen bzw. in weiteren Untersuchungen das Fach als sogenannte fixierte Facette in den G-Studien zu integrieren.<sup>95</sup>

95 Man kann im Rahmen von G-Studien verschiedene Arten von Fehlerquellen (Facetten) unterscheiden (zusef. Praetorius, 2014), Zufallsfacette und fixierte Facette. Im Rahmen der vorliegenden Arbeit wurde bislang nur die sogenannte Zufallsfacette berücksichtigt, d. h. eine zufällige Auswahl an Ratern hat die Merkmale der Klassenführung und des Unterrichtsklimas eingeschätzt. Die Facette liegt variiert vor (= Idealfall). Wenn man jedoch – wie im vorliegenden Fall – zusätzlich daran interessiert ist, Unterschiede zwischen den Fächern Deutsch und Mathematik zu untersuchen, bietet es sich an, in einer weiteren Studie eine fixierte Facette Unterrichtsfach mit lediglich zwei Ausprägungen (Deutsch / Mathematik) zu integrieren (vgl. Cronbach et al., 1972). Dies wurde bislang nicht berücksichtigt und sollte selbstverständlich auch für alle anderen Merkmale der Klassenführung und des Unterrichtsklimas überprüft werden.

Ähnliche Unterschiede in der Höhe der relativen G-Koeffizienten in Abhängigkeit vom Fach ergaben sich für die Merkmale „Zeitmanagement / effektive Zeitnutzung (Zeit)“ und „Übergangsmanagement (Ueb)“, allerdings zugunsten von Mathematik. Inhaltlich bedeutet dieser Unterschied, dass in der Videostudie Mathematik (2. Schuljahr) sowohl das „Zeitmanagement / effektive Zeitnutzung (Zeit)“ als auch das „Übergangsmanagement (Ueb)“ mit einer hohen Übereinstimmung zwischen den beiden Ratern eingeschätzt werden konnte, während dieselben zwei Rater unter Verwendung des gleichen Ratingsystem in der Videostudie Deutsch (1. Schuljahr) geringere Übereinstimmungen erzielten. Auch für das Merkmal „Regelklarheit bzw. -verwendung (Regl)“ zeigten sich in der vorliegenden Studie deutliche Unterschiede in der Höhe der relativen G-Koeffizienten zwischen den Fächern. Während die zwei Rater dieses Merkmal erneut für die Videostudie Mathematik (2. Schuljahr) reliabel einschätzen konnten ( $g = .95$ ), lag der relative G-Koeffizient für die Videostudie Deutsch (1. Schuljahr) nur bei .66.

Im Fall der zuvor erwähnten drei hoch inferenten Merkmale wird vermutlich die Konfundierung von Fach / Videostudie und Schuljahr deutlich. Es kann angenommen werden, dass vor allem im ersten Schuljahr (Fach Deutsch), das als eine Art Anpassungsphase definiert werden kann, die Einhaltung von Regeln, aber auch ein effektives Zeit- und Übergangsmanagement noch nicht so bedeutsam sind wie ein Schuljahr später (Fach Mathematik). Das Zeitmanagement im ersten Schuljahr zielt vermutlich noch nicht so sehr auf eine effektive Zeitnutzung und damit ein aktives Lernen bzw. eine höhere Aufmerksamkeit und Konzentration der Schüler ab. Damit stellt sich an dieser Stelle die Frage, ob die Unterschiede in der Höhe des relativen G-Koeffizienten ein Problem des Messinstruments oder des jeweiligen Konstrukts sind, das je nach Kontext und Zeitpunkt nicht vergleichbar zu sein scheint. In Bezug auf das Merkmal „Regelklarheit bzw. -verwendung (Regl)“ muss beispielsweise der Zeitpunkt der Erhebung / Beobachtung berücksichtigt werden. Hier kann unterschieden werden zwischen (1) dem Zeitpunkt, in dem die Schüler die Regeln und Prozeduren des Schulalltags lernen und festigen und (2) dem Zeitpunkt, an dem die Regeln und Prozeduren so gefestigt und verinnerlicht sind (vgl. Boostrom, 1991; Carter & Doyle, 2006), dass ein Verweis auf Regeln durch die Lehrperson im Unterricht eher selten zu beobachten ist. Folgt man dieser Argumentation, dann erfolgte die Operationalisierung des Merkmals „Regelklarheit bzw. -verwendung (Regl)“ im vorliegenden Fall eher für den Zeitpunkt (2). Dies könnte erklären, warum die Rater Schwierigkeiten bei der Einschätzung der Regelklarheit im Fach Deutsch (1. Schuljahr) hatten und somit der relative G-Koeffizient niedriger ausfällt.

Das Merkmal „Vorbereitung des Klassenraums (Vorb)“ konnte nur für die Videostudie Deutsch (1. Schuljahr) reliabel durch die zwei Rater eingeschätzt werden, wobei die Höhe des relativen G-Koeffizienten mit .67 nur knapp die Grenze von .65 überschreitet. Im Fach Mathematik (2. Schuljahr) konnte aufgrund der geringen Varianz zwischen den Videos kein relativer

G-Koeffizient berechnet werden. Die Rater schätzten für einen Großteil der Videos die Vorbereitung des Klassenraums durch die Lehrperson im Mittel mit „4“ ein (vgl. Kapitel 21). Um zukünftig ausreichend Varianz zu erzeugen, sollten unter anderem die Antwortstufen „1“ und „2“ deutlicher voneinander abgegrenzt werden. Bei abschließender Betrachtung dieser beiden Antwortstufen fiel auf, dass diese sich inhaltlich kaum unterscheiden. Für weitere Studien sollte zudem grundsätzlich überlegt werden, ob das Merkmal „Vorbereitung des Klassenraums (Vorb)“, das theoretisch eine wichtige Bedeutung im Zusammenhang mit einem effektiven ‚Classroom Management‘ erfährt (vgl. Kapitel 3.2), ein sinnvolles Merkmal für die Erfassung der Klassenführung mittels Videografie darstellt. Sowohl die standardisierten inhaltlichen Vorgaben durch PERLE als auch das Wissen der Lehrperson um die Aufzeichnung ihres Unterrichts haben vermutlich positive Auswirkungen auf die Ausprägung dieses Merkmals, egal für welches Fach sie erfasst wird. Außerdem kann angenommen werden, dass die Lehrperson selbst wenig Einfluss auf die Größe und Ausstattung des Klassenraums nehmen kann (Problem der materiellen und finanziellen Ausstattung einer Schule) und sich mit den gegebenen Situationen arrangiert.

Ähnlich wie bei dem Merkmal „Vorbereitung des Klassenraums (Vorb)“ zeigten sich auch für das Merkmal „Wettbewerb und Leistungsdruck (Wett)“ Unterschiede in der Höhe der relativen G-Koeffizienten zugunsten des Fachs Deutsch. Während in der Videostudie Deutsch (1. Schuljahr) die Rater zu einer deutlich reliablen Einschätzung gelangen ( $g_{relativ} = .80$ ), liegt der relative G-Koeffizient für das Fach Mathematik bei .61 und verfehlt somit die zuvor gesetzte Grenze für eine zufriedenstellende Beobachterübereinstimmung. Als Grund kann erneut ein ungünstigeres Verhältnis der „wahren“ Variation zu den systematischen und unsystematischen Fehlervarianzen für das Fach Mathematik (2. Schuljahr) aufgezeigt werden. Für beide Fächer zeigt sich, dass die Antwortstufen nicht ausreichend Varianz erzeugen, da sowohl in Deutsch (1. Schuljahr) als auch in Mathematik (2. Schuljahr) die Rater einen Großteil des beobachteten Unterrichts dahingehend einschätzten, dass kaum bis gar kein Wettbewerbs- bzw. Leistungsdruck ersichtlich ist. Dieses Ergebnis überrascht, da vor allem in den BIP-Schulen hohe Leistungsanforderungen an die Schüler gestellt werden. Beispielsweise werden bereits ab dem ersten Schuljahr in beiden Fächern Noten vergeben. Für diejenigen Schüler mit der Note 3 werden individuelle Förderpläne erstellt (vgl. Lipowsky et al., 2013)<sup>96</sup>. Hier stellt sich erneut die Frage, ob die Vorgaben an die Lehrpersonen für beide Videostudien zu dieser geringen Ausprägung des Merkmals führten. Die Engführung des Unterrichts und somit die Fokussierung des Unterrichts auf ein bestimmtes Ziel lassen es kaum zu, dass Lehrpersonen die Leistungen der Schüler öffentlich machen.

96 Einige BIP-Einrichtungen distanzieren sich jedoch zunehmend von diesem Ziel (ausführlich siehe Hadeler, 2013).

*Rater-Haupteffekt (VK Rater) – Einfluss des Rater-Bias*

Im Rahmen der vorliegenden Arbeit zeigte sich, dass der Rater-Haupteffekt, also der relative Anteil der Varianz, der auf Unterschiede in den Beurteilungen durch die Rater zurückgeführt werden kann (= systematische Messfehler), sowohl bei der fachunspezifischen als auch bei der fachspezifischen Analyse eher gering ist. Beurteilungsfehler durch die Rater stellten demnach in der vorliegenden Arbeit eine fast vernachlässigbare Fehlerquelle dar (vgl. Kapitel 13.1 und 13.2). Das bedeutet, dass sich die Rater in ihrer allgemeinen Wahrnehmung bezüglich einzelner Merkmale nur minimal unterscheiden (die Rater sind ähnlich streng/mild).

Es bleibt durch die hier durchgeführten Analysen jedoch unklar, welche Unterrichtssituationen die Rater im Rahmen des hoch inferenten Ratings für ihr Gesamturteil herangezogen haben. Zwei Rater können beispielsweise das Zeitmanagement bzw. die effiziente Zeitnutzung für eine Lehrperson übereinstimmend mit „3“ beurteilen, es kann jedoch nicht nachvollzogen werden, ob die Rater einzelne Situationen unterschiedlich gewichten bzw. unterschiedliche Situationen für ihr Urteil berücksichtigen. Solange diese am Ende zu einer übereinstimmenden Beurteilung führen, wird von einer hohen Reliabilität gesprochen (vgl. Lotz et al., 2013). Pietsch und Tosana (2008) kritisieren in diesem Zusammenhang, dass eine hohe Übereinstimmung zwischen den Ratern häufig nur wenig über die tatsächliche Qualität der Messung aussagt, da die Einigkeit der beiden Rater nicht zwangsläufig gleichbedeutend mit der tatsächlichen Realität ist. Helmke (2009) verweist zusätzlich auf die Möglichkeit des „kollektiven Irrtums“ (S. 300). Trotz dieser Kritik wird die übereinstimmende Beurteilung durch mehrere Rater als hinreichender Indikator für die Qualität der Messung angesehen.

*Interaktion zwischen Rater und beobachtetem Unterricht (VK Video x Rater)*

Wie bereits im Vorfeld deutlich wurde, entfällt sowohl bei der fachunspezifischen als auch bei der fachspezifischen Berechnung ein relativ großer Teil der Varianz auf die Interaktion zwischen Ratern und Video (beobachtetem Unterricht in Deutsch oder Mathematik, = unsystematischer Messfehleranteil). Damit ist dieser Fehleranteil für einen Großteil der hoch inferenten Merkmale der Klassenführung und des Unterrichtsklimas bedeutungsvoller als die Rater-Effekte an sich (vgl. Kapitel 13.2). Da bislang kaum Studien existieren, die solche Interaktionseffekte berichten (zusf. Praetorius et al., 2012), werden die Ergebnisse an dieser Stelle zusätzlich diskutiert. Die teilweise sehr hohen unsystematischen Messfehleranteile deuten darauf hin, dass sich die beiden Rater dahingehend uneinig sind, ob ein bestimmter beobachteter Unterricht gemittelt über alle eingeschätzten Merkmale gut oder weniger gut ist. Es kann vermutet werden, dass trotz eines intensiven Trainings implizite Theorien darüber existieren, was unter einer effektiven Klassenführung bzw. einem positiven Unterrichtsklima zu verstehen ist. Ähnlich argumentiert auch Praetorius (2014). Solche persönlichen Überzeugungen können ihrer Meinung nach in

die hoch inferente Beurteilung entsprechenden Unterrichts einfließen. In zukünftigen Studien sollten im Vorfeld implizite Theorien der Rater erfasst und im Rahmen des Trainings angesprochen und gegebenenfalls verändert bzw. erweitert werden. Dadurch lassen sich vermutlich Einflüsse solcher impliziter Theorien minimieren, jedoch nie ganz eliminieren (ebd.).

### *Zwischenfazit*

An dieser Stelle kann festgehalten werden, dass die hoch inferenten Beurteilungen der einzelnen Merkmale der Klassenführung und des Unterrichtsklimas mit wenigen Ausnahmen eine gute bis sehr gute Messqualität in beiden Fächern aufweisen. Die Brauchbarkeit des hoch inferenten Ratingsystems für beide Fächer konnte mithilfe der durchgeführten Analysen belegt werden. Demnach bedarf es bei der Beurteilung der meisten Merkmale der Klassenführung und des Unterrichtsklimas keiner fachdidaktischen Kenntnisse vonseiten der Rater, um eine entsprechend hohe Messqualität in verschiedenen Fächern zu erzielen. Allerdings sollte die Brauchbarkeit des Ratingsystems in weiteren Studien und für andere Stichproben erneut überprüft werden. Beispielsweise können auf Basis der hier gewonnenen Ergebnisse, aufgrund der Konfundierung von Fach und Schuljahr, keine allgemeinen Schlussfolgerungen für die Fächer gezogen werden. Hier sollten sich Studien anschließen, die das Ratingsystem in den Fächern Deutsch und Mathematik zum gleichen Messzeitpunkt einsetzen. Beispielsweise wäre ein Studiendesign denkbar, in dem sowohl im ersten als auch im zweiten Schuljahr der Unterricht in beiden Fächern gefilmt und mithilfe des hoch inferenten Ratingsystems hinsichtlich der Klassenführung und des Unterrichtsklimas eingeschätzt wird. Einerseits ließe sich mit diesem Vorgehen der Konfundierung zwischen Fach und Schuljahr entgegenwirken, andererseits könnten sich in diesem Zusammenhang weitere Analysen zur Stabilität der beobachteten Unterrichtsqualitätsmerkmale in Abhängigkeit vom Unterrichtsfach anschließen.

## **17.3 Methodische Einschränkungen und weitere Forschungsfragen**

Entsprechend dem hoch inferenten Ansatz wurde in der vorliegenden Arbeit eine Doppelstunde Deutsch (1. Schuljahr) bzw. Mathematik (2. Schuljahr) als Indikator für das gesamte Unterrichtsgeschehen im jeweiligen Fach herangezogen. Zur Diskussion gestellt werden muss an dieser Stelle, dass es fraglich ist, ob die einmalige Erfassung der Klassenführung und des Unterrichtsklimas in einer Unterrichtsstunde ausreicht, um die Qualität des Unterrichts über ein gesamtes Schuljahr bzw. für ein spezielles Fach abzubilden. Häufig wird in diesem Zusammenhang argumentiert, dass es notwendig sei, eine größere

Anzahl an Stunden in den entsprechenden Fächern zu beobachten, um allgemeingültige Aussagen über den gesamten Zeitraum des Anfangsunterrichts machen zu können und um Effekte auf Schülermerkmale zu untersuchen.

Dass eine aufgezeichnete Unterrichtsstunde durchaus als hinreichender Schätzer angesehen werden kann, zeigt eine aktuelle Studie von Praetorius et al. (2013). Praetorius et al. (2013) bestätigen, dass bereits eine aufgezeichnete Stunde ausreicht, um die Basisdimensionen Klassenführung und Unterrichtsklima hinreichend zuverlässig zu erfassen – im Gegensatz zur kognitiven Aktivierung, für die mindestens neun Stunden notwendig sind. Seidel et al. (2002) zeigen ergänzend, dass vor allem die effektive Zeitnutzung eine hohe Konformität in den über ein halbes Jahr hinweg aufgezeichneten Stunden aufweist. Für die Grundschule liefern Pianta et al. (2008) Hinweise für eine gewisse Stabilität von Unterrichtsmerkmalen über ein Schuljahr.

Allerdings argumentieren Kritiker, dass eine solche einmalige Erfassung eher anfällig gegenüber Verzerrungen aller Art ist, welche unter anderem aus der Tagesform der Lehrperson und der Schüler entstehen können (vgl. Kleinbub, 2010). Zudem ist die Ausgestaltung der Unterrichtsqualität abhängig von situationalen Merkmalen (z. B. Sozialform), Rahmenbedingungen (z. B. Klassenzusammensetzung) oder den übergreifenden Orientierungen der Lehrperson (z. B. leistungsbezogene Einstellungen) (vgl. Roßbach, 2002b). Jedoch bestätigte Roßbach (2002b), dass sowohl Strukturmerkmale als auch die Orientierung der Lehrperson, in diesem Fall der Klassenlehrer, die tatsächlich zu beobachtende Unterrichtsqualität determinieren, d. h. „das gleiche Unterrichtsverhalten und die gleichen Interaktionen der Lehrerinnen mit den Schülern können unter verschiedenen Rahmenbedingungen beobachtet werden [...]“ (S. 240).

Im Zusammenhang mit der Abhängigkeit der Unterrichtsqualität von situationalen Merkmalen des Unterrichts stellt sich die Frage, ob vor allem die Operationalisierung der Klassenführung für den Anfangsunterricht überdacht werden sollte, denn obwohl mithilfe des Ratingsystems größtenteils reliable Daten für den aufgezeichneten Unterricht in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) erhoben werden konnten, blieb bei der Entwicklung teilweise unberücksichtigt, dass sich der Unterricht in der Grundschule durch häufigen Methoden- und Sozialformwechsel auszeichnet und sich somit auch die Rolle der Lehrperson und vor allem die Anforderungen an die Klassenführung je nach Sozialform bzw. gewählter Unterrichtsmethode ändern (vgl. Carter & Doyle, 2006). Carter und Doyle (2006) unterscheiden beispielsweise zwischen Klassenführung in lehrerzentrierten Phasen (Lesson), in Phasen der Diskussion (Discussion and Sharing Time), in Schülerarbeitsphasen am Platz (Seatwork) und während der Gruppenarbeit (Group Cooperation). Da die Operationalisierung in Anlehnung an Videostudien aus der Sekundarstufe erfolgte, wurden beispielsweise Merkmale der Klassenführung eher im Sinne der direkten Instruktion bzw. nach Carter und Doyle (2006) tendenziell für lehrerzentrierten Unterricht entwickelt. Vor

allem in offenen Arbeitsphasen, in denen Schüler alleine oder in Gruppen arbeiten, besitzt die Klassenführung jedoch eine andere Charakteristik als in Phasen, in denen lehrerzentrierter Unterricht stattfindet (vgl. Bohl & Kucharz, 2010). Im Vordergrund stehen eher präventive Maßnahmen, wie die Festlegung von Regeln während der Schülerarbeitsphasen oder routinierte Abläufe. Zudem liegt der Fokus hier vor allem auf der Organisation der Aufgaben und des Materials. Bohl und Kucharz (2010) diskutieren anhand der Merkmale von Kounin (vgl. Kapitel 3.1, Tabelle 5) fünf Möglichkeiten der Klassenführung im offenen Unterricht. Beispielsweise würde die Allgegenwärtigkeit einer Lehrperson in offenen Unterrichtsphasen sichtbar, wenn die Lehrperson eine Gruppe berät, während sie eine andere unruhige Gruppe beobachtet oder wenn man den Eindruck hat, dass die Lehrperson die Schüler und / oder Schülergruppen punktuell oder systematisch beobachtet (vgl. Bohl & Kucharz, 2010). Übergänge zwischen einzelnen Stationen würden von den Schülern selbst initiiert werden und müssten im Vorfeld geregelt werden. Hier spielt vor allem die Festlegung von Regeln und die Routinisierung von solchen Arbeitsabläufen eine besondere Rolle. Im Vergleich zu traditionellem Unterricht, der weniger anspruchsvolle Lehrmethoden einsetzt, müssen für Formen des selbstständigen oder kooperativen Lernens deutlich mehr Regeln eingeführt werden bzw. es wird ein flexibleres Regelsystem benötigt (zusf. Doyle, 1986). In diesen Lernformen wird ein wesentlicher Teil der Klassenführung von der Lehrperson auf die Schüler übertragen. Somit bedarf es nicht mehr der permanenten Steuerung durch die Lehrperson, sondern den Schülern wird ein hohes Maß an selbstbestimmtem Lernen sowie Entscheidungs- und Handlungsspielräumen gegeben. Für Knauf (2001) müssen in solchen Lernformen unter anderem Regeln existieren, die der Sicherung einer guten Arbeitsatmosphäre dienen (gedämpfte Geräuschkulisse, Vermeiden von hektischem Herumlaufen), Regeln für Partnerarbeiten (z. B. wechselseitige Hilfe oder Vermeidung von Störungen) oder Regeln beim Auf- und Abbau von Stuhlkreisen).

Angesichts der Ausdifferenzierung der Lernsituationen ist die Einschätzung einer Unterrichtsstörung sowohl für die Lehrperson als auch für den Rater in solchen Phasen vermutlich viel schwieriger (vgl. Bohl & Kucharz, 2010). Während kleinste Störungen in lehrerzentrierten Arbeitsphasen von der Lehrperson bereits als störend empfunden werden können, ist ein gewisser Lautstärkepegel in offenen Unterrichtsphasen nicht ungewöhnlich. Die Frage „Ab wann soll die Lehrperson eingreifen?“ wird nochmals ausgeweitet, wenn es darum geht, einen Teil der Klassenführung in offenen Unterrichtsphasen an die Schüler abzugeben. Dies wurde jedoch bei der Erstellung des Ratingsystems nicht berücksichtigt. Zwar mussten die Rater ein Gesamturteil über die Stunden abgeben und man kann annehmen, dass massive Störungen in Schülerarbeitsphasen, die von der Lehrperson nicht unterbunden wurden, negativ in das Urteil der Rater einfließen, jedoch lag der Fokus nicht auf einer differenziellen Erfassung der Klassenführung in lehrer-

zentrierten bzw. schülerzentrierten Phasen. Um die intersituationale Stabilität der Klassenführung über verschiedene Sozialformen hinweg überprüfen zu können, sollten sich weitere Analysen anschließen. Als Analyseeinheit sollte dann nicht die gesamte Unterrichtsstunde dienen, sondern die im Vorfeld im Rahmen der Basiskodierung identifizierten Phasen des öffentlichen Unterrichts und der Schülerarbeitsphasen (Einzel-, Partner- und Gruppenarbeit). Dieses Vorgehen würde es zusätzlich erlauben, die Brauchbarkeit des hoch inferenten Ratingsystems für offenere Unterrichtsformen zu überprüfen. Da der Unterricht durch Sozialformen auch methodisch-organisatorisch strukturiert wird, könnten sich Analysen zur Stabilität der Klassenführung anschließen.

#### *Einbezug weiterer Datenquellen*

Im Rahmen der vorliegenden Arbeit wurden die Merkmale der Klassenführung und des Unterrichtsklimas lediglich aus der Perspektive der Beobachter bzw. durch die Analyse von Unterrichtsvideos erhoben. Die Verwendung weiterer Informationsquellen (Lehrpersonen und Schüler) ermöglicht jedoch eine Integration verschiedener Perspektiven auf den Unterricht. So kann die Qualität des Unterrichts zusätzlich aus dem Blickwinkel der Lehrpersonen und Schüler erfasst werden, indem Fragebögen eingesetzt oder Interviews durchgeführt werden.<sup>97</sup> Subjektive Angaben durch die Schüler können ergänzende Einblicke geben (vgl. Helmke et al., 1986). Denn beispielsweise reagieren nicht alle Schüler gleich auf das bestehende Klima einer Klasse, sondern interpretieren klimapositive und -negative Situationen unterschiedlich. So werden vermutlich einige Schüler auch im Anfangsunterricht bereits über Strategien verfügen, die einen Mangel an positiven Lehrer-Schüler-Beziehungen ausgleichen, während andere Schüler mit einer Reduzierung des eigenen Lernens reagieren bzw. negative Selbstwertgefühle entwickeln. Ein und dasselbe Unterrichtsklima in einer Klasse kann demnach von den Schülern unterschiedlich wahrgenommen werden und somit unterschiedlich auf die einzelnen Schüler wirken. Für die Erfassung des Unterrichtsklimas aus Schülersicht besteht in den ersten beiden Schuljahren jedoch die Herausforderung vor allem darin, standardisierte Fragebögen zu entwickeln, die punktuell das Unterrichtsklima in der Klasse erfassen und in denen die Beantwortung der Fragen wenig Abstraktion von den noch jungen Schülern verlangt. Vor allem Schüler in den Anfangsjahren ihrer Schulzeit können lediglich generell zutreffende Aussagen treffen, die sich auf die augenblickliche Situation oder auf die nahe Vergangenheit beziehen (z. B. Lipski, 2000). Eine auf einen längeren Zeitraum rückblickende Beurteilung bereitet eher Schwierigkeiten, da dies von den jungen Schülern die Abstraktion vom konkret Erlebten auf eine generalisierende Einschätzung verlangt.

---

97 Die Eignung von Lehrer- und Schüleraussagen als Datenquelle zur Erfassung der Unterrichtsqualität wird vielfach diskutiert (z.B. Clausen, 2002; Helmke, 2009; Praetorius, 2014). Auf die Vor- und Nachteile der Erfassung der Unterrichtsqualität aus Sicht dieser beiden Personengruppen wird an dieser Stelle nicht weiter eingegangen.

## 18 Teilstudien (Studie 1 bis 3)

Im folgenden Teil sollen die Ergebnisse der drei Teilstudien zusammengefasst und diskutiert werden, beginnend mit der Studie 1, deren Ziel es war, die Ausprägungen der Unterrichtsmerkmale für die vorliegende Stichprobe zu beschreiben. Dem schließt sich die Zusammenfassung und Diskussion der Ergebnisse der faktoranalytischen Prüfung (Studie 2, vgl. Kapitel 18.2) und der multikriterialen Effekte (Studie 3, vgl. Kapitel 18.3) an.

### 18.1 Ausprägung der Merkmale (Studie 1)

Da es bis zum jetzigen Zeitpunkt nur wenige videobasierte Studien in Deutschland gibt, anhand derer Informationen über die tatsächliche Ausgestaltung der Unterrichtsqualität im Anfangsunterricht abgeleitet werden können, ging es in der Studie 1 ausschließlich um eine Beschreibung auf Basis der einzelnen hoch inferenten Merkmale. Dementsprechend hatte diese Teilstudie 1 eher einen heuristischen bzw. explorativen Charakter.

Die deskriptiven Befunde der hoch inferenten Ratings bestätigen, dass die Merkmale der Klassenführung sowohl im Fach Deutsch (1. Schuljahr) als auch im Fach Mathematik (2. Schuljahr) im Mittel von den zwei Ratern deutlich positiv eingeschätzt werden. Für den Großteil der an PERLE teilnehmenden Lehrpersonen können für die Aspekte der Klassenführung Mittelwerte um „3“ und höher festgestellt werden. Demnach kann für die vorliegende Stichprobe gezeigt werden, dass es den teilnehmenden Lehrpersonen eher leichtfällt, ihre Klasse effektiv zu führen. Da aktuelle Modelle zur Unterrichtsqualität (vgl. Pietsch-Modell, Abbildung 4) zeigen, dass unter anderem eine effektive Klassenführung als eine Grundvoraussetzung bzw. als eine Art Basis guten Unterrichts gilt, damit beispielsweise kognitiv aktivierende Instruktionsprozesse stattfinden können (vgl. Kapitel 1.2.1), ist dieses Ergebnis positiv zu bewerten. Verschiedene Beobachtungsstudien konnten bereits ähnlich hohe Ausprägungen der Klassenführung für den Anfangsunterricht bestätigen (vgl. Kapitel 2.3).

Es bleibt an dieser Stelle zu diskutieren, ob die mangelnde Varianz in der Ausprägung der Klassenführung in den teilnehmenden Klassen eventuell durch den Einsatz von Unterrichtsvideografie begründet – die Lehrpersonen versuchen in diesem Fall, bestmöglichen Unterricht zu präsentieren – oder ob die mangelnde Variabilität auf das hoch inferente Ratingsystem zurückgeführt werden kann. Hier stellt sich vor allem die Frage, inwieweit das entsprechende Antwortformat nicht genügend Varianz erzeugt hat und somit die mögliche vorhandene Variabilität in der Effektivität der Klassenführung

nicht ausreichend abgebildet werden konnte. Zusätzlich sollte überprüft werden, ob nicht auch die inhaltlichen Vorgaben an die Lehrpersonen und die daraus resultierenden ähnlichen Inszenierungsmuster bzw. Unterrichtsverläufe (vgl. Kapitel 9.3) grundsätzlich zu keiner großen Variabilität in den Ausprägungen der Klassenführungsmerkmale geführt haben. Die Vorgaben stellten für die Lehrpersonen hinsichtlich der Klassenführung scheinbar keine große Herausforderung dar. Dies gilt für beide Fächer gleichermaßen.

Für die Merkmale des Unterrichtsklimas zeigte sich kein so einheitliches Bild wie bei der Klassenführung. Während die Merkmale der Wertschätzung („Anerkennung der Schüler durch die Lehrperson (Anes)“ und „gegenseitige Anerkennung der Schüler (Aneg)“) im Mittel ähnlich positiv von den beiden Ratern eingeschätzt wurden wie die Merkmale der Klassenführung (Mittelwert liegt zwischen 3.5 und 4), liegen die Mittelwerte der restlichen Merkmale zwischen 2 und 3. Diese Werte könnten ein Hinweis darauf sein, dass es den teilnehmenden Lehrpersonen im Mittel etwas „schwerer“ fällt, ein positives Unterrichtsklima zu schaffen, als ihre Klasse effektiv zu führen. Da sowohl die Aspekte der Klassenführung als auch des Unterrichtsklimas – mit Ausnahme der Merkmale „humorvolle Lernatmosphäre (Humo)“ und „Einsatz von Lob (Lob)“ im Fach Deutsch – mit Werten über der theoretischen Mitte von 2.5 eingeschätzt wurden, kann die Hypothese 1 an dieser Stelle bestätigt werden: Für das erste (Fach Deutsch) und das zweite Schuljahr (Fach Mathematik) schätzten die Rater die Aspekte der Klassenführung und des Unterrichtsklimas im Mittel deutlich positiv ein.

Die niedrigen Werte für das Merkmal „humorvolle Lernatmosphäre (Humo)“ bedeuten, dass sowohl in Deutsch (1. Schuljahr) als auch in Mathematik (2. Schuljahr) in den Klassen eine eher trockene Unterrichtsatmosphäre herrscht, in der die Lehrpersonen von sich aus eher wenig Humor einsetzen. Vermutlich sind hier indirekt Hinweise zu implizitem Wissen der Lehrpersonen über Humortheorien ersichtlich. So kann – wie bereits in Kapitel 4.2 dargestellt – vermutet werden, dass Lehrpersonen bewusst wenig Humor im Unterricht einsetzen, da sie implizit wissen, dass zum Verstehen von Humor ein komplexer kognitiver Prozess notwendig ist (vgl. Raaf, 2005) bzw. dass das Humorverständnis vom kognitiven Entwicklungsstand oder dem Stand der Sprachentwicklung der Schüler abhängig ist.

Um zu überprüfen, ob sich die mittlere Ausprägung auf den Merkmalen zwischen den Fächern unterscheidet, wurden in einem zweiten Analyseschritt im Rahmen der Studie 1 univariate Varianzanalysen auf Grundlage der Daten derjenigen PERLE-Lehrpersonen berechnet, die gleichzeitig an beiden Videostudien teilgenommen haben und für die entsprechende Einschätzungen der Rater vorlagen. Es zeigte sich in der Höhe der Einschätzungen der Merkmale durch die zwei Rater auf Basis der kleineren Stichprobe von 30 Lehrpersonen – mit Ausnahme des Merkmals „Einsatz von Lob (Lob)“ auf einem 10 %-Niveau – kein signifikanter Unterschied in der Höhe der Ausprägungen der einzelnen Ratings zwischen den zwei Fächern. Zusammenfassend kann somit

festgehalten werden, dass sowohl für die Merkmale der Klassenführung als auch für die des Unterrichtsklimas die Mittelwertsunterschiede in der Ausprägung zwischen den beiden Videostudien praktisch nicht bedeutsam sind. Dies weist auf eine gewisse Stabilität in der Höhe der Ausprägungen zwischen den Fächern / Schuljahren bei gleicher Lehrperson und Klasse hin.

Der nicht gefundene Unterschied in der Höhe der Mittelwerte lässt sich möglicherweise auch darauf zurückführen, dass es sich bei den Einschätzungen des Unterrichts der 30 Lehrpersonen um abhängige Daten handelt. Die aufgezeichneten Unterrichtsstunden in den Fächern Deutsch und Mathematik sind geschachtelt in der jeweiligen Lehrperson, was die Höhe der Mittelwertsunterschiede beeinflussen kann, denn die Ausprägungen der Merkmale der Klassenführung und des Unterrichtsklimas fallen ähnlicher aus, wenn es sich um ein und dieselbe Lehrperson handelt, als wenn man unterschiedliche Lehrpersonen miteinander vergleicht. Das gleiche gilt vermutlich auch für die Ausprägung der Merkmale, die gezielt das Schülerverhalten erfassen (z. B. „gegenseitige Anerkennung der Schüler (Aneg)“). Hier empfiehlt sich zusätzlich die Berechnung einer Varianzanalyse mit Messwiederholung, in der die Lehrperson als Messwiederholungsfaktor definiert wird. Zudem sollten sich weitere Analysen zur Stabilität der Unterrichtsqualitätsmerkmale anschließen. Beispielsweise könnten durch eine zusätzliche Berechnung korrelativer Zusammenhänge zwischen den Einschätzungen in Deutsch und Mathematik Aussagen über die normative Stabilität der einzelnen Merkmale der Klassenführung und des Unterrichtsklimas getroffen werden. Dadurch lässt sich unter anderem die Frage beantworten, ob sich die Rangordnung der Lehrpersonen zwischen den zwei Messzeitpunkten ändert ( $r \neq 0$ ) oder ob sie gleich bleibt ( $r = 0$ ).

Wie bereits erwähnt, ließ sich lediglich für das Merkmal „Einsatz von Lob (Lob)“ tendenziell ein signifikanter Unterschied zugunsten des Fachs Mathematik (2. Schuljahr) aufzeigen, d. h., die Rater kommen übereinstimmend zu dem Urteil, dass im Mathematikunterricht des zweiten Schuljahres „häufiger“ gelobt wird als im Deutschunterricht des ersten Schuljahres. Aufgrund der eher geringen Effektstärke von .051 kann dieser Unterschied jedoch als weniger bedeutsam eingestuft werden. Mit Mittelwerten knapp über 2 wird der „Einsatz von Lob (Lob)“ von den zwei Ratern in beiden Fächern als weniger häufig wahrgenommen. Damit können bisherige Befunde bestätigt werden (zusef. Brophy, 1981; vgl. Kapitel 4.2). In weiteren Analysen sollte untersucht werden, ob die Häufigkeit von Lob je nach Unterrichtsmethode variiert. Hier könnten sich niedrig inferente Kodierungen anschließen, wie sie beispielsweise im Kontext der WOLFF-Studie durchgeführt wurden (vgl. Kucharz, Schnebel & Helming, 2011). In der Studie konnte gezeigt werden, dass Lehrpersonen vor allem im lehrerzentrierten Unterricht besonders viel loben, womit erneut die Situationsspezifität einzelner Unterrichtsqualitätsmerkmale deutlich wird. Die Autorinnen begründen dieses Ergebnis damit, dass Lehrpersonen in solchen Phasen „zum Ausgleich der Instrukionalität und Zielori-

entierung ihres Unterrichtsstils besonderen Wert auf emotionale Bestätigung legen“ (Kucharz et al., 2011, S. 106).

An dieser Stelle bleibt grundsätzlich zu diskutieren, ob die theoretische Annahme, dass die „4“ auf der Ratingskala als die höchste Ausprägung des jeweiligen Merkmals gilt, überhaupt wünschenswert ist. Beispielsweise hat ein „Mehr“ an Lob keine positiveren Auswirkungen auf den Leistungsfortschritt von Schülern (vgl. Brophy, 1981). Vor allem in der Grundschule sollte Lob nicht in übertriebenen Maßen eingesetzt werden (vgl. Anderson et al., 1979). Einsiedler (1989) weist zudem darauf hin, dass Schüler aus häufigem Lob auch schlussfolgern können, die Lehrperson „wolle damit die Anstrengungsbereitschaft hervorheben, die das zu geringe Fähigkeitsniveau ausgleicht“ (S. 106). Auch in Bezug auf die „Fürsorglichkeit, Herzlichkeit und Wärme (Fuer)“ sollte diskutiert werden, wie viel an „Fürsorglichkeit, Herzlichkeit und Wärme“ für den Anfangsunterricht „angemessen“ ist. In beiden Fächern schätzten die Rater dieses Merkmal mit Werten zwischen 2.5 und 3.0 ein (vgl. Kapitel 21), demnach sind die Lehrpersonen der PERLE-Stichprobe sowohl im Fach Deutsch (1. Schuljahr) als auch im Fach Mathematik (2. Schuljahr) überwiegend freundlich und interessiert an den persönlichen (nicht fachlichen) und außerschulischen Belangen ihrer Schüler. Die Lehrpersonen nehmen sich demnach auch im Unterricht viel Zeit für die persönlichen Probleme ihrer Schüler. Im Sinne einer effektiven Zeitnutzung für unterrichtsrelevante Inhalte sollte dies jedoch angemessen erfolgen. So ist die höchstmögliche Ausprägung auf einem Merkmal nicht zwangsläufig auch die bestmögliche, vielmehr sollte in weiteren Studien das „vage Maß der Angemessenheit“ (Kleinbub, 2010, S. 287) berücksichtigt werden.

#### *Weiterführende Forschungsfragen*

Inwieweit sich der oben dargestellte Ergebnistrend auch für das Fach Kunst (1. Schuljahr) bestätigen lässt, bedarf weiterer Untersuchungen. Erste deskriptive Analysen für das Fach Kunst (1. Schuljahr) zeigen jedoch ähnlich hohe Werte für die Aspekte der Klassenführung und des Unterrichtsklimas wie in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr). Beispielsweise schätzen die Rater auch im Fach Kunst die Lernatmosphäre als eher trocken ein.

Obwohl sich auf Mittelwertebene keine Unterschiede in der Höhe der Ausprägungen der einzelnen Merkmale zwischen den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) zeigen ließen – mit Ausnahme des Merkmals „Einsatz von Lob (Lob)“ –, kann dennoch angenommen werden, dass es einzelne Klassen gibt, die durchgehend hohe Werte für die Merkmale der Klassenführung und des Unterrichtsklimas haben und andere, in denen durchgehend eher niedrigere Werte zu finden sind. Es stellt sich in diesem Zusammenhang die Frage, ob sich in weiteren Analysen verschiedene Profile von Klassen bestätigen lassen. So konnten beispielsweise Hamre und Pian-  
ta (2007) mithilfe des CLASS-Instruments (vgl. Kapitel 2.1) Kernprofile der

Unterrichtsqualität empirisch nachweisen. Ein Teil der in ihrer Studie untersuchten Klassen zeichneten sich durchgehend durch hohe Werte auf den drei Grunddimensionen aus (Emotional Support, Classroom Organization und Instructional Support), während ein ähnlich großer Anteil an Klassen identifiziert werden konnte, die für alle drei Domänen niedrige Werte hatten. Für das erste Schuljahr belegen Stuhlman und Pianta (2009) ähnliche Profile. Inwieweit sich auch auf Basis der PERLE-Daten Kernprofile identifizieren lassen, sollte in weiteren Analysen untersucht werden.

Zusätzlich könnte untersucht werden, ob sich die Klassen der BIP-Schulen und die der staatlichen Grundschulen in der Höhe der Ausprägung der einzelnen Merkmale unterscheiden. In den bisherigen Analysen blieb bislang unberücksichtigt, dass sich die Stichprobe aus zwei verschiedenen Schultypen (privat versus staatlich) zusammensetzt, die sich hinsichtlich der Ausprägung der einzelnen Merkmale unterscheiden können. Allein die Tatsache, dass es sich bei den BIP-Kreativitätsgrundschulen um Ganztagschulen handelt, an denen in den Fächern Deutsch und Mathematik zwei Lehrpersonen die Klasse unterrichten, hat vermutlich Auswirkungen auf Merkmale der Unterrichtsqualität.

Offen bleibt an dieser Stelle, welche Rolle Lehrerkognitionen und -überzeugungen (bzw. Beliefs) für das unterrichtliche Handeln und somit für die Ausgestaltung der Unterrichtsqualität haben bzw. welche handlungssteuernden Wirkungen diesen in der Umsetzung einer effektiven Klassenführung oder eines positiven Unterrichtsklimas zukommen. Dabei stellt sich beispielsweise die Frage, ob Unterricht von konstruktivistisch orientierten Lehrpersonen weniger strukturiert ist. Im Rahmen der COACTIV-Studie erwies sich die Effizienz der Klassenführung (aus Schülersicht) als unabhängig von den Überzeugungen der Lehrpersonen in der Sekundarstufe. Die Autoren werten dies als einen Hinweis darauf, dass „Unterricht von konstruktivistisch orientierten Lehrkräften nicht unbedingt weniger strukturiert ist“ (Voss, Kleickmann, Kunter & Hachfeld, 2011, S. 250). Konstruktivistische Überzeugungen gehen demnach nicht zwangsweise mit einer Laissez-faire-Haltung einher (vgl. ebd.). Für die Grundschule kommen Hartinger, Kleickmann und Hawelka (2006) zu einem ähnlichen Ergebnis. In ihrer Studie gaben konstruktivistisch orientierte Lehrpersonen mehr Freiräume, wobei der Unterricht nicht weniger strukturiert abläuft. Gleichzeitig fühlen sich die Schüler in einer solchen Lernumgebung selbstbestimmter und sind interessierter. Im Rahmen der PERLE-Studie liegen entsprechende Daten zu konstruktivistischen Lehrerüberzeugungen im Anfangsunterricht vor, womit sich weitere Analysen in diese Richtung anbieten.

## 18.2 Überprüfung der Konstruktvalidität (Studie 2)

Die einzelnen hoch inferenten Merkmale der Klassenführung und des Unterrichtsklimas konnten mithilfe mehrerer explorativer und konfirmatorischer Faktorenanalysen übergeordneten Faktoren zugeordnet werden. Die Analyse erfolgte in zwei Schritten, deren Ergebnisse an dieser Stelle noch einmal kurz zusammengefasst und anschließend diskutiert werden. In einem Vergleich mehrerer Modelle (Generalfaktormodell, 2-Faktoren-Modell, 3-Faktoren-Modell, vgl. Kapitel 15.1.3) konnte in einem ersten Schritt, ausgehend von der Gesamtstichprobe (Deutsch- und Mathematikvideos), kein allgemeines Modell für den Anfangsunterricht identifiziert werden, das im Sinne der Modellgüte perfekte Fit-Indizes aufwies. Die „beste“ Modellgüte erreichte in diesem Schritt das 2-Faktoren-Modell mit zusätzlichen Residualkorrelationen zwischen den „Feedbackmerkmalen“ (vgl. Modell 2b) und passt demnach besser auf die zugrunde liegenden hoch inferenten Daten als das Generalfaktormodell, womit Hypothese 2 grundsätzlich bestätigt werden kann. In diesem Modell 2b wurde bereits das Merkmal „gegenseitige Anerkennung der Schüler (Aneg)“ als Aspekt der Klassenführung modelliert, da sich im Vorfeld zeigte, dass diese eher mit den Merkmalen der Klassenführung korreliert als mit denen des Unterrichtsklimas. Dieses Ergebnis wurde an entsprechender Stelle bereits diskutiert (vgl. Kapitel 15.1). Obwohl die zusätzlichen Residualkorrelationen zwischen den „Feedbackmerkmalen“ auf einen latenten dritten Faktor hinweisen, erzielte das 3-Faktoren-Modell keine deutlich bessere Modellgüte. Dies liegt vermutlich an dem Merkmal „Einsatz von Lob (Lob)“, das sich im Verlauf der fachspezifischen EFA für beide Fächer als problematisch herausstellte und keinem der Faktoren zugeordnet werden konnte (vgl. Kapitel 15.2).

Ein möglicher Grund dafür, dass im Rahmen dieses ersten Modellvergleichs kein allgemeingültiges Modell für den Anfangsunterricht identifiziert werden konnte, resultiert vor allem aus der Stichprobenzusammensetzung. Da die konfirmatorischen Faktorenanalysen (vgl. Kapitel 15.1) auf Basis der Gesamtstichprobe, d. h., auf Basis aller Deutsch- und Mathematikvideos des ersten und zweiten Schuljahres berechnet wurden, kann vermutet werden, dass sich die Modellgüte verbessert, wenn fachspezifische konfirmatorische Faktorenanalysen durchgeführt werden. Da dies in der vorliegenden Arbeit aufgrund der zu geringen Stichprobe für die jeweiligen Fächer nicht möglich war, erfolgte in einem zweiten Schritt lediglich die Berechnung fachspezifischer explorativer Faktorenanalysen, deren Ergebnisse anschließend diskutiert werden.

Unberücksichtigt blieb zudem, dass die Fälle strenggenommen geschachtelt sind, d. h. für 30 Lehrpersonen liegen jeweils Werte für Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) vor (siehe auch Studie 1). Aufgrund der geringen Stichprobengröße konnte jedoch nicht geprüft werden, wie die Mo-

dellprüfungen ausfallen, wenn man von jenen Lehrpersonen, die beide Fächer unterrichten, nur jeweils einen Wert heranzieht. Sinnvoll wäre es zudem gewesen, die Faktorenstruktur getrennt für die Lehrpersonen zu prüfen, welche entweder nur Deutsch bzw. Mathematik oder aber beide Fächer unterrichten. Hier sollten sich weitere Analysen anschließen.

#### *Diskussion der Ergebnisse aus den fachspezifischen EFA*

Die Ergebnisse der fachspezifischen explorativen Faktorenanalysen sprechen insgesamt für eine gute Konstruktvalidität der hoch inferent erfassten Merkmale für beide Fächer. Dabei zeigte sich, dass sich die ursprünglich als allgemeine Unterrichtsmerkmale operationalisierten Basisdimensionen Klassenführung und Unterrichtsklima in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) unterschiedlich gestalten bzw. vermutlich unterschiedliche Bedeutung haben. Je nach Fach konnten Unterschiede in der Faktorenstruktur bestätigt werden. Im Fach Deutsch (1. Schuljahr) ergaben sich drei Faktoren (Klassenführung, Unterrichtsklima und Feedback). Der Faktor Klassenführung setzt sich aus allen, im Vorfeld theoretisch angenommen und reliabel eingeschätzten hoch inferenten Merkmalen der Klassenführung sowie aus dem Klimamerkmal „gegenseitige Anerkennung der Schüler (Aneg)“ zusammen. Auf dem Faktor Unterrichtsklima laden lediglich die Merkmale „humorvolle Lernatmosphäre (Humo)“, „Anerkennung der Schüler durch die Lehrperson (Anes)“ und „Fürsorglichkeit, Herzlichkeit und Wärme (Fuer)“. Der Feedback-Faktor besteht nur aus den zwei Merkmalen „Ermutigung durch die Lehrperson (Emut)“ und „sachlich-konstruktive Rückmeldung (Rück)“.

Für das Fach Mathematik (2. Schuljahr) ließen sich hingegen vier Faktoren ermitteln, wobei die zwei Faktoren Unterrichtsklima und Feedback mit denen in der Videostudie Deutsch (1. Schuljahr) tendenziell vergleichbar sind. Es zeigen sich nur leichte Unterschiede in den einzelnen Faktorladungen und Reliabilitäten. Allerdings sollten sich hier noch Analysen zur Messinvarianz der erhobenen Konstrukte über die zwei Messzeitpunkte anschließen, um zu überprüfen, ob die einzelnen Messungen über die Zeit hinweg wirklich vergleichbar sind (vgl. Geiser, 2010). Im Vergleich zur Videostudie Deutsch (1. Schuljahr) bilden die Merkmale der Klassenführung im Fach Mathematik zwei separate Faktoren: (1) effizientes Zeitmanagement und (2) effektiver Umgang mit Störungen. Der Faktor „effizientes Zeitmanagement“ besteht aus den Merkmalen „Zeitmanagement / effektive Zeitnutzung (Zeit)“, „Übergangsmanagement (Managing Transition) (Ueb)“, „Gruppenfokus (Grup)“ und „Allgegenwärtigkeit der Lehrperson (Allg)“. Auf dem Faktor „effektiver Umgang mit Störungen“ laden die Merkmale „effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)“, „Störungsfreiheit (Disz)“, „gegenseitige Anerkennung der Schüler (Aneg)“ und „Regelklarheit bzw. -verwendung (Regl)“.

Es stellt sich die Frage, ob die Unterschiede in der Struktur auch dann bestehen bleiben, wenn für die explorativen Faktorenanalysen nur diejenigen Fälle berücksichtigt werden, für die sowohl Daten aus der Videostudie Deutsch (1. Schuljahr) und der Videostudie Mathematik (2. Schuljahr) vorliegen. Sollte sich im Rahmen dieser weiteren Analysen zeigen, dass die Unterschiede in der Struktur (vor allem für die Merkmale der Klassenführung) bestehen bleiben, kann geschlussfolgert werden, dass es sich um fachspezifische strukturelle Unterschiede handelt.

#### *Methodische Einschränkungen im Rahmen der Studie 2*

Bezüglich der fachspezifischen explorativen aber auch der fachunspezifischen konfirmatorischen Faktorenanalysen ist anzumerken, dass es sich in beiden Videostudien um sehr kleine Stichproben von 50 bzw. 51 Fällen handelt. Es stand demnach keine hinreichend große Stichprobe als Datengrundlage zur Verfügung. Diese ist jedoch eine der wichtigsten Voraussetzungen für die stabile Schätzung der Ladungsmuster (vgl. Moosbrugger & Hartig, 2002, S. 143). Vor diesem Hintergrund sollten die Ergebnisse der fachspezifischen Faktorenanalysen nicht generalisiert interpretiert werden bzw. die Interpretation der faktoriellen Struktur sollte mit Zurückhaltung erfolgen. Es können anhand der fachspezifischen Faktorenanalysen keine Aussagen darüber gemacht werden, ob die auf Basis der EFA resultierenden Faktorenstrukturen (VSS = 3-Faktoren-Modell, VSM = 4-Faktoren-Modell) die empirischen Daten in dem jeweiligen Fach am besten abbilden (keine Angaben über Modellgüte möglich). Dies sollte auf Basis einer größeren Stichprobe von Unterrichtsvideos im Kontext konfirmatorischer Faktorenanalysen noch einmal überprüft werden. Unter der Bedingung, dass bei allen für die Faktorenanalyse verwendeten Merkmalen die Kommunalität<sup>98</sup> mindestens .60 beträgt, sind Stichproben von mindestens  $N = 60$  bereits ausreichend (vgl. MacCallum, Widaman, Zahng & Hong, 1999).

#### *Weiterführende Forschungsfragen*

Im Anschluss an die Ergebnisse der Studie 2 sollte zusätzlich analysiert werden, ob sich das Konzept der drei Basisdimensionen sowohl für das Fach Deutsch (1. Schuljahr) als auch für das Fach Mathematik (2. Schuljahr) bestätigen lässt. Hierfür sollte sich die videobasierte Erfassung der kognitiven Aktivierung in den beiden Fächern anschließen. Für das Fach Mathematik liegen bereits Daten aus einem hoch inferenten Rating zur kognitiven Aktivierung vor. So wurde beispielsweise im Rahmen dieses hoch inferenten Ratings der kognitiv-herausfordernde Umgang der Lehrperson mit Schülerbeiträgen, der Einsatz kognitiv aktivierender Aufgaben oder die Unterstützung kognitiver Selbstständigkeit von zwei Ratern reliabel eingeschätzt (vgl. Lau-

---

98 Kommunalität ist der Anteil der Varianz, den jede Variable (oder Dimension) mit anderen Variablen gemeinsam hat. Sie gibt an, in welchem Ausmaß eine Variable durch den entsprechenden Faktor aufgeklärt bzw. erfasst wird.

terbach, Gabriel & Lipowsky, 2013). Für das Fach Deutsch (1. Schuljahr) steht die hoch inferente Erfassung der kognitiven Aktivierung noch aus. Im Kontext der Dissertation von Miriam Lotz (2013d) wird derzeit das kognitive Aktivierungspotenzial der Leseübung mithilfe verschiedener niedrig und hoch inferenter Methoden analysiert.

Zudem sollte auf Basis einer größeren Stichprobe untersucht werden, ob sich die fachspezifischen Unterschiede in der Struktur der Klassenführung (Deutsch: Gesamtskala; Mathematik: zwei Teilskalen) auch im Rahmen von konfirmatorischen Faktorenanalysen bestätigen lassen. In diesem Zusammenhang können sich erneut Analysen für das Fach Kunst (1. Schuljahr) anschließen. Ergibt sich für das Fach Kunst (1. Schuljahr) beispielsweise eine ähnliche Struktur wie für das Fach Deutsch (1. Schuljahr), könnten die in der vorliegenden Arbeit bestätigten Unterschiede in der Struktur zwischen den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) eher mit dem Schuljahr bzw. dem Messzeitpunkt zusammenhängen, was erneut auf eine Konfundierung des Schuljahres mit dem entsprechenden Fach hinweisen würde. Somit wären die bestätigten Unterschiede in der Struktur der Unterrichtsqualitätsmerkmale kein fachspezifisches Phänomen, sondern abhängig davon, ob die Merkmale im ersten oder zweiten Schuljahr erfasst werden.

### **18.3 Multikriteriale Effekte der Klassenführung und des Unterrichtsklimas (Studie 3)**

Ziel der Studie 3 war es, den Stellenwert einer effektiven Klassenführung und eines positiven Unterrichtsklimas für die Entwicklung kognitiver und nicht-kognitiver Schülermerkmale im Anfangsunterricht in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) zu untersuchen. Zur Überprüfung wurden verschiedene Mehrebenenregressionsmodelle in *MPlus* gerechnet. Dabei konnten entgegengesetzte Wirkungen der beiden Unterrichtsqualitätsmerkmale für die Leistungs- und Selbstkonzeptentwicklung aufgezeigt werden. Die Ergebnisse werden nun für beide Fächer getrennt zusammengefasst und diskutiert.

#### *Ergebnisse zum Zusammenhang der Unterrichtsqualitätsmerkmale*

Obwohl im Vorfeld theoretisch ein Zusammenhang zwischen einer effektiven Klassenführung und einem positiven Unterrichtsklima angenommen wurde (vgl. Kapitel 1.2.2), zeigt sich dieser nur im Fach Deutsch (1. Schuljahr). Im Fach Mathematik (2. Schuljahr) konnte im Kontext der PERLE-Studie kein signifikanter Zusammenhang bestätigt werden. Hypothese 3 wird demnach teilweise verworfen. Im Fach Deutsch (1. Schuljahr) schätzen die Rater Klassen mit einem positiven Unterrichtsklima tendenziell auch positiv bzw. ef-

fektiv in Bezug auf die Klassenführung ein. Vermutlich wird hier erneut die Konfundierung von Fach und Messzeitpunkt ersichtlich. Zur Erinnerung: Die Videostudie Deutsch fand in der ersten Hälfte des ersten Schuljahres statt, während die Videos im Fach Mathematik ein Schuljahr später aufgezeichnet wurden. Da vor allem die ersten Monate des ersten Schuljahres allgemein als eine Art Eingewöhnungsphase für die Kinder und die Lehrperson zu sehen sind, kann vermutet werden, dass die Lehrpersonen zum Ausgleich einer effektiven Klassenführung bzw. der Einführung von Regeln etc. besonderen Wert auf die Schaffung einer angenehmen Lernatmosphäre legen, die es den Kindern erleichtert, sich in der neuen Rolle als Schulkind und den damit einhergehenden Verpflichtungen und Regeln wohlzufühlen.

#### *Bedeutung der Klassenführung für multikriteriale Schülermerkmale*

Die Ergebnisse zur Bedeutung der effektiven Klassenführung für die Leistungen und Selbstkonzepte der Schüler in Abhängigkeit vom untersuchten Unterrichtsfach werden anschließend diskutiert. Wie sich zeigte, können sowohl für den Bereich Rechtschreiben (1. Schuljahr) als auch im Fach Mathematik (2. Schuljahr) keine signifikanten Effekte der Klassenführung auf die Leistungen der Schüler bestätigt werden, nach Kontrolle vorheriger individueller Vorläuferfähigkeiten bzw. Leistungen im entsprechenden Fach und der gleichzeitigen Modellierung des Unterrichtsklimas auf Level 2. In Mathematik gilt dies für beide Teilskalen der Klassenführung gleichermaßen. Im Vorfeld wurde jedoch davon ausgegangen, dass eine effektive Klassenführung auch im Anfangsunterricht und unabhängig vom Fach einen positiven Einfluss auf die Leistungen der Schüler hat (vgl. Hypothesen 4 und 5). Da sich dies im Rahmen der vorliegenden Arbeit nicht zeigen ließ, müssen die Hypothesen 4 und 5 verworfen werden. Dies wirft die Frage auf, ob für die Leistungsentwicklung im Anfangsunterricht eine effektive Klassenführung grundsätzlich noch nicht so bedeutsam ist, wie beispielsweise in den darauffolgenden Schuljahren, für die positive Zusammenhänge bestätigt werden können (vgl. Kapitel 5.1.1).

Während in der vorliegenden Studie eine effektive Klassenführung keine Bedeutung für die Leistungsentwicklung im Rechtschreiben bzw. in Mathematik hat, zeigte sich für beide Fächer / Domänen tendenziell ein negativer Einfluss einer effektiven Klassenführung auf das *Selbstkonzept* der Schüler, nach Kontrolle früherer Selbstkonzepte auf Individualebene und der gleichzeitigen Modellierung des Unterrichtsklimas auf Level 2 (vgl. Modelle 1b, 2c, 2d). Der negative Effekt bleibt auch dann bestehen, wenn das Unterrichtsklima nicht gleichzeitig auf Level 2 modelliert wird. Die inhaltliche Interpretation fällt an dieser Stelle schwer, da durch die Verwendung der Skala „Klassenführung“ bzw. in Mathematik der Skalen „effizientes Zeitmanagement“ und „effektiver Umgang mit Störungen“ nicht genau gesagt werden kann, ob dieser Effekt durch das Zusammenspiel der Einzelmerkmale der Klassenführung zustande kommt oder eines der erfassten Einzelmerkmale hier beson-

ders ins Gewicht fällt. So zeigen Ergebnisse aus der Sekundarstufe, dass den einzelnen Merkmalen eine unterschiedliche Bedeutung zukommt. Während eine intensive Zeitznutzung und somit eine ausgeprägte Lehrstofforientierung sowie Disziplin und Strukturiertheit einen positiven Effekt auf das Selbstkonzept der Schüler in der Sekundarstufe haben, geht ein hohes Interaktionstempo und damit zusammenhängend eine gewisse Zielorientierung eher mit einem negativeren Selbstkonzept der Schüler einher (vgl. Kapitel 5.1.2). Weitere Analysen mit den Einzelmerkmalen können für die vorliegende Stichprobe Klarheit darüber bringen, welche Einzelmerkmale eventuell gegenläufige Effekte haben und ob die Zusammensetzung der Skala für die Analyse von Wirkungen sinnvoll ist. Beispielsweise kann angenommen werden, dass das Merkmal „Gruppenfokus (Grup)“ tendenziell mit einem geringeren Selbstkonzept der Schüler einhergeht. Eine positive Ausprägung dieses Merkmals bedeutet, dass die PERLE-Lehrpersonen versuchen, möglichst allen Schülern die Gelegenheit zu geben, sich aktiv am Unterrichtsgeschehen zu beteiligen, indem sie beispielsweise auch Schüler aufrufen, die sich nicht melden. Die Redeanteile der Schüler sind somit relativ gleichmäßig verteilt. Durch dieses Vorgehen werden deutlich mehr öffentliche Lerngelegenheiten geschaffen, die vermutlich zu mehr sozialen Vergleichen führen und somit negative Konsequenzen für das Selbstkonzept der Schüler haben können. Auch die öffentliche Zurechtweisung von störenden bzw. unaufmerksamen Schülern kann mit geringeren Selbstkonzepten der Schüler einhergehen. Allerdings beruht diese Erklärung auf der Annahme, dass das Verhalten, das die Lehrpersonen der PERLE-Stichprobe in den videografierten Stunden gezeigt haben, repräsentativ für ihr grundsätzliches Verhalten ist.

Eine andere Interpretation erfahren die tendenziell negativen Effekte der Klassenführung auf die Selbstkonzepte der Schüler, wenn man berücksichtigt, dass eine Steigerung des Selbstkonzepts nicht unbedingt als anzustrebender Zustand betrachtet wird. Vor dem Hintergrund, dass bisherige Befunde im Kontext der PERLE-Studie belegen, dass die Schüler ihre Fähigkeiten im Rechtschreiben und Rechnen im ersten und zweiten Schuljahr stark überschätzen (z. B. Gabriel et al., 2010) und sich erst im Verlauf der Grundschule realistischer beurteilen<sup>99</sup> (vgl. Kastens & Gabriel, in Vorb.), lassen sich die hier berichteten Ergebnisse auch so werten, dass eine effektive Klassenführung einen Beitrag zu einem geringeren und somit wahrscheinlich auch zu einem realistischeren Selbstkonzept in beiden Domänen leisten kann. Dabei stellt sich die Frage, ob dieser Befund sowohl für die leistungsschwächeren als auch leistungsstärkeren Schüler Gültigkeit hat. Kammermeyer und Martschinke (2003) diskutieren, dass vor allem bei leistungsschwächeren Schülern eine Überschätzung der eigenen Fähigkeiten vermutlich eher dazu führt,

---

99 Realistischer bedeutet in diesem Zusammenhang, dass sich die Einschätzung der eigenen Fähigkeiten zunehmend den tatsächlichen Leistungen annähert. Der Zusammenhang wird enger.

dass sie sich beispielsweise auch auf schwierigere Aufgaben einlassen und somit unterschiedliche Erfahrungen machen. Zudem führt die damit zusammenhängende hohe Motivation dazu, dass leistungsschwächere Schüler sich bei Misserfolgen nicht so schnell entmutigen lassen (vgl. Kammermeyer & Martschinke, 2003). Ein realistischeres Selbstkonzept kann demnach für leistungsschwächere Schüler zu einem Verlust der Lernmotivation und Anstrengungsbereitschaft führen. Im Vergleich dazu kann die leichte Überschätzung der eigenen Fähigkeiten für leistungsstärkere Schüler ein Risiko darstellen, weil notwendige Lernanstrengungen unterbunden werden bzw. die Schüler den Lernerfolg als selbstverständlich erachten (zusf. Helmke, 1998). Die Überprüfung dieser differenziellen Effekte konnte im Rahmen der vorliegenden Arbeit jedoch nicht mehr geleistet werden, sollte sich aber anschließen.

Grundsätzlich kann bisher weder für die Klassenführung noch für andere Unterrichtsqualitätsmerkmale definiert werden, in welcher Ausprägung und Intensität diese als produktiv gelten und wann eine hohe Ausprägung des Merkmals kontraproduktiv wird (vgl. Bohl & Kucharz, 2010). Im Rahmen der PERLE-Studie sind vor allem die Merkmale der Klassenführung in beiden Videostudien / Fächern sehr hoch ausgeprägt. In Anlehnung an frühere Ergebnisse von Helmke (1988) für das fünfte Schuljahr könnte angenommen werden, dass dieses Übermaß an Klassenführung auf Kosten affektiv-motivationaler Schülervariablen geht und damit der negative Effekt auf die Selbstkonzepte erklärbar ist.

#### *Bedeutung des Unterrichtsklimas für multikriteriale Schülermerkmale*

Hinsichtlich der Bedeutung des Unterrichtsklimas für die Leistungsentwicklung der Schüler zeigten sich in der vorliegenden Arbeit Unterschiede in Abhängigkeit vom Fach. Während im Bereich Rechtschreiben kein signifikanter Effekt gefunden werden konnte und somit bisherige Befunde aus der Grundschule (vgl. Hypothese 6, Kapitel 5.2.1) sowie aus der Sekundarstufe (z. B. Klieme et al., 2008) bestätigt werden können, hat das Unterrichtsklima in Mathematik (2. Schuljahr) tendenziell einen negativen Effekt auf die Leistungen der Schüler. Die negative Richtung des Beta-Koeffizienten überrascht, bedeutet er doch inhaltlich, dass ein positiv eingeschätztes Unterrichtsklima, operationalisiert über die „Fürsorglichkeit, Herzlichkeit und Wärme (Fuer)“ sowie die „Anerkennung der Schüler durch die Lehrperson (Anes)“ und durch eine „humorvolle Lernatmosphäre (Humo)“ tendenziell mit etwas geringeren Leistungen der Schüler im Fach Mathematik am Ende des zweiten Schuljahres einhergeht, nach Kontrolle der anderen Prädiktoren auf Level 1 und 2. Pianta et al. (2008) kommen in ihrer Studie zu ähnlichen Ergebnissen für das Fach Mathematik. Sie schlussfolgern, dass es im Fach Mathematik im Vergleich zu anderen Fächern eher um das Lehren und Einüben technischer Fertigkeiten geht und der Fokus weniger auf der emotionalen Unterstützung liegt (vgl. Kapitel 5.2.1).

Es stellt sich in Bezug auf die Zusammensetzung der Skala Unterrichtsklima die Frage, ob die Reduzierung auf die Merkmale „Fürsorglichkeit, Herzlichkeit und Wärme (Fuer)“, „Anerkennung der Schüler durch die Lehrperson (Anes)“ und „humorvolle Lernatmosphäre (Humo)“ (vgl. Kapitel 15) – die sich aus der Überprüfung der Reliabilität ergab – nicht dazu geführt hat, dass die Skala Unterrichtsklima letztendlich nur Ausdruck einer positiven Beziehungsgestaltung zwischen Lehrperson und Schülern und somit einer positiven Schülerorientierung (vgl. Kapitel 4.1.1) ist. Aspekte der Lernumwelt im Sinne unterstützender (z. B. Lob und Ermutigung) bzw. sachlich-konstruktiver Rückmeldungen wurden für die Skalenbildung nicht weiter berücksichtigt, da diese sowohl in Deutsch (1. Schuljahr) als auch in Mathematik (2. Schuljahr) auf einem dritten Faktor luden bzw. im Falle des Merkmals „Einsatz von Lob (Lob)“ keinem der Faktoren zugeordnet werden konnten. Da solche Lehrer-Schüler-Interaktionen Schüler in ihren Lernprozessen anregen, begleiten und unterstützen (vgl. Krammer, 2009), kann davon ausgegangen werden, dass diesen Merkmalen eher eine positive Bedeutung für die Leistungsentwicklung zukommt (auch im Fach Deutsch, 1. Schuljahr). Dies sollte in weiteren Analysen überprüft werden. Solche positiven Rückmeldungen im Sinne eines informierenden Feedbacks ermöglichen den Schülern zusätzlich „die realistische Wahrnehmung der eigenen Stärken und Schwächen und zeigen auf, wo konkret individuelle Verbesserungsmöglichkeiten liegen“ (Waldis et al., 2010b, S. 181), womit solchen positiven Rückmeldungen auch ein Zusammenhang mit dem Kompetenzerleben und dem Selbstkonzept der Schüler attestiert werden kann. Allerdings muss hier zwischen der reinen Mitteilung, ob ein Ergebnis richtig oder falsch ist, und individuellen Verbesserungsvorschlägen, die Lernende im Erreichen ihres Lernziels unterstützen, unterschieden werden.

Die hier verwendete Skala Unterrichtsklima besteht eher aus Verhaltensweisen, die als förderlich für die Persönlichkeitsentwicklung angesehen werden können (vgl. Krammer, 2009), was durch den positiven Effekt des Unterrichtsklimas auf das Selbstkonzept im Rechnen am Ende des zweiten Schuljahres bestätigt wird (vgl. Modelle 2c und 2d). Auch im Fach Deutsch (1. Schuljahr) zeigt sich tendenziell ein positiver Effekt des Unterrichtsklimas auf das Selbstkonzept im Schreiben vom Ende des ersten Schuljahres, dieser wird jedoch nach Kontrolle der übrigen Prädiktoren nicht signifikant. Hypothese 7 kann somit nur zum Teil bestätigt werden.

### *Methodische Einschränkungen im Rahmen der Studie 3*

Im Rahmen der Studie 3 wurden Mehrebenenregressionsmodelle in *MPlus* gerechnet, um die Wirkungen der beiden Unterrichtsqualitätsmerkmale Klassenführung und Unterrichtsklima auf kognitive und nicht-kognitive Merkmale im Anfangsunterricht in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) zu untersuchen. Nicht unerwähnt bleiben sollen an dieser Stelle methodische Kritikpunkte, die sich unter anderem erneut aus der

Stichprobengröße ergeben. So existieren nur vereinzelt Erkenntnisse zu Minimalanforderungen an die Stichprobengröße bei einer Mehrebenenmodellierung (zusf. Nezlek et al. 2006; Lüdtke et al., 2007). Für einfache 2-Ebenen-Modelle reichen 30 Level-2-Einheiten aus. Sobald Interaktionen zwischen den Ebenen untersucht werden sollen bzw. Varianzanteile von Interesse sind, sind auf Level 2 mindestens  $N = 50$  Untersuchungseinheiten erforderlich. Allerdings stößt diese Forderung vor allem bei der Datenerhebung mittels Videos an praktische und ökonomische Grenzen. Im Kontext der vorliegenden Untersuchung wurden im Fach Deutsch (1. Schuljahr)  $N = 37$  und in Mathematik (2. Schuljahr)  $N = 36$  Level-2-Einheiten herangezogen. Um valide Aussagen über die Effekte der Unterrichtsqualitätsmerkmale in der Grundschule zu überprüfen, sollten mindestens 50 bzw. bis zu 100 Level-2-Einheiten zur Verfügung stehen (zusf. Nezlek et al. 2006; Lüdtke et al., 2007). Die hier gefundenen Effekte der Unterrichtsqualitätsmerkmale sollten demnach nicht überinterpretiert werden. Es kann vermutet werden, dass die Effekte für die Gesamtpopulation deutlich geringer ausfallen.

Nicht nur die Stichprobengröße, sondern auch der Umgang mit fehlenden Werten sollte im Zusammenhang mit den Ergebnissen der Studie 3 diskutiert werden. Die vorliegenden Mehrebenenanalysen basieren durch die Voreinstellung in *MPlus* lediglich auf vollständigen Datensätzen. Das bedeutet, dass nur diejenigen Fälle auf Individualebene (Level 1) herangezogen wurden, die entsprechende Werte aufgewiesen haben. Alle anderen blieben im Sinne des fallweisen Ausschlusses (Listwise Deletion) unberücksichtigt. Mit diesem gewählten Vorgehen ist nach Hochweber (2010) „zum einen ein Verlust an Power verknüpft (der Möglichkeit, in der Population bestehende Effekte nachzuweisen), zum anderen kann der Fallausschluss zu verzerrten Effekten führen“ (S. 153). Da im vorliegenden Fall mindestens von MAR (Missing at Random) ausgegangen werden kann (vgl. Kapitel 11.5), ist der Fallausschluss jedoch tendenziell unproblematisch, da in diesem Fall ‚Missings‘ bei den Prädiktoren unabhängig von den Ausprägungen auf dem Kriterium auftreten (vgl. Hochweber, 2010). Für weitere Analysen und zum Vergleich mit den vorliegenden Ergebnissen, sollten sich jedoch Analysen auf Basis imputierter Daten anschließen. Hier stellt sich allerdings die Frage nach einem geeigneten Vorgehen bzw. Verfahren, denn bei der Imputation der Daten sollte die hierarchische Struktur mit berücksichtigt werden, da dies sonst zu einer Unterschätzung der Intraklassenkorrelation der relevanten Variablen führen könnte (vgl. van Buuren, 2005). Da die Intraklassenkorrelation an sich die Ausprägung der Mehrebenenstruktur darstellt, würde eine solche hierarchische Schachtelung durch ein imputationsbasiertes Verfahren unterschätzt werden.

Im Kontext der Studie 3 stellt sich zusätzlich die Frage, ob die Auswahl der Zielkriterien überdacht werden sollte. Einerseits wurden die Zielkriterien einige Monate nach der Aufzeichnung der Unterrichtsvideos erhoben, andererseits ist die Wahl der Zielkriterien (Leistungen und Selbstkonzept) in Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) insofern problematisch,

als sie selbst in einem gewissen Maß kumulativ sind. Der Zusammenhang zwischen den Leistungen und dem Selbstkonzept der Schüler nimmt im Verlauf des Anfangsunterrichts zu (vgl. van Aken, Helmke & Schneider, 1997; Kammermeyer & Martschinke, 2003), auch in der PERLE-Stichprobe (vgl. Kastens & Gabriel, in Vorb.). Dies hätte in den Mehrebenenregressionsmodellen insofern berücksichtigt werden müssen, als auf Individualebene (Level 1) gleichzeitig sowohl Leistungs- als auch Selbstkonzeptdaten modelliert werden. In diesem Zusammenhang könnten sich zusätzliche Analysen anschließen, die ähnlich wie bereits in der KILIA- bzw. SCHOLASTIK-Studie (vgl. Kammermeyer & Martschinke, 2003; Renkl, Helmke & Schrader, 1997) den Einfluss der Klassenführung und des Unterrichtsklimas auf den Zusammenhang zwischen Leistung und Selbstkonzept untersuchen.

Zusätzlich repräsentieren die abhängigen Variablen in beiden Fächern selbst eine „dauernde persönliche Entwicklung“ (Eder, 2011, S. 120). Eder (2011) merkt in diesem Zusammenhang kritisch an, dass sich sowohl Persönlichkeitsmerkmale (z. B. Selbstkonzept) als auch Leistungsmerkmale mehr oder weniger kontinuierlich über die Zeit entwickeln, sodass bei einer punktuellen Messung des Unterrichtsklimas bzw. anderer Unterrichtsqualitätsmerkmale wie der Klassenführung nur ein geringer Einfluss erwartbar wäre. Hier sollten sich Analysen anschließen, die kurzfristige Effekte der Unterrichtsqualitätsmerkmale untersuchen. Im Rahmen der PERLE-Videostudien wurden zusätzlich Leistungsdaten erfasst, die jedoch aus verschiedenen Gründen in der vorliegenden Arbeit nicht herangezogen wurden. Beispielsweise erzeugte der Gruppentest (vgl. Kapitel 9.1), der nach der aufgezeichneten Unterrichtseinheit im Fach Deutsch durchgeführt wurde, kaum Varianz, d. h. fast alle Schüler beantworteten alle Fragen richtig. Für Mathematik (2. Schuljahr) liegen jedoch entsprechende Leistungsdaten der Schüler vor und nach der Videostudie vor.

#### *Weiterführende Forschungsfragen*

Da sich im Bereich Rechtschreiben weder für die Klassenführung noch für das Unterrichtsklima signifikante Effekte auf die Leistungsentwicklung zeigten, sollten in weiteren Analysen zusätzliche Unterrichtsmerkmale oder fachspezifische Merkmale als Prädiktoren herangezogen werden, um die vorhandenen Unterschiede zwischen den Klassen (der ICC für die Rechtschreibleistungen liegt bei ca. 11 %, vgl. Kapitel 16) zu erklären. So konnten bisherige Analysen im Rahmen der PERLE-Studie bereits zeigen, dass im Bereich Rechtschreiben am Ende des ersten Schuljahres auch die Zugehörigkeit zu einer BIP-Klasse einen positiven Einfluss auf die Leistungsentwicklung der Schüler hat. Lotz et al. (2013) bestätigten im Bereich Rechtschreiben für das Ende des ersten Schuljahres Leistungsvorsprünge der BIP-Schüler gegenüber den Schülern an staatlichen Grundschulen, die sich im zweiten Schuljahr noch verstärken. Im Bereich Mathematik sieht dies ein wenig anders aus. Während sich ähnlich wie im Rechtschreiben im ersten Schuljahr Vorsprün-

ge der BIP-Schüler bestätigen ließen, setzt sich diese Entwicklung im zweiten Schuljahr nicht fort (vgl. Karst & Lipowsky, 2013). Allerdings sollte grundsätzlich gefragt werden, wieviele Unterschiede in den Leistungen im Fach Mathematik überhaupt durch Klassenvariablen aufgeklärt werden können, denn im Vergleich zum Bereich Rechtschreiben ist der ICC hier deutlich geringer (vgl. Kapitel 16).

Neben der Zugehörigkeit zu einer BIP-Klasse könnten zur Erklärung der Leistungsunterschiede im Rechtschreiben zwischen den Klassen weitere Unterrichtsqualitätsmerkmale herangezogen werden. Dabei stellt sich die Frage, ob die Ergebnisse aus der SCHOLASTIK-Studie (3./4. Schuljahr) auch für das erste Schuljahr bestätigt werden können. Im Rahmen der SCHOLASTIK-Studie erwiesen sich im Bereich Rechtschreiben neben der Klassenführung und dem sozialen Klima weder die Strukturiertheit, die individuelle fachliche Unterstützung, die Förderungsorientierung, die Variabilität der Unterrichtsformen noch die Klarheit als relevante Prädiktoren für die Leistungsentwicklung im dritten und vierten Schuljahr. Lediglich die Motivierungsqualität des Unterrichts, „die sich in aktiver Beteiligung der Schüler am Unterrichtsgeschehen und dem vergleichsweise seltenen Vorkommen passiver Formen der Unaufmerksamkeit [...] manifestiert“ (Helmke & Weinert, 1997, S. 248), stellte eine wichtige Bedingung für die Leistungsentwicklung im Bereich Rechtschreiben dar. Die Autoren schlussfolgerten daraus, dass individuelle Fortschritte im Rechtschreiben eher durch außerschulische Faktoren erklärt werden können. Inwieweit außerschulische Merkmale auch im ersten Schuljahr individuelle Rechtschreibleistungen erklären können, sollte noch untersucht werden. Vermutlich spielen jedoch eher fachspezifische Merkmale eine Rolle. Ergebnisse aus der EVES-Studie für das dritte und vierte Schuljahr zeigen beispielsweise, dass vor allem „die Zeit, die für die Besprechung und Erläuterung von Rechtschreibphänomenen, Regeln, Merkhilfen oder zur Kontrolle von Übungsaufgaben genutzt wurde, [...] mit einer positiven Leistungsentwicklung im Rechtschreiben in Verbindung [steht]“ (Roos & Schöler, 2009, S. 193). Für den Anfangsunterricht kann in der Studie von Klicpera und Gassteiger-Klicpera (1993) bestätigt werden, dass fachspezifische Merkmale gegenüber allgemeinen Unterrichtsmerkmalen bedeutsamer für die Leistungen im Rechtschreiben sind.

Im Rahmen der Studie 3 wurden lediglich Effekte einer effektiven Klassenführung und eines positiven Unterrichtsklimas für die Bereiche Rechtschreiben (1. Schuljahr) und Mathematik (2. Schuljahr) untersucht. Da zusätzlich Daten zur Entwicklung der Leseleistungen und der entsprechenden Selbstkonzepte im Lesen vorliegen, sollten sich zusätzliche Mehrebenenanalysen für den Bereich Lesen anschließen. Hier wäre interessant zu überprüfen, ob auch im Bereich Lesen eher fachspezifische Merkmale gegenüber allgemeinen Merkmalen von Unterricht eine Rolle für die Leistungsentwicklung spielen. Zudem liegen bislang kaum Studien vor, die Unterschiede in der Bedeutung der Unterrichtsqualitätsmerkmale für die Domänen Lesen und

Schreiben gleichzeitig untersuchen. Da nach Klicpera und Gasteiger-Klicpera (1993) in den ersten Schuljahren die Entwicklung der Lesekompetenz nicht zwangsläufig parallel zur Rechtschreibentwicklung verlaufen muss, könnten entsprechende Unterschiede in der Bedeutung angenommen werden. Für deutsche Grundschulen bestätigen Helmke et al. (2010) im Rahmen der Studie „VERA – Gute Unterrichtspraxis“ (vgl. Helmke et al. 2008) einen positiven Zusammenhang der effektiven Klassenführung mit der Lesekompetenz für das vierte Schuljahr ( $r = .62$ ). Das bedeutet: Je weniger Zeit für außerschulische Angelegenheiten verschwendet wurde bzw. je weniger Störungen im Unterricht vorkommen, desto höher war die Lesekompetenz der Viertklässler auf Klassenebene. Auch für das erste Schuljahr konnte der positive Einfluss einer effektiven Klassenführung mit den Leistungen der Schüler für die ersten Schuljahre im Rahmen einzelner Studien bestätigt werden (z. B. Klicpera & Gasteiger-Klicpera, 1993). In der Studie von Pianta et al. (2008) ergaben sich kleine positive Effekte der beobachteten emotionalen Qualität der Lehrer-Schüler-Interaktionen und dem Anstieg der Leseleistungen für die Gruppe der sogenannten „normalen Leser“ im dritten und fünften Schuljahr. Schüler, deren Leseleistungen sich normal entwickeln, profitieren demnach von einer positiven emotionalen Unterstützung durch die Lehrperson. Im ersten Schuljahr zeigte sich der umgekehrte Fall. Je höher die Beobachter die emotionale Unterstützung einschätzten, desto langsamer entwickelten sich die Leseleistungen der „normalen“ Schüler. Es konnte zudem im ersten Schuljahr eine bedeutende Interaktion zwischen der Qualität der emotionalen Unterstützung und der Quantität im Sinne einer aktiven Lernzeitnutzung bestätigt werden, d. h., dass sich Klassen mit einer hohen beobachteten emotionalen Unterstützung durch einen negativen Zusammenhang zwischen der aktiven Zeitnutzung und dem Lesewachstum auszeichneten.

#### *Unterricht als Produkt: Veränderter Blickwinkel?*

In den Mehrebenenregressionsmodellen in Studie 3 blieb bislang unberücksichtigt, dass guter Unterricht im Sinne einer effektiven Klassenführung und eines positiven Unterrichtsklimas selbst auch eine Folge vergangener Entwicklungsverläufe sein kann. So konnte Helmke (1992) im Kontext der Münchner Hauptschulstudie (5./6. Schuljahr) anhand der Berechnung eines Pfadmodells zeigen, dass das durchschnittliche mathematische Vorkenntnisniveau der Klasse zu Beginn des fünften Schuljahres einen starken Einfluss auf die Gestaltung bzw. Konzeption des Unterrichts hat. Je niedriger das kognitive Eingangsniveau der Klasse, desto größer ist die Anforderung an die Lehrperson, einen für alle verständlichen Unterricht zu halten und desto mehr Zeit muss sie für andere als fachliche Belange (z. B. für private, soziale aber auch disziplinarische Aktivitäten) aufwenden. Auch im Rahmen der DESI-Studie belegen die Autoren für das Fach Deutsch, dass eine schlechtere Ausgangslage einer Klasse positive Auswirkungen auf die Strukturiertheit und die Unterstützung hat. Außerdem werden in solchen Klassen mehr Lernge-

legenheiten angeboten, z. B. bezüglich des Wortschatzes, der Rechtschreibung oder des Briefeschreibens. Demnach reagieren Lehrpersonen im Fach Deutsch adaptiv auf schwierige Lernausgangslagen. Es konnte kein Effekt der Ausgangsleistung zu Beginn des neunten Schuljahres auf das Unterrichtsklima aufgezeigt werden (vgl. Klieme et al., 2008). Um Aussagen über die Bedeutung der Ausgangslage im Rechtschreiben und in Mathematik im Anfangsunterricht für die Unterrichtsgestaltung machen zu können, sollten zusätzliche Analysen folgen.

#### *Wirkungen einzelner Unterrichtsmerkmale auf Schülermerkmale*

Wie bereits im Vorfeld erwähnt, wurde im Rahmen der vorliegenden Arbeit nicht untersucht, ob einzelne Merkmale der Klassenführung oder des Unterrichtsklimas unterschiedliche Bedeutungen für die Leistungs- und Selbstkonzeptentwicklung der Schüler in den beiden Fächern haben. Obwohl in der vorliegenden Arbeit bestätigt werden konnte, dass sich die einzelnen hoch inferenten Merkmale der Klassenführung und des Unterrichtsklimas faktoranalytisch zu verschiedenen Skalen zusammenfügen lassen, sollte in weiteren Analysen auch die Wirkung der Einzelmerkmale auf multikriteriale Schülermerkmale untersucht werden. So könnte zum Beispiel für die Selbstkonzeptentwicklung in den beiden Fächern überprüft werden, welches Merkmal der Klassenführung als stärkster Prädiktor identifiziert werden kann. Hierbei ist es sinnvoll, fachspezifische Korrelationen der einzelnen Merkmale mit den abhängigen Variablen zu berechnen und die Einzelmerkmale entsprechend ihrer Größe des Zusammenhangs mit der abhängigen Variable in ein Mehrebenenregressionsmodell aufzunehmen, beginnend mit dem Merkmal, das den größten Zusammenhang aufweist. Diese Vorgehensweise ermöglicht es, ein Modell zu identifizieren, das mit möglichst wenig Variablen möglichst viel Varianz in der abhängigen Variable erklärt.

#### *Ergänzende Analysen mithilfe eines personenzentrierten Ansatzes*

Durch die Wahl des variablenzentrierten Ansatzes in der vorliegenden Arbeit gerät die Komplexität des Unterrichts aus dem Blick (vgl. Helmke et al., 2007). Deshalb empfiehlt Helmke (2009), dass ein variablenzentrierter Ansatz durch einen personenzentrierten ergänzt werden sollte. Im Hinblick auf die Forderung nach multikriterialer Zielerreichung in der Grundschule (vgl. Kapitel 5) bedeutet dies, dass beispielsweise in weiteren Analysen diejenigen Klassen identifiziert werden sollten, die als „erfolgreich“ im Sinne einer positiven Leistungs- und Selbstkonzeptentwicklung gelten. Hier können im Rahmen zusätzlicher Analysen in PERLE weitere Zielvariablen berücksichtigt werden (z. B. Interesse am jeweiligen Fach, Lernmotivation, Lernfreude etc.). In einem zweiten Schritt sollten diese Klassen und ihr Unterricht / ihre Lehrperson näher betrachtet werden. Hierbei sollte nicht nur die Klassenführung oder das Unterrichtsklima fokussiert werden. Die Identifikation kann in diesem Zusammenhang mithilfe von Clusteranalysen erfolgen. Für die Grundschule identifizierten beispielsweise Weinert und Helmke (1996) mithilfe die-

ser Methode im Rahmen der SCHOLASTIK-Studie für das 3./4. Schuljahr 13 Optimalklassen, die auf allen zugrunde gelegten Zielkriterien (Leistungsentwicklung in Arithmetik und im mathematischen Problemlösen, Egalisierung von Leistungsunterschieden in Mathematik, Entwicklung des Fähigkeits-selbstbildes sowie Entwicklung der Lernfreude in Mathematik) überdurchschnittliche Werte erzielten. Ein Vergleich dieser Optimalklassen mit den übrigen teilnehmenden Klassen ergab vor allem Unterschiede im Hinblick auf die Klarheit und Strukturiertheit des Unterrichts, die Effektivität der Klassenführung, die Förderung aufgabenbezogener Schüleraktivitäten, die individuelle Unterstützung der Schüler sowie die Variabilität der Unterrichtsformen. Keine Unterschiede konnten hingegen in der Ausprägung des sozialen Klassenklima oder in Bezug auf die Adaptivität des Unterrichts gefunden werden. In den Klassen mit multikriterialer Zielerreichung zeigte sich jedoch auch ein stark divergierendes Muster von Merkmalen der Unterrichtsqualität. Demnach muss Unterricht nicht in allen Merkmalen positiv ausgeprägt sein, um sowohl kognitive als auch affektiv-motivationale Zielbereiche zu fördern. Für den Bereich Rechtschreiben liefert die Studie „Lesen und Schreiben für alle (PLUS)“ von May (2001) ausführliche Porträts von erfolgreichen Klassen (1. Schuljahr). Die Studie bestätigt, dass sich die erfolgreichen Klassen vor allem durch die Gestaltung des Unterrichts von den anderen Klassen unterscheiden. In den lernförderlichen Klassen wird mehr Unterrichtszeit effektiv zum Lernen verwendet, die Lehrperson schenkt den Schülern mehr Aufmerksamkeit und es dominieren Interaktionsprozesse zwischen Lehrpersonen und Schülern, d. h. es gibt z. B. weniger Einzelarbeit. Auch das Sozialklima in diesen Klassen wird positiver von Beobachtern eingeschätzt (vgl. May, 2001).

Das bisherige Vorgehen in diesen Optimalklassenstudien lässt jedoch offen, ob es neben den Unterrichtsqualitätsmerkmalen weitere Komponenten des Unterrichts gibt, die gleichzeitig die kognitive und affektiv-motivationale Entwicklung der Schüler positiv beeinflussen (vgl. Lipowsky, 2009). Ausgeklammert blieben beispielsweise bislang eher fachspezifische Merkmale des Unterrichts oder die Zielsetzung des Unterrichts im Sinne der didaktischen Konzeption bzw. der Unterrichtskonzeption. Auch die Rolle der Öffnung von Unterricht in der Grundschule wurde bislang kaum im Zusammenhang mit Optimalklassen untersucht (Ausnahme KILIA-Studie, vgl. Kammermeyer & Martschinke, 2004). Auch hier sollten sich weitere Studien anschließen.

## VI Fazit

Da die Entwicklung von hoch inferenten Ratingsystemen bzw. entsprechenden Indikatoren beobachtbaren Verhaltens einen relativ hohen Zeitaufwand erfordert, leistet die vorliegende Arbeit einen Beitrag zur videografierten Unterrichtsforschung im Anfangsunterricht. Es konnte gezeigt werden, dass sich Beobachterratings grundsätzlich auch für den Anfangsunterricht eignen, um allgemeine, fachübergreifende Merkmale der Unterrichtsqualität – im vorliegenden Fall die Klassenführung und das Unterrichtsklima – zu erfassen. In Abhängigkeit vom Fach bzw. Schuljahr ergaben sich für die Rater bei der Einschätzung einzelner Merkmale jedoch Schwierigkeiten, was die Frage aufwirft, ob diese Merkmale in verschiedenen Fächern bzw. Schuljahren eine unterschiedliche Bedeutung haben und entsprechend um fach- bzw. schuljahr-spezifische Indikatoren ergänzt werden sollten.

Auch Unterschiede in der faktoriellen Struktur der hier untersuchten Merkmale in Abhängigkeit vom Unterrichtsfach weisen darauf hin, dass im Anfangsunterricht keineswegs von stabilen Strukturen gesprochen werden kann. Die Unterschiede in der Faktorenstruktur sowie in den Faktorladungen und Reliabilitäten lassen den Schluss zu, dass vor allem das Konstrukt der Klassenführung für beide Fächer eine unterschiedliche Bedeutung hat. Während für den Mathematikunterricht im zweiten Schuljahr die Unterscheidung zwischen den Merkmalen der effizienten Zeitnutzung und dem effektiven Umgang mit auftretenden Störungen von Bedeutung ist, scheint diese Unterteilung für das Fach Deutsch (1. Schuljahr) nicht gerechtfertigt. Das Konstrukt der Klassenführung ist demnach über die Fächer bzw. Schuljahre hinweg nicht stabil.

In Bezug auf die Bedeutung der beiden Unterrichtsqualitätsmerkmale zeigte sich, dass sich auch hier fachspezifische Analysen im Anfangsunterricht durchaus lohnen. Aus den hier dargestellten Befunden resultieren verschiedene Anknüpfungspunkte für zukünftige Studien. Allerdings ist aufgrund der methodischen Einschränkungen und der Zusammensetzung der zugrunde liegenden Stichprobe eine Überprüfung der Ergebnisse erforderlich. Vor allem die Brauchbarkeit des hier entwickelten fachunspezifischen hoch inferenten Ratingsystems gilt es, in weiteren Studien und für weitere Fächer zu überprüfen.



## VII Literatur

- Aebli, H. (1994). *Zwölf Grundformen des Lehrens: eine allgemeine Didaktik auf psychologischer Grundlage*. Stuttgart: Klett-Cotta.
- Ainsworth, M. D. S., Blehar, M. C., Waters, E. & Wall, S. (1987). *Patterns of attachment. A psychological study of the strange situation*. Hillsdale: Erlbaum.
- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52(3), 317–332.
- Aken, M. v., Helmke, A. & Schneider, W. (1997). Selbstkonzept und Leistung – Dynamik ihres Zusammenspiels: Ergebnisse aus dem SCHOLASTIK-Projekt. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 341–350). Weinheim: Psychologie Verlags Union.
- Anderson, C. S. (1982). The search for school climate: A review of the research. *Review of Educational Research*, 52(3), 368–420.
- Anderson, L. M., Evertson, C. M. & Brophy, J. E. (1979). An experimental study of effective teaching in first-grade reading groups. *The Elementary School Journal*, 79(4), 193–223.
- Arbinger, R. & Saldern, M. v. (1984). Schulische Umwelt und soziales Klima in Schulklassen. *Psychologie in Erziehung und Unterricht*, 31, 81–99.
- Arlin, M. (1979). Teacher transition can disrupt time flow in classroom. *American Educational Research Journal*, 16(1), 42–56.
- Arnold, K.-H., Koch-Priewe, B. & Lin-Klitzing, S. (2007). Allgemeine Didaktik, Fachdidaktik und Unterrichtsqualität. In K.-H. Arnold (Hrsg.), *Unterrichtsqualität und Fachdidaktik* (S. 19–50). Bad Heilbrunn: Klinkhardt.
- Aufschnaiter, S. v. & Welzel, M. (Hrsg.) (2001). *Nutzung von Videodaten zur Untersuchung von Lehr-Lern-Prozessen: Aktuelle Methoden empirischer pädagogischer Forschung*. Münster: Waxmann.
- Babu, S. & Mendro, R. (2003). *Teacher accountability: HLM based teacher effectiveness indices in the investigation of teacher effects on student achievement in a state assessment program*. Paper presented at the annual meeting of the American Educational Research Association. Chicago. Online verfügbar: [http://www.dallasisd.org/inside\\_disd/depts/evalacct/research/articles.htm](http://www.dallasisd.org/inside_disd/depts/evalacct/research/articles.htm). (Zugriff: 16.03.2010).
- Baumert, J. (1997). Zielkonflikte in der Grundschule: Kommentar. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 317–321). Weinheim: Psychologie Verlags Union.
- Baumert, J. & Köller, O. (2000). Unterrichtsgestaltung, verständnisvolles Lernen und multiple Zielerreichung im Mathematik- und Physikunterricht der gymnasialen Oberstufe. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS / III – Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn* (S. 271–315). Opladen: Leske & Budrich.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Dubberke, T., Jordan, A. et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180.
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I. et al. (Hrsg.) (1997). *TIMSS – Mathematisch naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske & Budrich.

- Bem, D. J. & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review*, 81(6), 506–520.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246.
- Bessoth, R. (1989). *Verbesserung des Unterrichtsklimas. Grundlagen, Aufbau und Einsatz von Instrumenten*. Neuwied: Luchterhand.
- Bloom, B. S. (1973). Individuelle Unterschiede in der Schulleistung: ein überholtes Problem. In W. Edelstein & D. Hopf (Hrsg.), *Bedingungen des Bildungsprozesses. Psychologische und pädagogische Forschungen zum Lehren und Lernen in der Schule* (S. 251–270). Stuttgart: Klett.
- Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw Hill.
- Blumberg, E., Möller, K. & Hardy, I. (2004). Erreichen motivationaler und selbstbezogener Zielsetzungen in einem schülerorientierten naturwissenschaftsbezogenen Sachunterricht – Bestehen Unterschiede in der Abhängigkeit von der Leistungsstärke? In E.-M. Lankes, W. Bos, N. Pfaffmeier & K. Schwippert (Hrsg.), *Heterogenität. Eine Herausforderung an die empirische Bildungsforschung* (S. 41–52). Münster: Waxmann.
- Bohl, T. & Kucharz, D. (2010). *Offener Unterricht heute. Konzeptionelle und didaktische Weiterentwicklung*. Weinheim: Beltz.
- Boostrom, R. (1991). The nature and function of classroom rules. *Curriculum Inquiry*, 21(2), 193–216.
- Borich, G. D. & Klinzing, H. G. (1986). Die Beobachtung von Unterrichtsprozessen. Einige Voraussetzungen und Empfehlungen zur Erfassung von Verhalten geringen Inferenzgrades. *Unterrichtswissenschaft*, 1, 65–79.
- Bortz, J. (1999). *Statistik. Für Sozialwissenschaftler*. (5. Aufl.). Berlin: Springer.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer-Medizin-Verlag.
- Bos, W., Bonsen, M., Baumert, J., Prenzel, M., Selter, C. & Walther, G. (Hrsg.) (2008). *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bremerich-Vos, A. (2006). Zur Videographie von (Deutsch-)Unterricht. Anmerkungen zur Vermittelbarkeit von linguistischer, sprachdidaktischer und pädagogisch-psychologischer Unterrichtsforschung. In I. Hosenfeld & Schrader F.-W. (Hrsg.), *Schulische Leistung. Grundlagen, Bedingungen, Perspektiven* (S. 243–262). Münster: Waxmann.
- Brennan, R. L. & Kane, M. T. (1977). An index of dependability for mastery tests. *Journal of Educational Measurement*, 14(3), 277–289.
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert, N. Birbaumer & C. F. Graumann (Hrsg.), *Psychologie des Unterrichts und der Schule* (S. 177–212). Göttingen: Hogrefe.
- Brophy, J. E. (1981). Teacher praise: A functional analysis. *Review of Educational Research*, 51(1), 5–32.
- Brophy, J. E. (1983). Classroom organization and management. *The Elementary School Journal*, 83(4), 265–285.
- Brophy, J. E. (2000). *Teaching* (Vol. 1). Brussels: International Academy of Education & International Bureau of Education.
- Brophy, J. & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching. A project of the American Educational Research Association* (pp. 328–375). New York: Macmillan.

- Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park: Sage Publications.
- Brückmann, M. (2009). *Sachstrukturen im Physikunterricht. Ergebnisse einer Videostudie*. Berlin: Logos Verlag.
- Buck, G. H. (1999). Smoothing the rough edge of classroom transition. *Intervention in School and Clinic*, 34(4), 224–227.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Bülter, H. & Meyer, H. (2004). Was ist ein lernförderliches Klima? Voraussetzungen und Wirkungen. *Zeitschrift der Gewerkschaft Erziehung und Wissenschaft*, 11, 31–36.
- Burnett, P. C. (2001). Elementary students' preference for teacher praise. *Journal of Classroom Interaction*, 36(1), 16–23.
- Burnett, P. C. (2002). Teacher praise and feedback and students' perceptions of classroom environment. *Educational Psychologist*, 22(1), 5–16.
- Burns, R. B. (1984). How time is used in elementary schools: the activity structure of classrooms. In L. W. Anderson (Ed.), *Time and school learning. Theory, research and practice* (pp. 91–127). London: Croom Helm.
- Buuren, S. van (2011). Multiple imputation of multilevel data. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 173–196). New York: Routledge.
- Buysse, E., Verschueren, K., Verachtert, P. & Van Damme, J. (2009). Predicting school adjustment in early Elementary school: Impact of Teacher-Child-Relationship quality and relational classroom climate. *The Elementary School Journal*, 110(2), 119–141.
- Cadima, J., Leal, T. & Burchinal, M. (2010). The quality of teacher-student-interactions: Associations with first graders' academic and behavioral outcomes. *Journal of School Psychology*, 48, 457–482.
- Caldwell, J. H., Huitt, W. G. & Graber, A. O. (1982). Time spent in learning: Implications from research. *The Elementary School Journal*, 82(5), 470–480.
- Cameron, C. E., McDonald Connor, C. & Morrison, F. J. (2005). Effects of variation in teacher organization and classroom functioning. *Journal of School Psychology*, 43, 61–5.
- Cameron, C. E., McDonald Connor, C., Morrison, F. J. & Jewkes, A. M. (2008). Effects of classroom organization on letter-word reading in first grade. *Journal of School Psychology*, 46(2), 173–192.
- Campbell, J., Kyriakides, L., Muijs, D. & Robinson, W. (2004). *Assessing teacher effectiveness. Developing a differentiated model*. London: Routledge Falmer.
- Carroll, J. B. (1963). A model of school learning. *Teacher College Record*, 64(8), 723–733.
- Carter, K. & Doyle, W. (2006). Classroom Management in Early Childhood and Elementary Classrooms. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of Classroom Management. Research, Practice, and Contemporary Issues* (pp. 373–406). New York: Routledge.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464–504.
- Chott, P. O. (1999). Ansätze zur Förderung einer ‚Fehlerkultur‘. Lernförderung in der Schule durch Fehlerprophylaxe und Fehlermanagement. *Pädagogisches Forum*, 3, 238–248.
- Christensen, C. M. (1960). Relationship between pupil achievement, pupil affect-need, teacher warmth, and teacher permissiveness. *Journal of Educational Psychology*, 51(3), 169–174.
- Clare, L., Valdés, R., Pascal, J. & Steinberg, J. R. (2001). *Teachers' assignments as indicators of instructional quality in elementary schools* (CSE Technical Report No. 545). Los Angeles (National Center for Research on Evaluation, Standards, and Student Testing.). Online verfügbar: <http://www.cse.ucla.edu/products/reports/tr545.pdf>. (Zugriff: 31.01.2013).

- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität*. Münster: Waxmann.
- Clausen, M., Reusser, K. & Klieme, E. (2003). Unterrichtsqualität auf der Basis hoch-inferenter Unterrichtsbeurteilungen. Ein Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft*, 31(2), 122–141.
- Clausen, M., Schnabel, K. & Schröder, S. (2002). Konstrukte der Unterrichtsqualität im Expertenurteil. *Unterrichtswissenschaft*, 30(3), 246–260.
- Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research*, 77(1), 113–143.
- Corvacho del Toro, I., Berner, N. E. & Mösko, E. (2013). Richtlinien zu den Film- und Tonaufnahmen der PERLE-Videostudien. In M. Lotz, F. Lipowsky & G. Faust (Hrsg.), *Technischer Bericht zu den PERLE-Videostudien*. Materialien zur Bildungsforschung. Bd. 23/3 (S. 53–65). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPPF).
- Creemers, B. P. M. (1994a). Effective instruction: An empirical basis for theory of educational effectiveness. In D. Reynolds (Ed.), *Advances in school effectiveness research and practice* (pp. 189–205). Oxford: Pergamon.
- Creemers, B. P. M. (1994b). *The effective classroom*. London: Cassell.
- Cronbach, L. J., Gleser, G. C., Nanda, H. & Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: John Wiley.
- Dalehefte, I. M. (2001). *Lernmotivation im Physikunterricht. Eine Videostudie zur Untersuchung motivationsunterstützender Bedingungen im Unterrichtsverlauf*. Unveröffentlichte Diplomarbeit, Christian-Albrechts-Universität zu Kiel.
- Davis, H. A. (2001). The quality and impact of relationships between elementary school students and teachers. *Contemporary Educational Psychology*, 26(4), 431–453.
- Deci, E. & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihrer Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, 39(2), 223–238.
- DESI-Konsortium (Hrsg.) (2008). *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie*. Weinheim: Beltz. Online verfügbar: [http://www.pedocs.de/volltexte/2010/3149/pdf/978\\_3\\_407\\_25491\\_7\\_1A\\_D\\_A.pdf](http://www.pedocs.de/volltexte/2010/3149/pdf/978_3_407_25491_7_1A_D_A.pdf) (Zugriff: 31.01.2013).
- Dinkmeyer, D. & Dreikurs, R. (1970). *Ermutigung als Lernhilfe*. Stuttgart: Klett Verlag.
- Ditton, H. (1998). *Mehrebenenanalyse. Grundlagen und Anwendungen des Hierarchisch Linearen Modells*. Weinheim: Juventa.
- Ditton, H. (2002). Unterrichtsqualität – Konzeptionen, methodische Überlegungen und Perspektiven. *Unterrichtswissenschaft*, 30(3), 197–212.
- Ditton, H. (2000). Qualitätskontrolle und Qualitätssicherung in Schule und Unterricht. *Zeitschrift für Pädagogik*, 41. Beiheft, 73–92.
- Doucet, S. A. & Wilsdorf, A. (2005). *Lucy rettet Mama Krokodil*. Hamburg: Oetinger.
- Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Ed.), *Handbook of research on teaching. A project of the American Educational Research Association* (pp. 392–431). New York: Macmillan.
- Dreesmann, H. (1978). *Das Unterrichtsklima als situative Bedingung für kognitive Prozesse und das Leistungsverhalten von Schülern*. Dissertation, Universität Heidelberg.
- Dreesmann, H. (1980). Unterrichtsklima als Bedingung für Lernmotivation. *Unterrichtswissenschaft*, 3, 243–251.
- Dreesmann, H. (1982). *Unterrichtsklima. Wie Schüler den Unterricht wahrnehmen*. Weinheim: Beltz.

- Dreesmann, H., Eder, F., Fend, H., Pekrun, R., Saldern, M. v. & Wolf, B. (1992). Schulklima. In K. Ingenkamp, R. S. Jäger, H. Petillon & B. Wolf (Hrsg.), *Empirische Pädagogik 1970–1990. Eine Bestandsaufnahme der Forschung in der Bundesrepublik* (S. 655–682). Weinheim: Deutscher Studien Verlag.
- Dubs, R. (2002). Der Rückzug der Schule und der Lehrkräfte auf die Kernkompetenzen – eine berechtigte Forderung, ein neues Schlagwort oder ein korrigierter Pendelschlag? *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 98(1), 1–7.
- Duijvenvoorde, A. C. K., Zanolie, K., Rombouts, S. A. R. B., Raijmakers, M. E. J. & Crone, E. A. (2008). Evaluating the negative or valuing the positive? Neural mechanisms supporting feedback-based learning across development. *The Journal of Neuroscience*, 17(38), 9495–9503.
- Dunn, R. J. & Harris, L. G. (1998). Organizational dimensions of climate and the impact on school achievement. *Journal of Instructional Psychology*, 25(2), 100–114.
- Eckerth, M., Hanke, P. & Hein, A. K. (2012). Schulische Bedingungen des Lehrens und Lernens im Anfangsunterricht der Grundschule – ausgewählte Ergebnisse aus dem FiS-Projekt. In F. Hellmich, S. Förster & F. Hoya (Hrsg.), *Bedingungen des Lehrens und Lernens in der Grundschule. Bilanz und Perspektiven* (S. 65–68). Wiesbaden: Springer.
- Eckes, T. (2004). Beurteilerübereinstimmung und Beurteilerstrenge. Eine Multifacetten-Rasch-Analyse von Leistungsbeurteilungen im „Test Deutsch als Fremdsprache“ (TestDaF). *Diagnostica*, 50(2), 65–77.
- Eder, F. (1986). Schulumwelt und Schulzufriedenheit. *Zeitschrift für Erziehungswissenschaftliche Forschung*, 20(1), 83–103.
- Eder, F. (1996). *Schul- und Klassenklima. Ausprägungen, Determinanten und Wirkungen des Klimas an höheren Schulen*. Wien: Studien-Verlag.
- Eder, F. (1998). *Der Linzer Fragebogen zum Schul- und Klassenklima für die 8.–13. Klasse*. Göttingen: Hogrefe.
- Eder, F. (2001). Schul- und Klassenklima. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 578–586). Weinheim: Beltz PVU.
- Eder, F. (2002). Unterrichtsklima und Unterrichtsqualität. *Unterrichtswissenschaft*, 30(3), 213–229.
- Eder, F. (2011). Klassenklima. In E. Kiel & K. Zierer (Hrsg.), *Unterrichtsgestaltung als Gegenstand der Wissenschaft. Basiswissen Unterrichtsgestaltung* (S. 113–127). Baltmannsweiler: Hohengehren.
- Eder, F. & Mayr, J. (2000). *Linzer Fragebogen zum Schul- und Klassenklima für die 4.–8. Klassenstufe (LFSK 4–8)*. Göttingen: Hogrefe.
- Ehmke, T., Siegle, T. & Hohensee, F. (2005). Soziale Herkunft im Ländervergleich. In M. Prenzel (Hrsg.), *PISA 2003. Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche?* (S. 235–268). Münster: Waxmann.
- Eikenbusch, G. (2001). *Qualität im Deutschunterricht der Sekundarstufe I und II*. Berlin: Cornelsen.
- Einsiedler, W. (1988). Schulanfang und Persönlichkeitsentwicklung. *Grundschule*, 10, 20–24.
- Einsiedler, W. (1997). Unterrichtsqualität und Leistungsentwicklung: Literaturüberblick. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 227–240). Weinheim: Beltz.
- Einsiedler, W. (2000). Von Erziehungs- und Unterrichtsstilen zur Unterrichtsqualität. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Pädagogisch-psychologische Aspekte des Lehrens und Lernens in der Schule* (S. 109–128). Opladen: Leske & Budrich.
- Einsiedler, W. (2001). *Ergebnisse und Probleme der Unterrichtsforschung im Primarbereich*. Universität Nürnberg-Erlangen: IFG – Institut für Grundschulforschung (Berichte und Arbeiten aus dem Institut für Grundschulforschung).

- Einsiedler, W. (2003). Unterricht in der Grundschule. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Meyer & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland. Strukturen und Entwicklungen im Überblick* (S. 285–341). Reinbek: Rowohlt Taschenbuch Verlag.
- Einsiedler, W. (2012). 20 Jahre empirisch-quantitative Grundschulforschung: Rückblick und Ausblick. In F. Hellmich, S. Förster & F. Hoya (Hrsg.), *Bedingungen des Lehrens und Lernens in der Grundschule. Bilanz und Perspektiven* (S. 19–38). Wiesbaden: Springer.
- Emmer, E. T., Evertson, C. M. & Anderson, L. L. M. (1980). Effective classroom management at the beginning of school year. *The Elementary School Journal*, 89(5), 219–231.
- Evertson, C. M. (1989). Classroom organization and management. In M. C. Reynolds (Ed.): *Knowledge base for the beginning teacher* (pp. 59–70). Oxford: Pergamon.
- Evertson, C. M. & Emmer, E. T. (2009). *Classroom Management for Elementary Teachers*. Upper Saddle River, NJ: Pearson Education, Inc.
- Evertson, C. M., Emmer, E. T., Sanford, J. P. & Clements, B. S. (1983). Improving classroom management: An experiment in elementary school classrooms. *The Elementary School Journal*, 84(2), 173–188.
- Evertson, C. M., Emmer, E. T. & Worsham, M. E. (2006). *Classroom Management for Elementary Teachers*. Boston: Allyn & Bacon.
- Evertson, C. M. & Weinstein, C. S. (Eds.) (2006). *Handbook of Classroom Management. Research, Practice, and Contemporary Issues*. Mahwah, N.J: Lawrence Erlbaum Associates.
- Ewerhardy, A. (2010). *Zusammenhänge zwischen Verständnisorientierung von naturwissenschaftsbezogenem Sachunterricht und Fortschritten im Verständnis naturwissenschaftlicher Konzepte bei Lernenden der Grundschule*. Dissertation, Westfälische Wilhelms-Universität Münster. Online verfügbar: [http://miami.uni-muenster.de/servlets/DerivateServlet/Derivate-6142/diss\\_ewerhardy.pdf](http://miami.uni-muenster.de/servlets/DerivateServlet/Derivate-6142/diss_ewerhardy.pdf) (Zugriff: 31.01.2013)
- Ewerhardy, A., Kleickmann, T. & Möller, K. (2009). Zusammenhänge zwischen Verständnisorientierung von naturwissenschaftsbezogenem Sachunterricht und Fortschritten im Verständnis naturwissenschaftlicher Konzepte bei Grundschulern. In D. Höttecke (Hrsg.), *Chemie- und Physikdidaktik für die Lehramtsausbildung. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Schwäbisch Gmünd 2008* (S. 398–400). Münster: Lit-Verlag.
- Eye, A. v. (2006). An alternative to Cohen's Kappa. *European Psychologist*, 11(1), 12–24.
- Faust, G., Lipowsky, F. & Gleich, A.-K. (2011). Unterrichtsqualität in der Grundschule. Kognitive Aktivierung in der PERLE-Videostudie „Sprache“. *Die Grundschulzeitschrift*, 245.246, 48–52.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014). Student ratings of teaching quality in Primary School: Dimensions and Prediction of Student Outcomes. *Learning and Instruction*, 29(1), 1–9.
- Fend, H. (1977). *Schulklima. Soziale Einflußprozesse in der Schule*. Weinheim: Beltz.
- Fend, H. (1997). *Der Umgang mit Schule in der Adoleszenz*. Bern: Huber.
- Fend, H. (1998, 2001). *Qualität im Bildungswesen. Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistung*. Weinheim: Juventa.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117–140.
- Fetzer, M. (2007). *Interaktion am Werk. Eine Interaktionstheorie fachlichen Lernens, entwickelt am Beispiel von Schreibanlässen im Mathematikunterricht der Grundschule*. Bad Heilbrunn: Klinkhardt.
- Föh, M.-J. (2005). Theoretische und empirische Untersuchung der mündlichen Kommunikation und Kooperation bei der gemeinsamen Arbeit am Computer im Deutschunterricht. In J. Stückrath & R. Strobel (Hrsg.), *Deutschunterricht empirisch. Beiträge zur Überprüfbarkeit von Lernfortschritten im Sprach-, Literatur- und Medienunterricht* (S. 257–273). Baltmannsweiler: Hohengehren.

- 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 of Technology.
- Fraser, B. J., Walberg, H. J., Welch, W. W. & Hattie, J. (1987). Syntheses of educational productivity research. *International Journal of Educational Research*, 11, 145–252.
- Fromm, S. (2008). Kapitel 15. Multiple lineare Regressionsanalyse. In N. Baur & S. Fromm (Hrsg.), *Datenanalyse in SPSS für Fortgeschrittene. Ein Arbeitsbuch* (S. 345–369). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Fudge, D. L., Skinner, C. H., Williams, J. L., Cowden, D., Clark, J. & Bliss, S. L. (2008). Increasing on-task behavior in every student in second-grade classroom during transitions: Validating the color wheel system. *Journal of School Psychology*, 46, 575–592.
- Gabriel, K., Kastens, C., Poloczek, S., Schoreit, E. & Lipowsky, F. (2010). Entwicklung des mathematischen Selbstkonzepts – Der Einfluss des Klassenkontextes. *Zeitschrift für Grundschulforschung*, 3(1), 65–82.
- Gabriel, K. & Lipowsky, F. (2013a). Hoch inferentes Rating: Klassenführung in Deutsch, Kunst und Mathematik. In M. Lotz, F. Lipowsky & G. Faust (Hrsg.), *Technischer Bericht zu den PERLE-Videostudien*. Materialien zur Bildungsforschung. Bd. 23/3 (S. 145–168). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF).
- Gabriel, K. & Lipowsky, F. (2013b). Hoch inferentes Rating: Unterrichtsklima in Deutsch, Kunst und Mathematik. In M. Lotz, F. Lipowsky & G. Faust (Hrsg.), *Technischer Bericht zu den PERLE-Videostudien*. Materialien zur Bildungsforschung. Bd. 23/3 (S. 169–190). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF).
- Gabriel, K., Mösko, E. & Lipowsky, F. (2010). Selbstkonzeptentwicklung von Jungen und Mädchen im Anfangsunterricht – Ergebnisse aus der PERLE-Studie. In F. Hellmich (Hrsg.), *Selbstkonzepte im Grundschulalter. Modelle, empirische Ergebnisse, pädagogische Konsequenzen* (S. 133–158). Stuttgart: Kohlhammer.
- Gautschi, P., Moser, D. V., Reusser, K. & Wiher, P. (Hrsg.) (2007). *Geschichtsunterricht heute. Eine empirische Analyse ausgewählter Aspekte*. Bern: h.e.p. Verlag.
- Geiser, C. (2010). *Datenanalyse mit MPLUS. Eine anwendungsorientierte Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Göbel, K. (2007). *Qualität im interkulturellen Englischunterricht. Eine Videostudie*. Münster: Waxmann.
- Gold, B. & Holodynski, M. (2011). Klassenführung. In E. Kiel & K. Zierer (Hrsg.), *Unterrichtsgestaltung als Gegenstand der Wissenschaft* (S. 133–151). Baltmannsweiler: Schneider Verlag Hohengehren.
- Good, T. L. (1979). Teacher effectiveness in the elementary school. *Journal of Teacher Education*, 30(2), 52–64.
- Götz, T., Lohrmann, K., Ganser, B. & Haag, L. (2005). Einsatz von Unterrichtsmethoden – Konstanz oder Wandel? *Empirische Pädagogik*, 19, 342–360.
- Gräsel, C. & Mandl, H. (2009). Qualitätskriterien von Unterricht: Ein zentrales Thema der Unterrichts- und Lehr-Lern-Forschung. In H. J. Apel & W. Sacher (Hrsg.), *Studienbuch Schulpädagogik* (S. 241–259). Heilbrunn: Klinkhardt.
- Greb, K., Faust, G. & Lipowsky, F. (2007). Projekt PERLE. Persönlichkeits- und Lernentwicklung von Grundschulkindern. *Diskurs Kindheits- und Jugendforschung*, 2, 100–104.
- Greb, K., Lipowsky, F. & Faust, G. (2009a). Nina und Michael, Miró und ein Nussknacker! Persönlichkeits- und Lernentwicklung von Grundschulkindern. *Die Grundschulzeitschrift*, 23, 18–21.

- Greb, K., Lipowsky, F. & Faust, G. (2009b). Einleitung. In F. Lipowsky, G. Faust & K. Greb (Hrsg.), *Dokumentation der Erhebungsinstrumente des Projekts „Persönlichkeits- und Lernentwicklung von Grundschulkindern“ (PERLE) – Teil 1. Materialien zur Bildungsforschung*. Bd. 23/1 (S. 5–19). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF).
- Grell, J. (2002). Direktes Unterrichten. Ein umstrittenes Unterrichtsmodell. In J. Wiechmann (Hrsg.), *Zwölf Unterrichtsmethoden. Vielfalt für die Praxis* (S. 35–49). Weinheim: Beltz.
- Gresser, A., Pohl, K., Corvacho del Toro, I., Greb, K. & Faust, G. (2009). Modul 2: Deutsch. In F. Lipowsky, G. Faust & K. Greb (Hrsg.), *Dokumentation der Erhebungsinstrumente des Projekts PERLE – „Persönlichkeits- und Lernentwicklung von Grundschulkindern“*. Materialien zur Bildungsforschung. Bd. 23/1 (S. 34–40). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF).
- Gruehn, S. (1995). Vereinbarkeit kognitiver und nichtkognitiver Ziele im Unterricht. *Zeitschrift für Pädagogik*, 41(4), 531–553.
- Gruehn, S. (1996). Lassen sich Leistungs- und Motivationsförderung im Unterricht vereinbaren? In J. Baumert, S. Gruehn, S. Heyn, O. Köller & K. Schnabel (Hrsg.), *Bildungsverläufe und psychosoziale Entwicklung im Jugendalter. Dokumentation* (S. 35–42). Berlin: Max-Planck-Institut für Bildungsforschung.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen. Schüler als Quellen der Unterrichtsbeschreibung*. Münster: Waxmann.
- Gudjons, H. (2003). *Frontalunterricht – neu entdeckt. Integration in offene Unterrichtsformen*. Bad Heilbrunn: Klinkhardt.
- Guay, F., Marsh, H. W. & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their casual ordering. *Journal of Educational Psychology*, 95(1), 124–136.
- Güttler, P. O. (2003). *Sozialpsychologie. Soziale Einstellungen, Vorurteile, Einstellungsänderungen*. München: Oldenbourg.
- Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, 19(2), 149–161.
- Hadeler, S. (2013). *Fordern und Fördern – Leistungsanforderungen und Differenzierung in der Lern- und Förderumwelt privater Grundschulen. Eine qualitative Untersuchung*. Dissertation, Universität Kassel.
- Hamre, B. K., Mashburn, A. J., Pianta, R. C. & Downer, J. T. (2006). *Validation of 3-factor model for classroom quality across preschool to fifth grade*. Manuskript submitted for publication.
- Hamre, B. K. & Pianta, R. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, 72, 625–638.
- Hamre, B. K. & Pianta, R. C. (2007). Learning opportunities in preschool and early elementary classrooms. In R. C. Pianta, M. J. Cox & K. L. Snow (Eds.), *School readiness and the transition to kindergarten in the era of accountability* (pp. 49–83). Baltimore: Paul H. Brookes Publishing.
- Hamre, B. K., Pianta, R. C., Mashburn, A. J. & Downer, J. T. (2007). *Building a Science of Classrooms: Application of the CLASS Framework in Over 4000 U.S. Early Childhood and Elementary Classrooms*. New York: Foundation for Child Development. Online verfügbar: <http://fed-us.org/sites/default/files/BuildingAScienceOfClassroomsPiantaHamre.pdf> (Zugriff: 31.01.2013).
- Hanke, P. (2007). Förderung der Bildungsprozesse von Kindern in der Schuleingangsphase – Konzeption des FiS-Projektes. In K. Möller, C. Beinbrech, P. Hanke, A. K. Hein, T. Schages & R. Kleickmann (Hrsg.), *Qualität von Grundschulunterricht. Entwickeln, erfassen und bewerten* (S. 187–190). Wiesbaden: VS Verlag für Sozialwissenschaften.

- Harnischfeger, A. & Wiley, D. E. (1977). Kernkonzepte des Schullernens. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 9(3), 207–228.
- Hartig, J. & Kühnbach, O. (2006). Schätzung von Verfahren mit „plausible values“ in mehrdimensionalen Rasch-Modellen. In A. Ittel & H. Merkens (Hrsg.), *Veränderungsmessung und Längsschnittstudien in der empirischen Erziehungswissenschaft* (S. 27–44). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hartinger, A., Kleickmann, T. & Hawelka, B. (2006). Der Einfluss von Lehrervorstellungen zum Lernen und Lehren auf die Gestaltung des Unterrichts und auf motivationale Schülermerkmale. *Zeitschrift für Erziehungswissenschaft*, 9(1), 110–126.
- Hascher, T. (2004). *Wohlbefinden in der Schule*. Münster: Waxmann.
- Hascher, T. & Oser, F. (1996). Lernen, Motivation und Sozialklima. In C. Spiel (Hrsg.), *Motivation und Lernen aus der Perspektive lebenslanger Entwicklung* (S. 175–184). Münster: Waxmann.
- Hattie, J. A. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Havers, N. (1998). Disziplinschwierigkeiten vermeiden. Gute Organisation des Unterrichtsablaufs. *Grundschule*, 4, 8–9.
- Hein, A. K., Eckerth, M. & Hanke, P. (2010). Grundschul- und fachdidaktische Merkmale jahrgangsbezogenen und jahrgangsübergreifenden Unterrichts in der Schuleingangsphase. In K.-H. Arnold, K. Hauenschild, B. Schmidt & B. Ziegenmeyer (Hrsg.), *Zwischen Fachdidaktik und Stufendidaktik. Perspektiven für die Grundschulpädagogik* (S. 51–54). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Heinze, A. (2004). Zum Umgang mit Fehlern im Unterrichtsgespräch der Sekundarstufe. Theoretische Grundlegung, Methode und Ergebnisse einer Videostudie. *Journal für Mathematik-Didaktik*, 25(3/4), 221–244.
- Heinze, A. (2005). Mistake-handling activities in the mathematics classroom. In H. L. Chick & J. L. Vincent (Eds.), *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education* (pp. 105–112). Melbourne: PME. Online verfügbar: <http://www.emis.de/proceedings/PME29/PME29RRPapers/PME29Vol3Heinze.pdf> (Zugriff: 31.01.2013)
- Helmke, A. (1988). Leistungssteigerung und Ausgleich von Leistungsunterschieden in Schulklassen: unvereinbare Ziele? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 20(1), 45–76.
- Helmke, A. (1992). Unterrichtsqualität und Unterrichtseffekte -- Ergebnisse der Münchner Studie. *Der Mathematikunterricht*, 38, 40–58.
- Helmke, A. (1998). Vom Optimisten zum Realisten? Zur Entwicklung des Fähigkeitsselbstkonzepts vom Kindergarten bis zur 6. Klassenstufe. In F. E. Weinert (Hrsg.), *Entwicklung im Kindesalter* (S. 117–132). Weinheim: Beltz.
- Helmke, A. (1999). Direkte Instruktion – effektiver Unterricht? *Bildung Real*, 38, 59–72.
- Helmke, A. (2006). Was wissen wir über guten Unterricht? Über die Notwendigkeit einer Rückbesinnung auf den Unterricht als dem „Kerngeschäft“ von Schule. *Pädagogik*, 2, 42–45.
- Helmke, A. (2007). Aktive Lernzeit optimieren. Was wissen wir über effiziente Klassenführung? *Pädagogik*, 5, 44–48.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Kallmeyer.
- Helmke, A., Helmke, T., Heyne, N., Hosenfeld, A., Hosenfeld, I., Schrader, F.-W. et al. (2008). Zeitznutzung im Grundschulunterricht: Ergebnisse der Unterrichtsstudie „VERA – Gute Unterrichtspraxis“. *Zeitschrift für Grundschulforschung*, 1(1), 23–36.

- 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, B. Schober, P. Wagner & R. Reimann (Hrsg.), *Bildungspsychologie* (S. 101–105). Göttingen: Hogrefe.
- Helmke, A., Helmke, T., Kleinbub, I., Nordheider, I., Schrader, F.-W. & Wagner, W. (2007). Die DESI-Videostudie. Unterrichtstranskripte für die Lehrerausbildung nutzen. *Der fremdsprachliche Unterricht Englisch*, 41(90), 37–45.
- Helmke, A., Helmke, T. & Schrader, F.-W. (2007). Unterrichtsqualität: Brennpunkte und Perspektiven der Forschung. In K.-H. Arnold (Hrsg.), *Unterrichtsqualität und Fachdidaktik* (S. 51–72). Bad Heilbrunn: Klinkhardt.
- Helmke, A., Helmke, T., Schrader, F.-W. & Wagner, W. (2007). *Der Ratingbogen der DESI-Videostudie*. Koblenz-Landau: Universität Koblenz-Landau.
- Helmke, A., Helmke, T., Schrader, F.-W., Wagner, W., Klieme, E., Nold, G. & Schröder, K. (2008). Wirksamkeit des Englischunterrichts. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 382–397). Weinheim: Beltz.
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2002). Unterricht, Mathematikleistung und Lernmotivation. In A. Helmke & R. S. Jäger (Hrsg.), *Das Projekt MARKUS. Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext* (S. 415–480). Landau: Verlag Empirische Pädagogik.
- Helmke, A. & Klieme, E. (2008). Unterricht und Entwicklung sprachlicher Kompetenzen. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 301–312). Weinheim: Beltz.
- Helmke, A. & Renkl, A. (1992). Das Münchner Aufmerksamkeitsinventar (MAI): Ein Instrument zur systematischen Verhaltensbeobachtung der Schüleraufmerksamkeit im Unterricht. *Diagnostica*, 38(2), 130–141.
- Helmke, A. & Renkl, A. (1993). Unaufmerksamkeit in Grundschulklassen: Problem der Klasse oder des Lehrers? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 25(3), 185–205.
- 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.
- Helmke, A. & Schrader, F.-W. (1990). Zur Kompatibilität kognitiver, affektiver und motivationaler Zielkriterien des Schulunterrichts – Clusteranalytische Studien. In M. Knopf & W. Schneider (Hrsg.), *Entwicklung. Allgemeine Verläufe, individuelle Unterschiede, pädagogische Konsequenzen* (S. 180–200). Göttingen: Hogrefe.
- Helmke, A. & Schrader, F.-W. (1993). Was macht erfolgreichen Unterricht aus? Ergebnisse der Münchner Studie über Schulleistungen. *Praxis Schule* 5–10, 1, 11–13.
- Helmke, A. & Schrader, F.-W. (1997). Unterrichtsbeurteilungen durch externe Beobachter. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 510–518). Weinheim: Beltz.
- Helmke, A. & Schrader, F.-W. (2008). Merkmale der Unterrichtsqualität: Potenzial, Reichweite und Grenzen. *Seminar – Lehrerbildung und Schule*, 14(3), 17–47.
- Helmke, A. & Weinert, F. E. (1997a). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert, N. Birbaumer & C. F. Graumann (Hrsg.), *Psychologie des Unterrichts und der Schule* (S. 71–176). Göttingen: Hogrefe.
- Helmke, A. & Weinert, F. E. (1997b). Unterrichtsqualität und Leistungsentwicklung: Ergebnisse aus dem SCHOLASTIK-Projekt. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 241–251). Weinheim: Beltz.

- Helmke, T., Helmke, A., Schrader, F.-W., Wagner, W., Nold, G. & Schröder, K. (2008). Die Videostudie des Englischunterrichts. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 345–364). Weinheim: Beltz.
- Hennemann, T. & Hillenbrand, C. (2010). Klassenführung – Classroom Management. In B. Hartke, K. Koch & K. Diehl (Hrsg.), *Förderung in der schulischen Eingangsstufe* (S. 255–279). Stuttgart: Kohlhammer.
- Hidi, S., Renninger, K. A. & Krapp, A. (2004). Interest, a motivational variable that combines affective and cognitive functioning. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion, and cognition. Integrative perspectives on intellectual functioning and development* (pp. 89–115). Mahwah: Erlbaum.
- Hiebert, J., Gallimore, R., Garnier, H., Giwin, K. B., Hollingworth, H., Jacobs, J. et al. (2003). *Teaching mathematics in seven countries. Results from the TIMSS 1999 video study*. Washington, DC: U.S. Government Printing Office. Online verfügbar: <http://nces.ed.gov/pubs2003/2003013.pdf> (Zugriff: 31.01.2013).
- Hochweber, J. (2010). *Was erfassen Mathematiknoten?* Münster: Waxmann.
- Hodel, J. & Waldis, M. (2007). Sichtstrukturen im Geschichtsunterricht – die Ergebnisse der Videoanalyse. In P. Gautschi, D. V. Moser, K. Reusser & P. Wiher (Hrsg.), *Geschichtsunterricht heute. Eine empirische Analyse ausgewählter Aspekte* (S. 91–142). Bern: h.e.p.
- Hodel, J., Waldis, M., Gautschi, P. & Reusser, K. (2006). Die Erfassung von Sichtstrukturen und Qualitätsmerkmalen im Geschichtsunterricht. Methodologische Überlegungen am Beispiel der Videostudie „Geschichte und Politik im Unterricht“. In H. G. Arndt & M. Sauer (Hrsg.), *Geschichtsdidaktik empirisch. Untersuchungen zum historischen Denken und Lernen* (S. 155–188). Berlin: Lit-Verlag.
- Hofer, M. (1985). Zu den Wirkungen von Lob und Tadel. *Bildung und Erziehung*, 38(4), 415–427.
- Hofman, D. A. & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: implications for research in organizations. *Journal of Management*, 24(5), 634–664.
- Hofmann, N. (2007). *Unterrichtsexpertise und Rechtschreibleistungen – Eine empirische Untersuchung in Heidelberger Grundschulen*. Unveröffentlichte Dissertation, Pädagogische Hochschule Heidelberg. Online verfügbar: [http://opus.bs-z-bw.de/phhd/volltexte/2008/7502/pdf/Unterrichtsexpertise\\_und\\_Rechtschreibleistung.pdf](http://opus.bs-z-bw.de/phhd/volltexte/2008/7502/pdf/Unterrichtsexpertise_und_Rechtschreibleistung.pdf) (Zugriff: 31.01.2013).
- Holodynski, M. (1998). Klassenführung. Eine notwendige Aufgabe erfolgreichen Unterrichtens. *Grundschule*, 9, 22–24.
- Holtappels, H. G. (1985). Schülerprobleme und abweichendes Verhalten aus der Schülerperspektive. *Zeitschrift für Sozialisationsforschung und Erziehungssoziologie*, 5, 291–223.
- Homburg, C. & Baumgartner, H. (1995). Beurteilung von Kausalmodellen: Be-standsaufnahme und Anwendungsempfehlungen. *Marketing – Zeitschrift für Forschung und Praxis*, 17(3), 162–176.
- Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods*, 5(1), 64–86.
- Hoyt, W. T. & Kerns, M.-D. (1999). Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods*, 4(4), 403–424.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55.
- Hugener, I. (2006). Überblick über die Beobachtungsinstrumente. In E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“*. Teil 3: *Videoanalysen*. Materialien zur Bildungsforschung. Bd. 15 (S. 41–54). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF). Online verfügbar: [http://www2.dipf.de/gfpf/publikationen/MatBild\\_Bd15.pdf](http://www2.dipf.de/gfpf/publikationen/MatBild_Bd15.pdf) (Zugriff: 31.01.2013).

- Hugener, I. & Krammer, K. (2010). Differenzierende Maßnahmen zur Individualisierung des Unterrichts. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität. Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 91–106). Münster: Waxmann.
- Hugener, I., Pauli, C. & Reusser, K. (2006). Videoanalysen. In E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“ (Teil 3)*. Frankfurt a.M.: GPPF/DIPF.
- Hugener, I., Rakoczy, K., Pauli, C. & Reusser, K. (2006). Videobasierte Unterrichtsforschung: Integration verschiedener Methoden der Videoanalyse für eine differenzierte Sicht auf Lehr-Lernprozesse. In S. Rahm, I. Mammes & M. Schratz (Hrsg.), *Schulpädagogische Forschung, Unterrichtsforschung, Perspektiven innovativer Ansätze* (S. 41–53). Innsbruck: Studien Verlag.
- Hughes, J. N., Cavell, T. A. & Jackson, T. (1999). Influence of the teacher-student relationship on childhood conduct problems: A prospective study. *Journal of Clinical Child Psychology*, 28(2), 173–184.
- Jacobs, J. K., Kawanaka, T. & Stigler, J. (1999). Integrating qualitative and quantitative approaches to the analysis of video data on classroom teaching. *International Journal of Educational Research*, 31, 717–724.
- Jank, W. & Meyer, H. (2009). *Didaktische Modelle*. Berlin: Cornelsen Verlag.
- Jerome, E. M., Hamre, B. K. & Pianta, R. C. (2009). Teacher-child relationships from Kindergarten to sixth grade: Early childhood predictors of teacher-perceived conflict and closeness. *Social Development*, 18(4), 915–945.
- Jerusalem, M. (1993). *Die Entwicklung von Selbstkonzepten und ihre Bedeutung für Motivationsprozesse im Lern- und Leistungsbereich*. Antrittsvorlesung, Humboldt- Universität Berlin. Online verfügbar: <http://edoc.hu-berlin.de/humboldt-vl/jerusalem-matthias/PDF/Jerusalem.pdf> (Zugriff: 31.01.2013).
- Jerusalem, M. & Schwarzer, R. (1991). Entwicklung des Selbstkonzepts in verschiedenen Lernumwelten. In R. Pekrun & H. Fend (Hrsg.), *Schule und Persönlichkeitsentwicklung* (S. 115–128). Stuttgart: Enke Verlag.
- Jones, V. (1996). Classroom management. In J. P. Sikula, T. J. Buttery & E. Guyton (Eds.), *Handbook of research on teacher education. A project of the Association of Teacher Educators* (pp. 503–521). New York: Macmillan Library Reference USA.
- Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 294–316). Newbury Park: Sage Publications.
- Kammermeyer, G. & Martschinke, S. (2003). Schulleistung und Fähigkeitsselbstbild im Anfangsunterricht – Ergebnisse aus dem KILIA-Projekt. *Empirische Pädagogik*, 17(4), 486–503.
- Kammermeyer, G. & Martschinke, S. (2004). KILIA – Selbstkonzept und Leistungsentwicklung im Anfangsunterricht. In G. Faust, M. Götz, H. Hacker & H.-G. Roßbach (Hrsg.), *Anschlussfähige Bildungsprozesse im Elementar- und Primarbereich* (S. 204–217). Bad Heilbrunn: Klinkhardt.
- Kammermeyer, G. & Martschinke, S. (2006). Selbstkonzept- und Leistungsentwicklung in der Grundschule – Ergebnisse aus der KILIA-Studie. *Empirische Pädagogik*, 20(3), 245–259.
- Kammermeyer, G. & Martschinke, S. (2009). Qualität im Anfangsunterricht – Ergebnisse der KILIA-Studie. *Unterrichtswissenschaft*, 37(1), 35–54.
- Karst, K. & Lipowsky, F. (2013). Leistungsentwicklung im Fach Mathematik und deren Determinanten. In F. Lipowsky, G. Faust & C. Kastens (Hrsg.), *Persönlichkeits- und Lernentwicklung an staatlichen und privaten Grundschulen. Ergebnisse der PERLE-Studie zu den ersten beiden Schuljahren* (S. 51–75). Münster: Waxmann.

- Kassner, D. (2002). *Humor im Unterricht. Bedeutung – Einfluss – Wirkungen. Können schulische Leistungen und berufliche Qualifikationen durch pädagogischen Humor verbessert werden?* Baltmannsweiler: Schneider-Verlag Hohengehren.
- Kastens, C. & Gabriel, K. (in Vorb.) *Academic Self-Concept Development in Elementary School: A Domain-Specific Perspective.*
- Kastens, C., Gabriel, K. & Lipowsky, F. (2013). Selbstkonzeptentwicklung im Anfangsunterricht. In F. Lipowsky, F. Faust & C. Kastens (Hrsg.), *Persönlichkeits- und Lernentwicklung an staatlichen und privaten Grundschulen. Ergebnisse der PERLE-Studie zu den ersten beiden Schuljahren* (S. 99–128). Münster: Waxmann.
- Kaufmann, L., Nuerk, H.-C., Graf, M., Krinzinger, H., Delazer M. & Willmes, K. (2009). *TEDI-MATH: Test zur Erfassung numerisch-rechnerischer Fertigkeiten vom Kindergarten bis zur 3. Klasse.* Zürich: Hans-Huber-Verlag.
- Kleickmann, T., Vehmeyer, J. & Möller, K. (2010). Zusammenhänge zwischen Lehrervorstellungen und kognitivem Strukturieren im Unterricht am Beispiel von Scaffolding-Maßnahmen. *Unterrichtswissenschaft*, 38(3), 210–228.
- Kleinbub, I. D. (2010). *Unterrichtsqualität im Leseunterricht. Eine videobasierte Analyse in vierten Klassen.* Trier: Wissenschaftlicher Verlag.
- Klicpera, C. & Gasteiger-Klicpera, B. (1993). *Lesen und Schreiben – Entwicklung und Schwierigkeiten. Die Wiener Längsschnittuntersuchungen über die Entwicklung, den Verlauf und die Ursachen von Lese- und Schreibschwierigkeiten in der Pflichtschulzeit.* Bern: Huber.
- Klieme, E. (2002). Was ist guter Unterricht? Ergebnisse der TIMSS-Videostudie im Fach Mathematik. In W. Bergsdorf, J. Court, M. Eckert & H. Hoffmeister (Hrsg.), *Herausforderungen der Bildungsgesellschaft. 4. Ringvorlesung der Universität Erfurt* (S. 89–113). Weimar: Rhino Verlag.
- Klieme, E. (2006). Empirische Unterrichtsforschung: Aktuelle Entwicklungen, theoretische Grundlagen und fachspezifische Befunde. *Zeitschrift für Pädagogik*, 52(6), 765–773.
- Klieme, E. (2010). *Standards der Unterrichtsqualität – Kann es das geben?* Vortrag am 20. November 2010 an der Universität Zürich.
- Klieme, E., Jude, N., Rauch, D., Ehlers, H., Helmke, A., Eichler, W. et al. (2008). Alltagspraxis, Qualität und Wirksamkeit des Deutschunterrichts. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 319–344). Weinheim: Beltz. Online verfügbar: [http://www.pedocs.de/volltexte/2010/3149/pdf/978\\_3\\_407\\_25491\\_7\\_1A\\_D\\_A.pdf](http://www.pedocs.de/volltexte/2010/3149/pdf/978_3_407_25491_7_1A_D_A.pdf) (Zugriff: 31.01.2013).
- Klieme, E., Lipowsky, F., Rakoczy, K. & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts „Pythagoras“. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms BIQUA* (S. 127–146). Münster: Waxmann.
- Klieme, E., Pauli, C. & Reusser, K. (Hrsg.) (2006). *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“.* Teil 3: *Videoanalysen.* Materialien zur Bildungsforschung. Bd. 15. Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GPF). Online verfügbar: [http://www2.dipf.de/gfpf/publikationen/MatBild\\_Bd15.pdf](http://www2.dipf.de/gfpf/publikationen/MatBild_Bd15.pdf) (Zugriff: 31.01.2013).
- Klieme, E. & Rakoczy, K. (2003). Unterrichtsqualität aus Schülerperspektive: Kulturspezifische Profile, regionale Unterschiede und Zusammenhänge mit Effekten von Unterricht. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000 – Ein differenzierter Blick auf Länder der Bundesrepublik Deutschland* (S. 333–359). Opladen: Leske & Budrich.

- Klieme, E. & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54(2), 222–237.
- Klieme, E. & Reusser K. (2003). Unterrichtsqualität und mathematisches Verständnis im internationalen Vergleich – Ein Forschungsprojekt und erste Schritte zur Realisierung. *Unterrichtswissenschaft*, 31(3), 194–205.
- Klieme, E., Schümer, G. & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: Aufgabekultur und Unterrichtsgestaltung. In Bundesministerium für Bildung und Forschung (Hrsg.), *TIMSS – Impulse für Schule und Unterricht. Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente* (S. 43–58). Bonn: BMBF.
- Klippert, H. (2008). *Besser lernen. Kompetenzvermittlung und Schüleraktivierung im Schulalltag*. Stuttgart: Klett.
- Knauf, T. (2001). *Einführung in die Grundschuldidaktik. Lernen, Entwicklungsförderung und Erfahrungswelten in der Primarstufe*. Stuttgart: Kohlhammer.
- Kobarg, M. & Seidel, T. (2003). Prozessorientierte Lernbegleitung im Physikunterricht. In T. Seidel, M. Prenzel, R. Duit & M. Lehrke (Hrsg.), *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“* (S. 151–200). Kiel: IPN.
- Kounin, J. S. (1970). *Discipline and group management in classrooms*. New York: Holt, Rinehart & Winston.
- Kounin, J. S. (1976). *Techniken der Klassenführung*. 1. Aufl. Stuttgart: Klett.
- Kounin, J. S. (2006). *Techniken der Klassenführung*. 2. Aufl. Münster: Waxmann.
- Krajewski, K., Küspert, P. & Schneider, W. (2002). *Deutscher Mathematiktest für erste Klassen (DEMAT 1+)*. Göttingen: Hogrefe.
- Krajewski, K., Liehm, S. & Schneider, W. (2004). *Deutscher Mathematiktest für zweite Klassen (DEMAT 2+)*. Göttingen: Hogrefe.
- Kramis, J. (1990). Bedeutsamkeit, Effizienz, Lernklima. Grundlegende Gütekriterien für Unterricht und Didaktische Prinzipien. *Beiträge zur Lehrerbildung*, 8(3), 279–296.
- Krammer, K. (2009). *Individuelle Lernunterstützung in Schülerarbeitsphasen. Eine videobasierte Analyse des Unterstützungsverhaltens von Lehrpersonen im Mathematikunterricht*. Münster: Waxmann.
- Kucharz, D. (2012). *Entwicklung der Unterrichtskompetenz von Grundschullehrer/innen im Berufseinstieg. Die internationale Vergleichsstudie ALPHA*. Vortrag auf der 1. Frankfurter Tagung zu Videoanalysen in der Unterrichts- und Bildungsforschung. Quantitativ & Qualitativ, Frankfurt am Main, 22.–23. Februar 2012.
- Kucharz, D., Küster, O., Müller, K. & Schnebel, S. (2009). *Wirkungen offener Lehr-Lernformen (WOLFF) – Abschlussbericht der Pilotstudie*.
- Kucharz, D., Schnebel, S. & Helming, V. (2011). *Projektbericht. Individuelle Lernunterstützung in Phasen des offenen Unterrichts. Eine videobasierte Beschreibung von lernbegleitenden Maßnahmen innerhalb Lehrer-Schüler-Interaktionen*. Weingarten: Pädagogische Hochschule.
- Kuliga, M. G. (1985). Selbstkonzepte Jugendlicher in zwei unterschiedlichen Lernumwelten. In H. A. Stiksrud (Hrsg.), *Adoleszenz und Postadoleszenz. Beitrag zur angewandten Jugendpsychologie* (S. 82–95). Eschborn: Fachbuchhandlung für Psychologie.
- Künsting, J., Post, S., Greb, K., Faust, G. & Lipowsky, F. (2010). Leistungsheterogenität im mathematischen Anfangsunterricht – Ein Risiko für die Leistungsentwicklung? *Zeitschrift für Grundschulforschung*, 3(1), 46–64.
- Kunter, M. (2005). *Multiple Ziele im Mathematikunterricht*. Münster: Waxmann.

- Kunter, M., Baumert, J. & Köller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction*, 17, 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., Dubberke, T., Baumert, J., Blum, W., Brunner, M., Jordan, A. et al. (2006). Mathematikunterricht in den PISA-Klassen 2004: Rahmenbedingungen, Formen und Lehr-Lernprozesse. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 161–194). Münster: Waxmann.
- Kunter, M., Klusmann, U., Dubberke, T., Baumert, J., Blum, W., Brunner, M. et al. (2007). Linking aspects of teacher competence to their instruction. Results from the COACTIV project. In M. Prenzel (Ed.), *Studies on the educational quality of schools. The final report on the DFG priority program* (pp. 39–59). Münster: Waxmann.
- Kunter, M. & Voss, T. (2011). Das Modell der Unterrichtsqualität in COACTIV: Eine multikriteriale 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.
- Kuß, A. (2012). *Marktforschung. Grundlagen der Datenerhebung und Datenanalyse*. Wiesbaden: Springer Gabler Verlag.
- La Paro, K. M., Pianta, R. C. & Stuhlman, M. (2004). The Classroom Assessment Scoring System: Findings from the Prekindergarten year. *The Elementary School Journal*, 104(5), 409–426.
- Lange, B., Kuffner, H. & Schwarzer, R. (1983). *Schulangst und Schulverdrossenheit. Eine Längsschnittanalyse von schulischen Sozialisationswirkungen*. Opladen: Westdeutscher Verlag.
- Langer, I. & Schulz von Thun, F. (2007). *Messung komplexer Merkmale in Psychologie und Pädagogik. Ratingverfahren*. Münster: Waxmann.
- Lankes, E.-A., Bos, W., Mohr, I., Plaßmeier, N. & Schwippert, K. (2003). Lehr- und Lernbedingungen in den Teilnehmerländern. In W. Bos., E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 29–67). Münster: Waxmann.
- Lauterbach, C., Gabriel, K. & Lipowsky, F. (2013). Hoch inferentes Rating: Kognitive Aktivierung im Mathematikunterricht. In M. Lotz, F. Lipowsky & G. Faust (Hrsg.), *Technischer Bericht zu den PERLE-Videostudien*. Materialien zur Bildungsforschung. Bd. 23/3 (S. 405–421). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF).
- Leist, S., Töpfer, T., Bardowiecks, S., Pietsch, M. & Tosana, S. (2010). *Handbuch zum Unterrichtsbeobachtungsbogen der Schulinspektion Hamburg* (Version 5). Hamburg: Institut für Bildungsmonitoring Schulinspektion. Online verfügbar: <http://www.schulinspektion.hamburg.de/index.php/file/download/1731> (Zugriff: 31.01.2013)
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. *Zeitschrift für Pädagogik*, 51. Beiheft, 47–70.
- Lipowsky, F. (2007a). Unterrichtsqualität in der Grundschule – Ansätze und Befunde der nationalen und internationalen Forschung. In K. Möller, C. Beinbrech, P. Hanke, A. K. Hein, T. Kleickmann & R. Schages (Hrsg.), *Qualität von Grundschulunterricht. Entwickeln, erfassen und bewerten* (S. 35–49). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Lipowsky, F. (2007b). Was wissen wir über guten Unterricht? Im Fokus: die fachliche Lernentwicklung. In G. Becker, A. Feindt, H. Meyer, M. Rothland, L. Stäudel und E. Terhart (Hrsg.), *Guter Unterricht. Maßstäbe & Merkmale – Wege und Werkzeuge*. Friedrich Jahresheft XXV (S. 26–30). Seelze: Friedrich.

- Lipowsky, F. (2009). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 73–101). Berlin: Springer.
- Lipowsky, F., Faust, G. & Greb, K. (Hrsg.) (2009). *Dokumentation der Erhebungsinstrumente des Projekts „Persönlichkeits- und Lernentwicklung von Grundschulkindern“ (PERLE) – Teil 1. Materialien zur Bildungsforschung*. Bd. 23/1. Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF).
- Lipowsky, F., Faust, G., Kastens, C. & Post, S. (2013). Die PERLE-Studie: Überblick und Hintergründe. In F. Lipowsky, G. Faust & C. Kastens (Hrsg.), *Persönlichkeits- und Lernentwicklung an staatlichen und privaten Grundschulen. Ergebnisse der PERLE-Studie zu den ersten beiden Schuljahren* (S. 9–28). Münster: Waxmann.
- Lipowsky, F., Pauli, C. & Rakoczy, K. (2008). Schülerbeteiligung und Unterrichtsqualität. In M. Gläser-Zikuda & J. Seifried (Hrsg.), *Lehrerexpertise. Analyse und Bedeutung unterrichtlichen Handelns* (S. 67–90). Münster: Waxmann.
- Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E. & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction*, 19, 527–537.
- Lipowsky, F., Rakoczy, K., Pauli, C., Reusser, K. & Klieme, E. (2007). Gleicher Unterricht – Gleiche Chancen für alle? Die Verteilung von Schülerbeiträgen im Klassenunterricht. *Unterrichtswissenschaft*, 35(2), 125–174.
- Lipski, J. (2000). Zur Verlässlichkeit der Angaben von Kindern bei standardisierten Befragungen. In F. Heinzel (Hrsg.), *Methoden der Kindheitsforschung. Ein Überblick über Forschungszugänge zur kindlichen Perspektive* (S. 77–86). Weinheim: Juventa.
- Lotz, M. (2010). Kognitiv aktivierende Leseübungen im Anfangsunterricht der Grundschule. Eine Videostudie. In G. Rupp, J. Boelmann & D. Frickel (Hrsg.), *Aspekte literarischen Lernens. Junge Forschung in der Deutschdidaktik* (S. 145–163). Berlin: LIT-Verlag.
- Lotz, M. (2013a). Die Kodierung der Lektionsdauer. In M. Lotz, F. Lipowsky & G. Faust (Hrsg.), *Technischer Bericht zu den PERLE-Videostudien*. Materialien zur Bildungsforschung, Bd. 23/3 (S. 107–114). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF).
- Lotz, M. (2013b). Die Kodierung der Sozialform. In M. Lotz, F. Lipowsky & G. Faust (Hrsg.), *Technischer Bericht zu den PERLE-Videostudien*. Materialien zur Bildungsforschung Bd. 23/3 (S. 123–142). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF).
- Lotz, M. (2013c). Die Kodierung von Klassenteilung und Teamteaching. In M. Lotz, F. Lipowsky & G. Faust (Hrsg.), *Technischer Bericht zu den PERLE-Videostudien*. Materialien zur Bildungsforschung Bd. 23/3 (S. 115–121). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF).
- Lotz, M. (2013d). *Kognitive Aktivierung im Leseunterricht der Grundschule – Eine Videostudie zur Gestaltung und Qualität von Leseübungen im ersten Schuljahr*. Dissertation, Otto-Friedrich-Universität Bamberg.
- Lotz, M., Berner, N. E. & Gabriel, K. (2013). Auswertung der PERLE-Videostudien und Überblick über die Beobachtungsinstrumente. In M. Lotz, F. Lipowsky & G. Faust (Hrsg.), *Technischer Bericht zu den PERLE-Videostudien*. Materialien zur Bildungsforschung Bd. 23/3. Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF).
- 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: VS Verlag für Sozialwissenschaften.

- Lotz, M. & Corvacho del Toro, I. (2013). Die Videostudie im Fach Deutsch: „Lucy rettet Mama Krokodil“. In M. Lotz, F. Lipowsky & G. Faust (Hrsg.), *Technischer Bericht zu den PERLE-Videostudien*. Materialien zur Bildungsforschung Bd. 23/3 (S. 29–36). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF).
- Lotz, M. & Faust, G. (2010). *Gleicher Lehrer – gleicher Unterricht? Schüler- und lehrerorientierte Sozialformen in den Fächern Deutsch und Mathematik*. Poster präsentiert auf der 19. Jahrestagung der Kommission „Grundschulforschung und Pädagogik der Primarstufe“ der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE, Sektion Schulpädagogik). 29.09.–01.10.2010. Weingarten, 2010.
- Lotz, M., Gabriel, K. & Lipowsky, F. (2013). Niedrig und hoch inferente Verfahren der Unterrichtsbeobachtung – Analysen zu deren gegenseitiger Validierung. *Zeitschrift für Pädagogik*, 2, 357–380.
- Lotz, M., Lipowsky, F. & Faust, G. (Hrsg.) (2013). *Technischer Bericht zu den PERLE-Videostudien*. Materialien zur Bildungsforschung. Bd. 23/3. Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF).
- Lotz, M., Schoreit, E. & Kempfer, I. (2013). Die Entwicklung des Leseverständnisses und der Rechtschreibfähigkeit. In F. Lipowsky, G. Faust & C. Kastens (Hrsg.), *Persönlichkeits- und Lernentwicklung an staatlichen und privaten Grundschulen. Ergebnisse der PERLE-Studie zu den ersten beiden Schuljahren* (S. 77–97). Münster: Waxmann.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. Probleme und Lösungen. *Psychologische Rundschau*, 58(2), 103–117.
- Maas, C. J. M. & Hox, J. J. (2005). Sufficient sample sizes for multilevel modelling. *Methodology*, 1(3), 85–91.
- MacCallum, R. C., Widaman, K. F., Zahng, S. & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84–99.
- Madson, C. H., Becker, W. C. & Thomas, D. R. (1977). Rules, praise, and ignoring: Elements of elementary classroom control. In K. O’Leary & S. O’Leary (Eds.), *Classroom management: The successful use of behavior modification* (pp. 63–84). New York: Pergamon.
- Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*, 82, 623–636.
- Marsh, H. W. & Hau, K.-T. (1996). Assessing goodness of fit: Is parsimony always desirable? *Journal of Experimental Education*, 64(4), 364–390.
- Marsh, H. W. & Shavelson, R. J. (1985). Self-concept: Its multifacted, hierarchical structure. *Educational Psychologist*, 20, 107–123.
- Martschinke, S. (2006). Anfangsunterricht in der Grundschule. In K.-H. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht* (S. 535–543). Bad Heilbrunn: Klinkhardt.
- Martschinke, S., Kammermeyer, G., Frank, A. & Mahrhofer, C. (2002). *Heterogenität im Anfangsunterricht – Welche Voraussetzungen bringen Schulanfänger mit und wie gehen Lehrerinnen damit um?* Friedrich-Alexander-Universität Erlangen-Nürnberg: IFG – Institut für Grundschulforschung (Berichte und Arbeiten aus dem Institut für Grundschulforschung).
- Marzano, R. J. (2003). *Classroom management that works: Research-based strategies for every teacher*. Virginia USA: Association for Supervision and Curriculum Development.
- May, P. (2001). *Lernförderlicher Unterricht. Teil I: Untersuchungen zur Wirksamkeit von Unterricht und Förderunterricht für den schriftsprachlichen Lernerfolg*. Frankfurt am Main: Verlag Peter Lang.
- Mayr, J. (2006). Klassenführung auf der Sekundarstufe II: Strategien und Muster erfolgreichen Lehrerhandelns. *Schweizerische Zeitschrift für Bildungswissenschaften*, 28(2), 227–242.

- McDonald Connor, C., Ponitz, C. C., Phillips, B. M., Travis, Q. M., Glasney, S. & Morrison, F. J. (2010). First graders' literacy and self-regulation gains: The effect of individualizing student instruction. *Journal of School Psychology, 48*(5), 433–455.
- Mehlhorn, G. & Mehlhorn, H.-G. (2003). Kreativitätspädagogik – Entwicklung eines Konzepts in Theorie und Praxis. *Bildung und Erziehung, 56*(1), 23–45.
- Mehlhorn, H.-G. (2001). Begabungsentwicklung durch Kreativitätsförderung in Kindergarten und Schule. *Materialien des Forums Bildung, 7*, 22–44.
- Meyer, H. (2004). *Was ist guter Unterricht?* Berlin: Cornelsen.
- Meyer, L., Seidel, T. & Prenzel, M. (2006). Wenn Lernsituationen zu Leistungssituationen werden: Untersuchung zur Fehlerkultur in einer Videostudie. *Schweizerische Zeitschrift für Bildungswissenschaften, 28*, 21–42.
- Möller, K., Jonen, A., Hardy, I. & Stern, E. (2002). Die Förderung von naturwissenschaftlichem Verständnis bei Grundschulkindern durch Strukturierung der Lernumgebung. *Zeitschrift für Pädagogik, 45. Beiheft*, 176–191.
- Moos, R. H. (1979). *Evaluating educational environments*. San Francisco: Jossey-Bass.
- Moos, R. H. & Trickett, E. (1974). *Classroom Environment Scale Manual*. Palo Alto, California: Consulting Psychologists Press.
- Moosbrugger, H. & Hartig, J. (2002). Factor analysis in personality research: Some artifacts and their consequences for psychological assessment. *Psychologische Beiträge, 44*, 136–158.
- Moosbrugger, H. & Schermelleh-Engel, K. (2006). Exploratorische (EFA) und Konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 307–324). Berlin: Studien Verlag.
- Moser, U., Berweger, S. & Lüchinger-Hutter, L. (2004). *LEst 4–7: Lern- und Entwicklungsstand bei 4- bis 7-Jährigen*. Unveröffentlichter Test, Kompetenzzentrum für Bildungsevaluation und Leistungsmessung an der Universität Zürich.
- Murrey, C. & Greenberg, M. T. (2000). Children's relationship with teachers and bonds with school. An investigation of patterns and correlates in middle childhood. *Journal of School Psychology, 38*(5), 423–445.
- Muthén, L. & Muthén, B. O. (1998–2007). *MPlus User's Guide*. Los Angeles: Muthén & Muthén.
- Neuenschwander, M. P. (2005). *Unterrichtssystem und Unterrichtsqualität. Konturen einer Unterrichtstheorie für die Sekundarstufe und ihre empirische Bewährung*. Bern: Haupt.
- Neuenschwander, M. P. (2006). Klassenführung – Konzepte und neue Forschungsbefunde. *Schweizerische Zeitschrift für Bildungswissenschaften, 28*(2), 189–203.
- Nezlek, J. B., Schröder-Abé, M. & Schütz, A. (2006). Mehrebenenanalysen in der psychologischen Forschung. Vorteile und Möglichkeiten der Mehrebenenmodellierung mit Zufallskoeffizienten. *Psychologische Rundschau, 57*(4), 213–223.
- National Institute of Child Health and Human Development, Early Child Care Research Network (NICHD ECCRN) (2002). The relation of global first-grade classroom environment to structural classroom features and teacher and students behaviors. *The Elementary School Journal, 102*(5), 367–387.
- Noddings, N. (1992). *The challenge to care in schools: An alternative approach to education*. New York: Teacher Collage Press.
- Nolting, H.-P. (2008). *Störungen in der Schulklasse. Ein Leitfaden zur Vorbeugung und Konfliktlösung*. Weinheim: Beltz.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- O'Connor, E. (2010). Teacher-child relationships as dynamic systems. *Journal of School Psychology, 48*, 187–218.

- Ohle, A., McElvany, N., Horz, H., Ullrich, M., Schnotz, W. & Hochpöchler, U. (2012). *Entwicklung und Überprüfung von Kompetenzmodellen zur integrativen Verarbeitung von Texten und Bildern (BiTe III-Video)*. Poster präsentiert auf der Frankfurter Tagung zu Videoanalysen in der Unterrichts- und Bildungsforschung. 22.02.–23.02.2012. Frankfurt am Main.
- O'Leary, K. & O'Leary, S. (Hrsg.) (1977). *Classroom management: The successful use of behavior modification*. New York: Pergamon.
- Ophardt, D. (2008). Die Herstellung von Ordnung als Zumutung oder als Auftrag? Rekonstruktionen professioneller Orientierungen zum Klassenmanagement. In Y. Ehrenspeck, G. Haan & F. Thiel (Hrsg.), *Bildung: Angebot oder Zumutung?* (S. 243–258). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Ophardt, D. & Thiel, F. (2008). Klassenmanagement als Basisdimension der Unterrichtsqualität. In M.K.W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungsperspektiven und methodische Zugänge* (S. 259–284). Wiesbaden: Verlag für Sozialwissenschaften.
- Oser, F., Hascher, T. & Spychinger, M. (1999). Lernen aus Fehlern. Zur Psychologie des „negativen“ Wissens. In W. Althof & F. Oser (Hrsg.), *Fehlerwelten. Vom Fehlermachen und Lernen aus Fehlern* (S. 11–41). Opladen: Leske & Budrich.
- Oser, F. & Spychiger, M. (2005). *Lernen ist schmerzhaft. Zur Theorie des Negativen Wissens und zur Praxis der Fehlerkultur*. Weinheim: Beltz.
- Oser, F., Spychiger, M., Mahler, F., Gut, K., Hascher, T., Büeler, U. et al. (1998). *Lernen Menschen aus Fehlern? Zur Entwicklung einer Fehlerkultur in der Schule. Wissenschaftlicher Zwischenbericht (2) an den Schweizerischen Nationalfond zur Förderung wissenschaftlicher Forschung Abteilung I: Geistes- und Sozialwissenschaften*. Freiburg: Universitätsverlag.
- Oswald, F., Pfeifer, B., Ritter-Berlach, G. & Tanzer, N. (1989). *Schulklima. Die Wirkungen der persönlichen Beziehungen in der Schule*. Wien: Universitätsverlag.
- Pauli, C. & Reusser, K. (2006). Von international vergleichenden Video Surveys zur videobasierten Unterrichtsforschung und -entwicklung. *Zeitschrift für Pädagogik*, 52(6), 774–798.
- Pauli, C. & Reusser, K. (2010). Selbst- und Unterrichtswahrnehmung der Lehrpersonen. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität. Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 143–170). Münster: Waxmann.
- Pekrun, R. (1985). Schulischer Unterricht, schulische Bewertungsprozesse und Selbstkonzeptentwicklung. *Unterrichtswissenschaft*, 13(3), 220–248.
- Perry, K. E., Donohue, K. M. & Weinstein, R. S. (2007). Teaching practices and the promotion of achievement and adjustment in first grade. *Journal of School Psychology*, 45(3), 269–292.
- Perrez, M., Huber, G. L. & Geißler, K. A. (2001). Psychologie der pädagogischen Interaktion. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie – Ein Lehrbuch* (S. 357–413). Weinheim: Beltz.
- Peterson, P. L. & Fennema, E. (1985). Effective teaching, student engagement in classroom activities, and sex-related differences in learning mathematics. *American Educational Journal*, 22(3), 309–335.
- Petko, D. (2006). Kameraskript. In E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“*. Teil 3: *Videoanalysen*. Materialien zur Bildungsforschung. Bd. 15 (S. 15–37). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF). Online verfügbar: [http://www2.dipf.de/gfpf/publikationen/MatBild\\_Bd15.pdf](http://www2.dipf.de/gfpf/publikationen/MatBild_Bd15.pdf) (Zugriff: 31.01.2013).

- Petko, D., Waldis, M., Pauli, C. & Reusser, K. (2003). Methodologische Überlegungen zur videogestützten Forschung in der Mathematikdidaktik. *Zentralblatt für Didaktik der Mathematik*, 35(6), 265–280.
- Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R. & Morrison, F. J. (2008). Classroom effects on children's achievement trajectories in Elementary School. *American Educational Research Journal*, 45(2), 365–397.
- 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.
- Pianta, R. C. & La Paro, K. M. (2003). Improving early school success. *Educational Leadership*, 60(7), 24–29.
- Pianta, R. C., La Paro, K. M. & Hamre, B. K. (2008). *Classroom Assessment Scoring System. Manual*. Baltimore: Brookes.
- Pianta, R. C. & Steinberg, M. S. (1992). Teacher-child relationships and the process of adjusting to school. *New Direction for Child Development*, 57, 61–80.
- Pianta, R. C., Steinberg, M. S. & Rollins, K. B. (1995). The first two years of school: Teacher-child relationship and deflections in children's classroom adjustment. *Development and Psychopathology*, 7(2), 295–312.
- Pietsch, M. (2008). Struktur und Entwicklung von Unterrichtsqualität. Das Qualitätsstufenmodell des Unterrichts der Schulinspektion Hamburg. *Hamburg macht Schule*, 3–4, 52–53.
- Pietsch, M. (2010). Evaluation von Unterrichtsstandards. *Zeitschrift für Erziehungswissenschaft*, 13, 121–148.
- Pietsch, M., Schnack, J. & Schulze, P. (2009). Unterricht zielgerichtet entwickeln. Die Hamburger Schulinspektion entwickelt ein Stufenmodell für die Qualität von Unterricht. *Pädagogik*, 61(2), 38–43.
- Pietsch, M. & Tosana, S. (2008). Beurteilereffekte bei der Messung von Unterrichtsqualität. Das Multi-Facetten-Rasch-Modell und die Generalisierbarkeitstheorie als Methode der Qualitätssicherung in der externen Evaluation von Schule. *Zeitschrift für Erziehungswissenschaft*, 11(3), 430–452.
- Poloczek, S. (2007). *Zur Struktur des schulischen Selbstkonzeptes von Erstklässlerinnen und Erstklässlern*. Unveröffentlichte Diplomarbeit, Johann Wolfgang Goethe-Universität, Frankfurt am Main.
- Poloczek, S., Greb, K. & Lipowsky, F. (2008). *Lassen sich schon am Anfang der ersten Klasse verschiedene Facetten im schulischen Selbstkonzept unterscheiden?* Vortrag im Rahmen der 71. Tagung der Arbeitsgruppe für Empirische Pädagogische Forschung (AEPF). Kiel.
- Poloczek, S., Greb, K., Faust, G. & Lipowsky, F. (2009a). Lehrerfragebogen. In F. Lipowsky, G. Faust & K. Greb (Hrsg.), *Dokumentation der Erhebungsinstrumente des Projekts „Persönlichkeits- und Lernentwicklung von Grundschulkindern“ (PERLE) – Teil 1* (S. 76–126). Frankfurt am Main: Materialien zur Bildungsforschung (23/1).
- Poloczek, S., Greb, K., Faust, G. & Lipowsky, F. (2009b). Schulisches Selbstkonzept (Zusatzerhebung). In F. Lipowsky, G. Faust & K. Greb (Hrsg.), *Dokumentation der Erhebungsinstrumente des Projekts „Persönlichkeits- und Lernentwicklung von Grundschulkindern“ (PERLE) – Teil 1* (S. 72–75). Frankfurt am Main: Materialien zur Bildungsforschung (23/1).
- Powell, J. P. & Anderson, L. W. (1985). Humor and teaching in higher education. *Studies in Higher Education*, 10, 79–90.
- Praetorius, A.-K. (2014). *Messung von Unterrichtsqualität durch Ratings*. Münster: Waxmann.
- Praetorius, A.-K., Lenske, G. & Helmke, A. (2012). Observer ratings of instructional quality: Do they fulfill what they promise? *Learning and Instruction*, 22(6), 387–400.

- Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K. & Klieme, E. (2013). *Eine für alle ... Kann man Unterrichtsqualität mit einer Unterrichtsstunde pro Lehrkraft erfassen?* Vortrag auf der 2. Frankfurter Tagung zu Videoanalysen in der Unterrichts- und Bildungsforschung, 21.02.2013.
- Raaf, B. (2005). *Humor im Englischunterricht*. Berlin: Langenscheidt.
- Rakoczy, K. (2006). Motivationsunterstützung im Mathematikunterricht. Zur Bedeutung von Unterrichtsmerkmalen für die Wahrnehmung von Schülerinnen und Schülern. *Zeitschrift für Pädagogik*, 52(6), 822–843.
- Rakoczy, K. (2008). *Motivationsunterstützung im Mathematikunterricht. Unterricht aus der Perspektive von Lernenden und Beobachtern*. Münster: Waxmann.
- Rakoczy, K., Klieme, E., Drollinger-Vetter, B., Lipowsky, F., Pauli, C. & Reusser, K. (2007). Structure as a quality feature in mathematics instruction: Cognitive and motivational Effects of a structured organization of the learning environment vs. a structured presentation of learning content. In M. Prenzel (Ed.), *Studies on the educational quality of schools. The final report on the DFG Priority Program* (pp. 102–121). Münster: Waxmann.
- 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, Lernverhalten und mathematisches Verständnis“*. Teil 3: *Videoanalysen*. Materialien zur Bildungsforschung. Bd. 15 (S. 206–233). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF). Online verfügbar: [http://www2.dipf.de/gfpf/publikationen/MatBild\\_Bd15.pdf](http://www2.dipf.de/gfpf/publikationen/MatBild_Bd15.pdf) (Zugriff: 31.01.2013).
- Ramm, G. C., Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D. et al. (2006). *PISA 2003: Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models*. Thousand Oaks, CA: Sage.
- Reimann, G. (2004). *Kinder brauchen Anerkennung. Loben und belohnen – aber richtig*. Freiburg: Herder.
- Renkl, A. & Helmke, A. (1993). Prinzip, Nutzen und Grenzen der Generalisierungstheorie. *Empirische Pädagogik*, 7(1), 63–85.
- Renkl, A., Helmke, A. & Schrader, F. W. (1997). Schulleistung und Fähigkeitsselbstbild – Universelle Beziehungen oder kontextspezifische Zusammenhänge? Ergebnisse aus dem SCHOLASTIK-Projekt. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 374–383). Weinheim: Beltz.
- Reusser, K. (1999). Schülerfehler – die Rückseite des Spiegels. In W. Althof & F. Oser (Hrsg.), *Fehlerwelten. Vom Fehlermachen und Lernen aus Fehlern* (S. 203–231). Opladen: Leske & Budrich.
- Reyer, T. (2004). *Oberflächenmerkmale und Tiefenstrukturen im Unterricht. Exemplarische Analysen im Physikunterricht der gymnasialen Sekundarstufe*. Berlin: Logos-Verlag.
- Rheinberg, F. & Hoss, J. (1979). Störungen und Mitarbeit im Unterricht. Eine Erkundungsstudie zu Kounin's Kategorisierung des Lehrerverhaltens. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 6(3), 244–249.
- Rimmele, R. (2002). *Videograph – Multimedia-Player zur Kodierung von Videos*. Kiel: Leibniz-Institut für die Pädagogik der Naturwissenschaften.
- Rimm-Kaufman, S. E., Early, D., Cox, M., Saluja, G., Pianta, R., Bradley R. et al. (2002). Early behavioral attributes and teachers' sensitivity as predictors of competent behavior in the kindergarten classroom. *Journal of Applied Developmental Psychology*, 23, 451–470.
- Rißland, B. (2002). *Humor und seine Bedeutung für den Lehrerberuf*. Bad Heilbrunn: Klinkhardt.

- Roorda, D. L., Koomen, H. M. Y., Split, J. L. & Oort, F. J. (2011). The influence of affective teacher-student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research*, 81(4), 493–529.
- Roos, J. & Schöler, H. (2009). *Entwicklung des Schriftspracherwerbs in der Grundschule*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Rosenshine, B. V. (1970). Evaluation of classroom instruction. *Review of Educational Research*, 40, 279–300.
- Rosenshine, B. V. (1980). How time is spent in Elementary School. In C. Denham & A. Lieberman (Eds.), *Time to learn. A review of the beginning teacher evaluation study* (pp. 107–126). Washington, D. C.: National Institute of Education.
- Roßbach, H.-G. (2002a). Unterrichtsqualität im 2. Schuljahr – Ergebnisse einer empirischen Untersuchung. *Unterrichtswissenschaft*, 30(3), 230–245.
- Roßbach, H.-G. (2002b). Unterrichtsqualität in Grundschulklassen und ihre Auswirkungen auf die Schüler. In H. Petillon (Hrsg.), *Individuelles und soziales Lernen in der Grundschule – Kindperspektive und pädagogische Konzepte* (S. 217–222). Opladen: Leske & Budrich.
- Roßbach, H.-G., Kluczniok, K. & Isenmann, D. (2008). Erfahrungen aus internationalen Längsschnittuntersuchungen. In H.-G. Roßbach & S. Weinert (Hrsg.), *Bildungsforschung Band 24. Kindliche Kompetenzen im Elementarbereich: Förderbarkeit, Bedeutung und Messung* (S. 7–88). Berlin: Bundesministerium für Bildung und Forschung (BMBF). Referat Bildungsforschung.
- Rost, D. H. (2007). *Interpretation und Bewertung pädagogisch-psychologischer Studien. Eine Einführung*. Weinheim: Beltz.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*. Bern: Huber.
- Rowan, B. R., Correnti, R. & Miller, R. J. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the Prospects Study of elementary schools. *Teachers College Record*, 104(8), 1525–1567.
- Rutter, M., Höhn, K.-R. & Hentig, H. v. (1980). *Fünfzehntausend Stunden. Schulen und ihre Wirkung auf die Kinder*. Weinheim: Beltz.
- Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well being. *American Psychologist*, 55(1), 68–78.
- Saldern, M. v. (1987). *Sozialklima von Schulklassen. Überlegungen und mehrbenenanalytische Untersuchungen zur subjektiven Wahrnehmung von Lernumwelten*. Frankfurt am Main: Lang.
- Saldern, M. v. (1991). Die Lernumwelt aus der Sicht von Lehrern und Schülern. *Psychologie in Erziehung und Unterricht*, 38, 190–198.
- Saldern, M. v. (2000). Unterrichtsklima, Partizipation und soziale Interaktion. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Pädagogisch-psychologische Aspekte des Lehrens und Lernens in der Schule* (S. 565–581). Opladen: Leske & Budrich.
- Saldern, M. v. & Littig, K. E. (1985). Die Konstruktion der Landauer Skalen zum Sozialklima. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 17(2), 138–149.
- Saldern, M. v. & Littig, K. E. (1987). *Landauer Skalen zum Sozialklima. LASSO 4–13*. Weinheim: Beltz („Deutsche Schultests“).
- Satow, L. (1999a). *Klassenklima und Selbstwirksamkeitserwartung. Eine Längsschnittstudie in der Sekundarstufe I*. Dissertation, Freie Universität Berlin. Online verfügbar: [http://www.diss.fu-berlin.de/diss/servlets/MCRFileNodeServlet/FUDISS\\_derivate\\_00000000271/00\\_satow.pdf;jsessionid=9BA805C181180767FCC2F84DE4082F78?hosts=\(Zugriff: 31.01.2013\)](http://www.diss.fu-berlin.de/diss/servlets/MCRFileNodeServlet/FUDISS_derivate_00000000271/00_satow.pdf;jsessionid=9BA805C181180767FCC2F84DE4082F78?hosts=(Zugriff: 31.01.2013)).
- Satow, L. (1999b). Zur Bedeutung des Unterrichtsklimas für die Entwicklung schulbezogener Selbstwirksamkeitserwartungen. Eine Mehrebenenanalyse mit latenten Variablen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 31, 171–179.

- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research – Online*, 8(2), 23–74.
- Schermelleh-Engler, K. & Werner, C. (2006). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 113–133). Berlin, Heidelberg: Studien Verlag.
- Schoy-Lutz, M. (2005). *Fehlerkultur im Mathematikunterricht. Theoretische Grundlegung und evaluierte unterrichtspraktische Erprobung anhand der Unterrichtseinheit „Einführung in die Satzgruppe des Pythagoras“*. Hildesheim: Franzbecker.
- Schrader, F.-W., Helmke, A. & Dotzler, H. (1997). Zielkonflikte in der Grundschule: Ergebnisse aus dem SCHOLASTIK-Projekt. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 299–316). Weinheim: Beltz.
- Schulmeiß, I., Seidel T. & Meyer L. (2003). Vermischung von Lern- und Leistungssituationen. In T. Seidel, M. Prenzel, R. Duit & M. Lehrke (Hrsg.), *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“* (S. 229–238). Kiel: IPN.
- Schwarzer, R. (1983). Unterrichtsklima als Sozialisationsbedingung für Selbstkonzeptentwicklung. *Unterrichtswissenschaft*, 2, 129–148.
- Schwarzer, R. (1984). Lernumweltbedingungen der Selbstkonzeptentwicklung. In R. Schwarzer (Hrsg.), *Selbstbezogene Kognitionen: Trends in der Selbstkonzeptforschung. Band 2 des Berichts über das Forschungsvorhaben „Entwicklung des Selbstkonzepts und selbstbezogener Kognitionen in Abhängigkeit von sozialen Vergleichssituationen in schulischen Umwelten“* (S. 5–92). Berlin: Institut für Psychologie Freie Universität Berlin.
- Seidel, T. (2003a). Kapitel 6. Sichtstrukturen – Organisation unterrichtlicher Aktivitäten. In T. Seidel, M. Prenzel, R. Duit & M. Lehrke (Hrsg.), *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“* (S. 113–127). Kiel: IPN.
- Seidel, T. (2003b). *Lehr-Lernskripts im Unterricht. Freiräume und Einschränkungen für kognitive und motivationale Lernprozesse. Eine Videostudie im Physikunterricht*. Münster: Waxmann.
- Seidel, T. (2003c). Videobasierte Kodierverfahren in der IPN Videostudie Physik – ein methodischer Überblick. In T. Seidel, M. Prenzel, R. Duit & M. Lehrke (Hrsg.), *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“* (S. 99–111). Kiel: IPN.
- Seidel, T. (2005). Video analysis strategies of the IPN Video Study – a methodological overview. In T. Seidel, M. Prenzel & M. Kobarg (Eds.), *How to run a video study. Technical report of the IPN Video Study* (pp. 70–78). Münster: Waxmann.
- Seidel, T. (2009). Klassenführung. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 135–148). Heidelberg: Springer Verlag.
- Seidel, T., Dalehefte, I. M. & Meyer, L. (2001). Videoanalysen – Beobachtungsschemata zur Erfassung von Sicht-Strukturen im Physikunterricht. In M. Prenzel, R. Duit, M. Euler, M. Lehrke & T. Seidel (Hrsg.), *Erhebungs- und Auswertungsverfahren des DFG-Projekts „Lehr-Lern-Prozesse im Physikunterricht – eine Videostudie“* (S. 41–58). Kiel: IPN.
- Seidel, T. & Prenzel, M. (2003a). Gelegenheitsstrukturen beim Klassengespräch und ihre Bedeutung für die Lernmotivation – Videoanalysen in Kombination mit Schülerelbststeinschätzungen. *Unterrichtswissenschaft*, 31, 142–165.
- Seidel, T. & Prenzel, M. (2003b). Mit Fehlern umgehen – Zum Lernen motivieren. *Praxis der Naturwissenschaften – Physik in der Schule*, 52(1), 30–34.
- Seidel, T. & Prenzel, M. (2004). Muster unterrichtlicher Aktivitäten im Physikunterricht. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule. Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsentwicklung* (S. 177–194). Münster: Waxmann.

- Seidel, T., Prenzel, M., Duit, R., Euler, M., Geiser, H., Hoffmann, L. et al. (2002). „Jetzt mal alle nach vorne schauen!“ – Lehr-Lernskripts im Physikunterricht und damit verbundene Bedingungen für individuelle Lernprozesse. *Unterrichtswissenschaft*, 30(1), 52–77.
- Seidel, T., Prenzel, M., Duit, R. & Lehrke, M. (Hrsg.) (2003). *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“*. Leibniz-Institut für die Pädagogik der Naturwissenschaften. Kiel: IPN.
- Seidel, T., Prenzel, M., Rimmel, R., Herweg, C., Kobarg, M., Schwindt, K. et al. (2007). Science teaching and learning in German physics classrooms. Findings from the IPN Video Study. In M. Prenzel (Ed.), *Studies on the educational quality of schools. The final report on the DFG priority program* (pp. 79–99). Münster: Waxmann.
- Seidel, T. & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: Role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454–499.
- Shavelson, R. J. & Webb, N. M. (1991). *Generalizability theory. A primer*. Newbury Park, Calif.: Sage.
- Skinner, E. A. & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, 85(4), 571–581.
- Slavin, R. E. (1987). Cooperative learning and the cooperative school. *Educational Leadership*, 45, 7–13.
- Slavin, R. E. (1994). Quality, appropriateness, incentive, and time: A model of instructional effectiveness. *International Journal of Educational Research*, 21, 141–157.
- Spychiger, M., Mahler, F., Hascher T. & Oser F. (1998). *Fehlerkultur aus Sicht von Schülerinnen und Schülern. Der Fragebogen S-UFS. Entwicklung und erste Ergebnisse*. Schriftenreihe zum Projekt „Lernen Menschen aus Fehlern? Zur Entwicklung einer Fehlerkultur in der Schule“ Nr. ¾, Pädagogisches Institut der Universität Freiburg, CH.
- Steuer, G., Rosentritt-Brunn, G. & Dresel, M. (2013). Dealing with errors in mathematics classrooms: Structure and relevance of perceived error climate. *Contemporary Educational Psychology*, 36, 196–210.
- Stigler, J., Gonzales, P., Kawanaka, T., Knoll, S. & Serrano, A. (1999). *The TIMSS Videotape Classroom Study. Methods and findings from an exploratory research project on eight-grade mathematics instruction in Germany, Japan, and the United States*. Washington, DC.: U.S. Government Printing Office.
- Stipek, D. (1997). *The Early Childhood Classroom Observation Measure (ECCOM)*. Unveröffentlichtes Manuskript, University of California, Los Angeles.
- Stock, C. & Schneider, W. (2008). *DERET 1–2+*. *Deutscher Rechtschreibtest für das erste und zweite Schuljahr*. Göttingen: Hogrefe.
- Stückrath, J. & Strobel, R. (Hrsg.) (2005). *Deutschunterricht empirisch. Beiträge zur Überprüfbarkeit von Lernfortschritten im Sprach-, Literatur- und Medienunterricht*. Baltmannsweiler: Hohengehren.
- Stude, J. & Ohlhus, S. (2005). Schreibenlernen in interaktiven Kontexten. Dialogische Aneignungsmechanismen von Textkompetenz in schulischen Schreibprozessen. In J. Stückrath & R. Strobel (Hrsg.), *Deutschunterricht empirisch. Beiträge zur Überprüfbarkeit von Lernfortschritten im Sprach-, Literatur- und Medienunterricht* (S. 68–87). Baltmannsweiler: Hohengehren.
- Stuhlman, M. W. & Pianta, R. C. (2009). Profiles of educational quality in first grade. *Elementary School Journal*, 109(4), 323–342.

- Stumpff, T. & Großmann, H. (2009). Generalisierbarkeitstheorie. Kapitel 7. In H. Holling (Hrsg.), *Grundlagen und statistische Methoden der Evaluationsforschung* (S. 207–234). Göttingen: Hogrefe.
- Tausch, R. & Tausch, A. (1998). *Erziehungspsychologie*. Göttingen: Hogrefe.
- Teilprojekt 2 der „BiKS“-Forschergruppe (2009). *Manual zur Unterrichtsbeobachtung*. Bamberg: Otto-Friedrich-Universität.
- Thal, J. & Ebert, U. (2004). *Methodenvielfalt im Unterricht. Mit Lust stressarm und effektiv lernen*. München: Luchterhand.
- Tiedemann, J. & Billmann-Mahecha, E. (2004). Kontextfaktoren der Schulleistung im Grundschulalter. Ergebnisse aus der Hannoverschen Grundschulstudie. *Zeitschrift für Pädagogische Psychologie*, 18, 113–124.
- Toman, H. (2007). *Classroom-Management. Praxishilfen für das Classroom-Management*. Baltmannsweiler: Hohengehren.
- Tucker, L. R. & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1), 1–10.
- Ulewicz, M. & Beatty, A. (2001). *The power of video technology in international comparative research in education*. Washington: National Academy Press.
- Urban, D. & Mayerl, J. (2011). *Regressionsanalyse: Theorie, Technik und Anwendung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Van den Noortgate, W., Opendakker, M.-C. & Onghena, P. (2005). The effects of ignoring a level in multilevel analysis. *School Effectiveness and School Improvement*, 16(3), 281–303.
- Vehmeier, J. K. (2009). *Kognitiv anregende Verhaltensweisen von Lehrkräften im naturwissenschaftlichen Sachunterricht – Konzeptualisierung und Erfassung*. Dissertation, Westfälische Wilhelms-Universität Münster. Online verfügbar: [http://miami.uni-muenster.de/servelets/DerivateServlet/Derivate-5615/diss\\_vehmeier.pdf](http://miami.uni-muenster.de/servelets/DerivateServlet/Derivate-5615/diss_vehmeier.pdf) (Zugriff: 31.01.2013)
- Vehmeier, J., Kleickmann, T. & Möller, K. (2007). Zusammenhänge von Vorstellungen zum Lehren und Lernen mit unterrichtlichen Handlungen von Lehrkräften. In K. Möller, C. Beimbrech, P. Hanke, A. K. Hein, T. Kleickmann & R. Schages (Hrsg.), *Qualität von Grundschulunterricht. Entwickeln, erfassen und bewerten* (S. 317–320). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Veith, P. (2007). *Humor im Klassenzimmer. Soziale Kompetenzen stärken – Ermutigen – Motivieren*. Göttingen: Vandenhoeck & Ruprecht.
- Vispoel, W. P. (1995). Self-concept in artistic domains: An extension of the Shavelson, Hubner, and Stanton (1976) model. *Journal of Educational Psychology*, 87, 134–153.
- Voss, T., Kleickmann, T., Kunter, M. & Hachfeld, A. (2011). Überzeugungen von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Forschung zur professionellen Kompetenz von Lehrkräften – Ergebnisse des Projekts COACTIV* (S. 235–258). Münster: Waxmann.
- Walberg, H. J. (1981). A psychological theory of educational productivity. In F. H. Frarley und N. J. Gordon (Eds.), *Psychology and education* (pp. 214–229). Berkeley: McCutchan.
- Walberg, H. J. (1986). Synthesis of research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching. A project of the American Educational Research Association* (pp. 214–229). New York: Macmillan.
- Walberg, H. J. & Paik, S. J. (2000). *Effective educational practices*. Genf: International Academy of Education / International Bureau of Education.

- Waldis, M., Grob, U., Pauli, C. & Reusser, K. (2010a). Der Einfluss der Unterrichtsgestaltung auf Fachinteresse und Mathematikleistung. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität. Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 209–251). Münster: Waxmann.
- Waldis, M., Grob, U., Pauli, C. & Reusser, K. (2010b). Der schweizerische Mathematikunterricht aus der Sicht von Schülerinnen und Schülern und in der Perspektive hochinferenter Beobachterurteile. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität. Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 171–208). Münster: Waxmann.
- Wang, M. C., Haertel, G. D. & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *Journal of Educational Research*, 84(1), 30–43.
- Wang, M. C., Haertel, G. D. & Walberg, H. J. (1993a). Toward a knowledge base: why, how, for whom? *Review of Educational Research*, 63(3), 365–376.
- Wang, M. C., Haertel, G. D. & Walberg, H. J. (1993b). Toward a knowledge base for school learning. *Review of Educational Research*, 63(3), 249–294.
- Wannack, E. (2012). Classroom Management und seine Bedeutung für die Gestaltung von Spiel- und Lernaktivitäten. In F. Hellmich, S. Förster & F. Hoya (Hrsg.), *Bedingungen des Lehrens und Lernens in der Grundschule. Bilanz und Perspektiven* (S. 69–72). Wiesbaden: Springer.
- Wannack, E., Arnaldi, U. & Schütz, A. (2009). Überlegungen zur Didaktik des Kindergartens. *4 bis 8. Fachzeitschrift für Kindergarten und Unterstufe*, 99(9), 4–6.
- Wannack, E., Herger, K. & Barblan, A. (2011). *Classroom Management in der Schuleingangsstufe. Dokumentation der qualitativen Erhebung*. Online verfügbar: [http://www.schuleingangsstufe.phbern.ch/download/dokumentation\\_qualitativer\\_teil.pdf](http://www.schuleingangsstufe.phbern.ch/download/dokumentation_qualitativer_teil.pdf). (Zugriff: 31.01.2013).
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response models. *Psychometrika*, 54(3), 427–450.
- Webb, N. M., Shavelson, R. J. & Haertel, E. H. (2007). Reliability coefficients and generalizability theory. In C. Rao & S. Sinharay (Eds.), *Handbook of Statistics*, Bd. 26, (pp. 81–124). Amsterdam: Elsevier.
- Weinert, F. E. (1996). *Enzyklopädie der Psychologie. Pädagogische Psychologie. Psychologie des Lernens und der Instruktion*. Göttingen: Hogrefe.
- Weinert, F. E. (1998). Lehrerkompetenz als Schlüssel der inneren Schulreform. *Schulreport*, 2, 24–27.
- Weinert, F. E. & Helmke A. (1996). Der gute Lehrer: Person, Funktion oder Fiktion? *Zeitschrift für Pädagogik*, 34. Beiheft, 223–233.
- Weinert, F. E. & Helmke, A. (Hrsg.) (1997). *Entwicklung im Grundschulalter*. Weinheim: Beltz.
- Weinert, F. E., Schrader, F.-W. & Helmke, A. (1989). Quality of instruction and achievement outcomes. *International Journal of Educational Research*, 13(2), 895–912.
- Weinert, F. E., Schrader, F.-W. & Helmke, A. (1990a). Educational expertise: closing the gap between educational research and classroom practice. *School Psychology International*, 11(3), 1063–1080.
- Weinert, F. E., Schrader, F.-W. & Helmke, A. (1990b). Unterrichtsexpertise – ein Konzept zur Verringerung der Kluft zwischen theoretischen Paradigmen. In L.-M. Alisch, J. Baumert & K. Beck (Hrsg.), *Professionswissen und Professionalisierung* (S. 173–206). Braunschweig: Copy-Center Colmsee.
- Wentzel, K. R. (1997). Student motivation in Middle School: The role of perceived pedagogical caring. *Journal of Educational Psychology*, 89(3), 411–419.
- Wharton-McDonald, R., Pressley, M. & Hampston, J. M. (1998). Literacy instruction in nine first-grade classrooms: Teacher characteristics and student achievement. *Elementary School Journal*, 99(2), 101–128.

- Wicki, W. (2000). Humor und Entwicklung: Eine kritische Übersicht. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 32(4), 173–185.
- Wildemann, A. (2011). Qualität im sprachlichen Anfangsunterricht – fachdidaktische Perspektiven und empirische Notwendigkeiten. In K.-O. Bauer & N. Logemann (Hrsg.), *Unterrichtsqualität und fachdidaktische Forschung. Modelle und Instrumente zur Messung fachspezifischer Lernbedingungen und Kompetenzen* (S. 75–100). Münster: Waxmann.
- Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.
- Woolfolk, A. (2008). *Pädagogische Psychologie*. München: Pearson Studium.
- Ysewijn, P. (1997). GT – Programm für Generalisierbarkeitsstudien. Neuchâtel: Institut de recherche et de documentation pédagogique. Verfügbar unter <http://www.irdp.ch/methodo/generali.htm> (Zugriff: 05.12.2012)
- Zimmer, R. (1994). Bewegung und Sport. In G. Faust-Siehl & R. Portmann (Hrsg.), *Die ersten Wochen in der Schule* (S. 123–134). Frankfurt am Main: Arbeitskreis Grundschule.
- Zimmer, R. (1997). Bewegung, Spiel und Sport. Fachspezifische und fachübergreifende Lern- und Erfahrungsmöglichkeiten. In D. Haarmann (Hrsg.), *Grundschule. Ein Handbuch* (S. 323–334). Weinheim: Beltz.
- Ziv, A. (1988). Teaching and learning with humor: Experiments and replication. *Journal of Experimental Education*, 5(1), 5–15.
- Zumhasch, C. (2006). Das Unterrichtsklima. In K.-H. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht* (S. 144–147). Bad Heilbrunn: Klinkhardt.



## VIII Abbildungsverzeichnis

Abbildung 1.: Ausgewählte Entwicklungslinien der Modelle und Konzepte zur Unterrichtsqualität (adaptiert nach Einsiedler, 2012, S. 21) .....	14
Abbildung 2.: Gegenüberstellung der Merkmalskataloge von Brophy (2000), Meyer (2004) und Helmke (2009) .....	18
Abbildung 3.: Basisdimensionen der Unterrichtsqualität (nach Klieme, Schümer & Knoll 2001, S. 51) .....	19
Abbildung 4.: Das mehrdimensionale Stufenmodell (in Anlehnung an Pietsch et al., 2009, S. 39) .....	23
Abbildung 5.: Wirkungsgeflecht der Klassenführung (vgl. Helmke, 2007, S. 45) .....	25
Abbildung 6.: Überblick über die Faktoren des CLASS-Beobachtungssystems und deren Merkmale (vgl. Pianta et al., 2008, S. 88) .....	34
Abbildung 7.: Untersuchte Merkmale der Klassenführung im Überblick .....	66
Abbildung 8.: Untersuchte Merkmale des Unterrichtsklimas im Überblick .....	88
Abbildung 9.: Basisdimensionen der Unterrichtsqualität und deren vermutete Wirkmechanismen (nach Klieme, Lipowsky, Rakoczy & Ratzka, 2006, S. 131) .....	92
Abbildung 10: Beziehungsstruktur von Video (= Messobjekt), Rater und Merkmal (adaptiert nach Langer & Schulz von Thun, 2007, S. 15) .....	117
Abbildung 11: Design der PERLE-I-Studie (Laufzeit 2006 bis 2008, Einschulung bis Ende zweites Schuljahr) .....	137
Abbildung 12: Design der Videostudie Deutsch (2. Halbjahr 1. Schuljahr, März 2007) .....	139
Abbildung 13: Beispiel für das Antwortformat (Gruppenbefragung Schüler Videostudie Deutsch) .....	140
Abbildung 14: Design der Videostudie Mathematik (2. Halbjahr 2. Schuljahr, März 2008) .....	141
Abbildung 15: Ein Beispiel für den möglichen Aufbau der Kameras und Mikrophone im Klassenraum .....	144
Abbildung 16: Repräsentativität der Unterrichtsstunden in Deutsch (VSS, 1. Schuljahr) .....	153
Abbildung 17: Repräsentativität der Unterrichtsstunden in Mathematik (VSM, 2. Schuljahr) .....	153
Abbildung 18: Vorgehen bei der Entwicklung des hoch inferenten Ratingsystems für die Analyse der Unterrichtsvideos in PERLE (in Anlehnung an Hugener et al., 2006; Seidel, 2003c) .....	160
Abbildung 19: Auszug aus dem Softwareprogramm Videograph (vgl. Rimmel, 2002) .....	167
Abbildung 20: Vorgehen bei der Klassenteilung am Beispiel einer Mathematikstunde „Einführung in die Multiplikation“ .....	169
(in Anlehnung an Lotz, Berner & Gabriel, 2013)	

Abbildung 21: Drei Antwortmöglichkeiten im Rahmen der Selbstkonzepterhebung (vgl. Poloczek, 2007; Gabriel, Kastens, Poloczek, Schoreit & Lipowsky, 2010; Gabriel, Mösko & Lipowsky, 2010) .....	177
Abbildung 22: Mittelwerte getrennt für VSS (1. Schuljahr) und VSM (2. Schuljahr) .....	189
Abbildung 23: Schematische Darstellung der erwarteten Faktorenstruktur (2-Faktoren-Modell) .....	194
Abbildung 24: Das Generalfaktor-Modell (Modell 1) .....	198
Abbildung 25: Das 2-Faktoren-Modell (Modell 2a) .....	200
Abbildung 26: Das 3-Faktoren-Modell (Modell 3) .....	201
Abbildung 27: Modell 2b mit zugelassenen Residualkorrelationen zwischen den Merkmalen des Feedbacks .....	202
Abbildung 28: Mehrebenenregressionsmodell 1a: standardisierte Regressions- koeffizienten (AV = Rechtschreibleistungen / DERET Ende 1. SJ) .....	219
Abbildung 29: Mehrebenenregressionsmodell 1b: standardisierte Regressions- koeffizienten (AV = Selbstkonzept Schreiben Ende 1. SJ) .....	220
Abbildung 30: Mehrebenenregressionsmodell 2a (AV = Leistungen Mathematik Ende 2. SJ) .....	221
Abbildung 31: Mehrebenenregressionsmodell 2b (AV = Leistungen Mathematik Ende 2. SJ) .....	222
Abbildung 32: Mehrebenenregressionsmodell 2c (AV = Selbstkonzept Rechnen Ende 2. SJ) .....	223
Abbildung 33: Mehrebenenregressionsmodell 2d (AV = Selbstkonzept Rechnen Ende 2. SJ) .....	224
Abbildung 34: Ergebnis der konfirmatorischen Faktorenanalyse – Das Generalfaktor-Modell .....	325
Abbildung 35: Ergebnis der konfirmatorischen Faktorenanalyse – Das Modell 2 .....	325
Abbildung 36: Ergebnis der konfirmatorischen Faktorenanalyse – Das Modell 2b .....	326
Abbildung 37: Ergebnis der konfirmatorischen Faktorenanalyse – Das Modell 3 .....	326

## IX Tabellenverzeichnis

Tabelle 1: Komponenten der Unterrichtsqualität in der Grundschule – Eine Gegenüberstellung verschiedener Studien .....	33
Tabelle 2: Erhebungsmethoden und -instrumente der aktuellen Unterrichtsqualitätsforschung im Grundschulbereich .....	38
Tabelle 3: Überblick über videobasierte Studien in der Grundschule (Fach, Merkmale der Unterrichtsqualität und Erhebungsmethode) .....	42
Tabelle 4: Proaktive und reaktive Kriterien effektiven ‚Classroom Managements‘ nach Hennemann und Hillenbrand (2010, S. 259) .....	52
Tabelle 5: Überblick über die Merkmale einer effektiven Klassenführung nach Kounin (2006) .....	53
Tabelle 6: Übersicht über die Facetten des Klimabegriffs nach Eder (1998, S. 40) .....	69
Tabelle 7: Einteilung der erfassten Einzelmerkmale des Unterrichtsklimas in drei Komponenten in Anlehnung an LASSO (vgl. v. Saldern & Littig, 1985) .....	73
Tabelle 8: Bedeutung der Klassenführung für die Leistungen in der Grundschule – Eine Gegenüberstellung ausgewählter Studien .....	96
Tabelle 9: Bedeutung des Unterrichtsklimas für die Leistungen in der Grundschule – Eine Gegenüberstellung ausgewählter Studien .....	106
Tabelle 10: Verteilung der teilnehmenden Schulen / Klassen und Anzahl der Unterrichtsvideos nach Videostudie .....	146
Tabelle 11: Kennwerte für die Lehrerstichprobe je nach Videostudie .....	149
Tabelle 12: Kennwerte der Schülerstichprobe nach Schultyp und Fach .....	151
Tabelle 13: Überblick über die hoch inferenten Merkmale des Ratingsystems zur Erfassung der Klassenführung und des Unterrichtsklimas in der Grundschule .....	162
Tabelle 14: Aufbau des hoch inferenten Ratingsystems am Beispiel des Übergangsmanagements (Managing Transition) .....	164
Tabelle 15: Deskriptive Statistiken der arithmetischen Grundkenntnisse und Fähigkeiten (Ende 1. und Ende 2. Schuljahr) .....	173
Tabelle 16: Deskriptive Statistiken der schriftspracherwerbsspezifischen Vorläuferfähigkeiten Anfang 1. Schuljahr .....	174
Tabelle 17: Deskriptive Statistiken der Rechtschreibleistungen Ende 1. Schuljahr .....	175
Tabelle 18: Items und Antwortformate zur Erfassung der Selbstkonzepte im Rechnen und Schreiben .....	176
Tabelle 19: Interne Konsistenzen (Cronbachs Alpha) sowie Mittelwerte und Standardabweichungen (in Klammern) für die Selbstkonzeptskalen Rechnen und Schreiben zu drei Messzeitpunkten .....	177
Tabelle 20: Varianzkomponenten und relative G-Koeffizienten für die einzelnen Merkmale der Klassenführung und des Unterrichtsklimas ( $N_{VSS-VSM} = 97$ ) .....	182
Tabelle 21: Relative G-Koeffizienten für die einzelnen Merkmale der Klassenführung und des Unterrichtsklimas (getrennt nach Videostudie) .....	185

Tabelle 22: Varianzkomponenten für die einzelnen Merkmale der Klassenführung und des Unterrichtsklimas (getrennt nach Videostudie) .....	187
Tabelle 23: Unterschiede in den Ausprägungen der Klassenführungsmerkmale in der VSS (1. Schuljahr) und VSM (2. Schuljahr) .....	191
Tabelle 24: Unterschiede in den Ausprägungen der Unterrichtsklimamerkmale in der VSS (1. Schuljahr) und VSM (2. Schuljahr) .....	192
Tabelle 25: Korrelationsmatrix für die 14 verbliebenen Merkmale .....	195
Tabelle 26: Wertebereiche und Cut-off-Werte der verwendeten Fit-Indizes nach Bühner (2006) .....	198
Tabelle 27: Der Modellvergleich (Modell 1, 2, 2b und 3) und die Gütekriterien der konfirmatorischen Faktorenanalysen ( $N = 101$ ) .....	203
Tabelle 28: Korrelationsmatrix für die 14 verbliebenen Merkmale der Videostudie Deutsch (VSS) .....	206
Tabelle 29: Korrelationsmatrix für die 14 verbliebenen Merkmale der Videostudie Mathematik (VSM) .....	207
Tabelle 30: Faktorladungen der hoch inferenten Merkmale sowie interne Konsistenzen der extrahierten Faktoren in der Videostudie Deutsch ( $N = 50$ ) .....	208
Tabelle 31: Faktorladungen der hoch inferenten Merkmale sowie interne Konsistenzen der extrahierten Faktoren in der Videostudie Mathematik ( $N = 51$ ) .....	209
Tabelle 32: Skalenskennwerte für die Videostudie Deutsch (1. Schuljahr) und die Videostudie Mathematik (2. Schuljahr) .....	211
Tabelle 33: Zusammenhang zwischen den zwei Unterrichtsmerkmalen in der Videostudie Deutsch (VSS, März 2007) .....	214
Tabelle 34: Zusammenhang zwischen den drei Unterrichtsmerkmalen in der Videostudie Mathematik (VSM, März 2008) .....	214
Tabelle 35: Überblick über die Mehrebenenregressionsmodelle der Studie 3 .....	216
Tabelle 36: Intraklassenkorrelationen (ICC) für abhängige Variablen der Studie 3 .....	216
Tabelle 37: Merkmale des hoch inferenten Ratings zur Klassenführung (Quellen) .....	291
Tabelle 38: Merkmale des hoch inferenten Ratings zum Unterrichtsklima (Quellen) .....	306
Tabelle 39: Häufigkeitsverteilung der Klassenführungsmerkmale in der Videostudie Deutsch (VSS, 1. Schuljahr, $N = 50$ ) .....	322
Tabelle 40: Häufigkeitsverteilung der Unterrichtsklimamerkmale in der Videostudie Deutsch (VSS, 1. Schuljahr, $N = 50$ ) .....	322
Tabelle 41: Häufigkeitsverteilung der Klassenführungsmerkmale in der Videostudie Mathematik (VSM, 2. Schuljahr, $N = 51$ ) .....	323
Tabelle 42: Häufigkeitsverteilung der Unterrichtsklimamerkmale in der Videostudie Mathematik (VSM, 2. Schuljahr, $N = 51$ ) .....	323
Tabelle 43: Prüfung auf Normalverteilung der Merkmale in VSS (Kolmogorov-Smirnov-Test) .....	324
Tabelle 44: Prüfung auf Normalverteilung der Merkmale in VSM (Kolmogorov-Smirnov-Test) .....	324

## X Anhang

### 19 Das hoch inferente Ratingsystem

Dieses hoch inferente Ratingssystem zur Erfassung der Klassenführung und des Unterrichtsklimas im Anfangsunterricht wird im Technischen Bericht des PERLE-Projekts ähnlich publiziert (vgl. Gabriel & Lipowsky, 2013a; b).

#### 19.1 Merkmale der Klassenführung

Unter Rückgriff auf theoretische Ansätze und bereits existierende Manuale aus der Sekundarstufe wurden hoch inferente Merkmale entwickelt.

Tabelle 37: Merkmale des hoch inferenten Ratings zur Klassenführung (Quellen)

Merkmale	Quellen
Allgegenwärtigkeit d. Lehrperson (Allg)	<ul style="list-style-type: none"> <li>– Eigenentwicklung in Anlehnung an Kounin (1976; 2006; „Withitness“)</li> <li>– Gruehn (2000) („störungspräventive Überwachung der Schülertätigkeiten“), adaptiert</li> </ul>
Gruppenfokus (Grup)	<ul style="list-style-type: none"> <li>– Eigenentwicklung in Anlehnung an Kounin (1976; 2006; „Group Focus“)</li> </ul>
Zeitmanagement / effiziente Zeitnutzung (Zeit)	<ul style="list-style-type: none"> <li>– CLASS („Classroom Assessment Scoring System“, Pianta et al., 2008), adaptiert</li> <li>– Eigenentwicklung in Anlehnung Helmke, A., Helmke, T., Schrader &amp; Wagner (2007)</li> <li>– Klieme (2002, TIMSS), adaptiert</li> </ul>
Übergangsmanagement (Managing Transition) (Ueb)	<ul style="list-style-type: none"> <li>– CLASS („Classroom Assessment Scoring System“, Pianta et al., 2008), adaptiert</li> </ul>
Regelklarheit bzw. -verwendung (Regl)	<ul style="list-style-type: none"> <li>– Teilprojekt 2 der „BiKS“-Forschergruppe (2009)</li> <li>– CES („Classroom Environment Scale Manual“, Moos &amp; Trickett, 1974), adaptiert</li> </ul>
Störungsfreiheit (Disz)	<ul style="list-style-type: none"> <li>– Rakoczy &amp; Pauli (2006), adaptiert</li> <li>– Ramm et al. (2006, PISA 2003, „Disciplinary Climate“), adaptiert</li> <li>– Clausen (2002), adaptiert</li> <li>– Helmke et al. (2007), adaptiert</li> </ul>
Effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)	<ul style="list-style-type: none"> <li>– Helmke et al. (2007), adaptiert</li> </ul>
Vorbereitung des Klassenraums (Vorb)	<ul style="list-style-type: none"> <li>– Eigenentwicklung in Anlehnung an Evertson et al. (2006)</li> </ul>

In Tabelle 37 ist notiert, auf welche Quellen bei der Entwicklung der Dimensionen Bezug genommen wurde. Das hoch inferente Ratingsystem zur Erfassung der Klassenführung in der Grundschule umfasst insgesamt acht Merkmale. Anschließend werden die Merkmale der Klassenführung vorgestellt.

### *19.1.1 Allgegenwärtigkeit der Lehrperson (Allg)*

#### *Grundidee*

Das Merkmal erfasst, inwiefern die Lehrperson im Klassenzimmer Präsenz zeigt bzw. im Sinne Kounins (1976, 2006) „allgegenwärtig“ ist. Die Schüler haben das Gefühl, dass die Lehrperson sie immer und überall im Blick hat. Die Lehrperson vermittelt den Schülern, dass sie über ihr Tun stets informiert ist. Das beinhaltet die Fähigkeit der Lehrperson, mehrere Dinge gleichzeitig zu tun (Multitasking), sei es durch Anweisungen, Bemerkungen oder nur durch Blicke. Die Lehrperson kann sich mehreren Verhaltensproblemen widmen, ohne die Übersicht und die Kontrolle über die Situation zu verlieren. Sie hat sprichwörtlich „Augen im Hinterkopf“. Die Allgegenwärtigkeit einer Lehrperson zeigt sich besonders in Phasen, in denen sie mit einzelnen Schülern kommuniziert. In diesem Fall ist darauf zu achten, ob sie auch noch den Rest der Klasse im Blick hat.

#### *Indikatoren*

- Die Lehrperson hat die Aktivitäten der Schüler immer im Blick.
- Die Lehrperson zeigt im ganzen Klassenzimmer Präsenz.
- Die Lehrperson minimiert Einzelkontakte zwischen Schülern sehr schnell.
- Die Lehrperson richtet ihren Blick vorwiegend zu ihren Schülern, kehrt ihnen nicht allzu häufig den Rücken zu. Beispielsweise richtet die Lehrperson auch beim Schreiben an die Tafel immer wieder den Blick zur Klasse.
- Die Allgegenwärtigkeit kann sich auch durch gelegentliche Bewegungen durch den Klassenraum äußern, z. B. während Einzelarbeitsphasen.
- Die Lehrperson kontrolliert während des Unterrichtsgesprächs gleichzeitig das Schülerverhalten einzelner Schüler und erkennt somit den / die Verursacher von Störungen sofort.
- Beim Einsammeln oder / und Austeilen von Arbeitsmaterialien lässt die Lehrperson ihren Blick in der ganzen Klasse hin und her wandern.

#### *Negativindikatoren*

- Während die Lehrperson mit einzelnen Schülern kommuniziert, hat sie die restlichen Schüler nicht mehr im Blick.

- Die Lehrperson kehrt ihren Schülern häufig den Rücken zu. Dadurch übersieht sie oft einzelne Schülermeldungen bzw. sieht sie erst sehr spät.

*Antwort: Gesamteindruck*

- Eine „4“ wird vergeben, wenn die Lehrperson ihre Schüler ständig im Blick hat. Jede Aktivität der Schüler scheint von ihr gesehen zu werden (Meldungen oder Einzelgespräche). Nicht erwünschtes Verhalten unterbindet sie sofort.
- Eine „3“ wird vergeben, wenn es der Lehrperson nicht immer, aber häufig gelingt „allgegenwärtig“ zu sein. Vereinzelt sieht sie Meldungen nicht oder hat in Einzelgesprächen die anderen Schüler der Klasse nicht konsequent im Blick.
- Eine „2“ wird vergeben, wenn die Lehrperson bei Einzelgesprächen etc. die Aktivitäten der restlichen Klasse selten im Blick hat. Viele Aktivitäten der Schüler bleiben somit unbeobachtet bzw. ungeahndet.
- Eine „1“ wird vergeben, wenn die Lehrperson ihren Schülern sehr häufig den Rücken zukehrt und sich in Einzelgesprächen ganz dem Schüler zuwendet. Sie bekommt in diesen Situationen sehr selten mit, was in der restlichen Klasse vor sich geht.

*Anmerkungen*

„Allgegenwärtigkeit der Lehrperson“ ist besonders in solchen Situationen zu erkennen, in denen die Lehrperson etwas an die Tafel schreibt, ihr weiteres Arbeitsmaterial sortiert etc. Ein besonderes Augenmerk ist auf Einzelgespräche zwischen der Lehrperson und einem Schüler / einer kleinen Schülergruppe zu legen.

### *19.1.2 Gruppenfokus (Grup)*

*Grundidee*

Diese Merkmal bezieht sich auf das gleichzeitige und aktive Einbeziehen aller Schüler in den Unterricht durch Aktivierung und Wertung der Beiträge („Gruppenfokus“; Kounin, 1976, S. 117). Das Merkmal beinhaltet zusätzlich das sogenannte „Rechenschaftsprinzip“ (Kounin, 1976, S. 126). Dabei geht es darum, dass während einer Stunde immer wieder andere Schüler drankommen und nicht nur einige wenige. Bei der Art der Ansprache müssen also alle damit rechnen, in den kommenden Sekunden dranzukommen („Gruppenmobilisierung“ oder „Klassenaktivierung“).

*Indikatoren*

- Die Lehrperson stellt die Fragen an die ganze Klasse und dabei wandert ihr Blick von Schüler zu Schüler. Durch das Pausieren und Sich-Um-

schaufen schafft die Lehrperson eine gewisse Spannung, bevor sie einzelne Schüler aufruft.

- Die Lehrperson ruft auch Schüler auf, die sich nicht melden. Das ist besonders in Situationen ersichtlich, in denen die Lehrperson eine Frage stellt und die Schüler auffordert sich zu melden.
- Die Lehrperson ruft nicht immer die gleichen Schüler auf. Die Antwortmöglichkeiten sind relativ gleichmäßig verteilt.
- Die Lehrperson reagiert nicht nur auf die guten Schüler, die als Stichwortgeber fungieren, sondern auch auf leistungsschwächere Schüler, die verstärkt Hilfe bei der Antwort brauchen.
- Die Lehrperson beteiligt andere Schüler an den Antworten des aufgerufenen Schülers: „*Max, du passt auf, wie Katharina diese Aufgabe löst, und sagst mir dann, was sie richtig oder falsch gelöst hat.*“
- Die Lehrperson fordert auch Meldungen ein, wenn sie den Eindruck hat, dass sich immer die gleichen Schüler melden. Sie nimmt somit nicht immer diejenigen Schüler dran, die sich als erste melden.

#### *Negativindikatoren*

- Die Lehrperson ruft häufig die gleichen Schüler auf. Die Antwortmöglichkeiten sind ungleichmäßig verteilt.
- Die Lehrperson ruft die Schüler häufig in Sitzreihenfolge auf. Die Schüler wissen genau, wann sie an der Reihe sind.
- Es wird ersichtlich, dass die Lehrperson verstärkt auf die guten Schüler reagiert, die als Stichwortgeber fungieren. Sie arbeitet mit Zwischenrufen.
- Die Lehrperson nimmt nur diejenigen Schüler dran, die sich melden.
- Die Lehrperson nimmt häufig die Schüler dran, die sich als erste melden.

#### *Antwort: Gesamteindruck*

- Eine „4“ wird vergeben, wenn die Lehrperson keinem festen Schema folgt und nicht nur diejenigen Schüler aufruft, die sich immer melden. Zu keinem Zeitpunkt fungieren immer die gleichen Schüler als Stichwortgeber. Auch Schüler, die sich nicht melden, werden drangenommen, wobei die Lehrperson sich Zeit nimmt und schwächeren Schülern bei der Antwort hilft.
- Eine „3“ wird vergeben, wenn die Lehrperson bemüht ist, die Antwortmöglichkeiten gleichmäßig zu verteilen, sich jedoch in einzelnen Situationen auf gute Schüler als Stichwortgeber verlässt.
- Eine „2“ wird vergeben, wenn die Lehrperson häufig die gleichen Schüler drannimmt, um im Unterricht schneller voranzukommen. Sie ruft häufig diejenigen Schüler auf, die sich melden. Eine „2“ wird außerdem

- vergeben, wenn die Lehrperson in manchen Situationen nach einem festen Schema vorgeht.
- Eine „1“ wird vergeben, wenn einige Schüler während des Unterrichts kein einziges Mal drangenommen werden. Es scheinen immer die gleichen Schüler von der Lehrperson aufgerufen zu werden. Dabei folgt die Lehrperson häufig einem festen Schema.

### *Anmerkungen*

Bei dieser Dimension sollten besonders die Phasen des öffentlichen Unterrichtsgesprächs („OEU“/“OEUSK“; vgl. Lotz, 2013b) beobachtet werden. Dabei liegt der Fokus auf der Lehrperson und wie sie mit Meldungen bzw. dem Drannehmen/Aufrufen von Schülern umgeht. Welche Schüler sich melden und welche die Lehrperson aufruft, kann unter Umständen mithilfe des Transkripts nachvollzogen werden. Es geht in dieser Dimension darum, eine gewisse „Ungerechtigkeit“ in der Verteilung von Lern- und Antwortgelegenheiten aufseiten der Lehrperson einzuschätzen. Diese Ungerechtigkeit kann auch von den Schülern wahrgenommen werden und darüber auch das Unterrichtsklima beeinflussen.

Ein Beispiel für eine gute Gruppenmobilisierung ist: Die Lehrperson kündigt eine Aufgabe mit den Worten an: „*Jetzt passt mal alle gut auf. Bestimmt kann da jeder was dazu sagen ...*“ Eine „kollektive“ Mobilisierung erfolgt nicht, wenn die Lehrperson z. B. vor einer Frage bzw. Aufgabe einen Schülernamen nennt oder eine Reihenfolge des „Drankommens“ vorgibt (z. B. „*Erst Max, dann Katharina und dann Felix!*“).

### *19.1.3 Zeitmanagement/effektive Zeitnutzung (Zeit)*

#### *Grundidee*

Ein Schwerpunkt der Klassenführung als Form des Managements von Lernzeit ist die maximale Bereitstellung von Zeit, in der sich die Schüler aktiv mit den Lerninhalten auseinandersetzen können. Aus diesem Grund steht im Vordergrund dieses Merkmals die Einschätzung, ob der Unterrichtsablauf auf eine effektive Unterrichts- und Zeitplanung der Lehrperson schließen lässt. Das Merkmal beschreibt, wie effizient die zur Verfügung stehende Unterrichtszeit von der Lehrperson genutzt wird, z. B. ob die Stunde pünktlich beginnt, kein Leerlauf entsteht oder ob das Material und die Unterlagen, die im Laufe der Stunde benötigt werden, bereit liegen.

#### *Indikatoren*

- Die Stunde wirkt geplant und organisiert, das heißt, der Unterrichtsablauf lässt auf eine angemessene Unterrichts- und Zeitplanung schließen.

- Die Lehrperson wirkt vorbereitet. Der Ablauf und Inhalt scheinen durchdacht. Das wird z. B. ersichtlich, wenn die Lehrperson mit den Schülern den Ablauf der Stunde bzw. einzelner Unterrichtsphasen bespricht.
- Die Stunde beginnt und endet pünktlich. Auch die Pause sollte pünktlich enden.
- Es wird keine Zeit verschwendet.
- Die Übergänge zwischen den einzelnen Unterrichtsphasen erfolgen kurz und reibungslos.
- Die Lehrperson hat für die einzelnen Unterrichtsabschnitte genügend Zeit eingeplant.
- Die Lehrperson macht klare Zeitansagen. Die Schüler wissen, wieviel Zeit sie für die einzelnen Aufgaben zur Verfügung haben: „*So, noch fünf Minuten.*“
- Es gibt keine unnötigen Warte- oder Leerzeiten, in denen die Schüler nicht wissen, was sie machen sollen.
- Die Lehrperson hat Differenzierungsmaßnahmen eingeplant, damit schnelle Schüler nicht untätig warten müssen.
- Der Unterricht wird für fachliche Themen genutzt. Sogenannte *non-instructional activities* werden aus dem eigentlichen Unterrichtsgeschehen ausgelagert (organisatorische Tätigkeiten sind z. B. das Einsammeln von Milchgeld oder das Verteilen von Elternbriefen).

*Negativindikatoren (Indikatoren für „Zeitverschwendung“)*

- Die Unterrichtsstunde beginnt unpünktlich.
- Es dauert zu Beginn der Stunde sehr lange, bis alle Schüler zur Arbeit bereit sind.
- Einem oder mehreren Schüler/n fehlt etwas, um mit der Arbeit anfangen zu können (fehlendes Material oder langes Suchen von Material).
- Im Unterricht wird viel Zeit mit anderen Sachen vertrödelt, z. B. dem Klären von organisatorischen, nicht lernrelevanten Inhalten.
- Während Gruppen- oder Stillarbeitsphasen erfolgen zusätzliche Erklärungen und das Nachschieben von Informationen durch die Lehrperson, was auf keine gute Vorbereitung hindeutet.
- Auch das Tafelbild lässt Planungsmängel vermuten: Die Lehrperson verschwendet Zeit beim Tafelbild, z. B. wischt sie einzelne Teile weg, um sie an anderer Stelle neu zu schreiben usw.
- Die Lehrperson scheint einzelne Unterrichtsabschnitte abzukürzen oder abzubrechen.
- Die Lehrperson wirkt am Stundenende hektisch.

*Antwort: Gesamteindruck*

- Eine „4“ wird vergeben, wenn die zur Verfügung stehende Zeit optimal genutzt wird und der Stundenablauf auf eine gute Unterrichts- und Zeitplanung schließen lässt.
- Bei einer „3“ wird nahezu die gesamte Unterrichtszeit effektiv genutzt. Vereinzelt sind z. B. nur kurze nicht fachliche Aktivitäten zu erkennen. Es ist gelegentlich eine kurze Leerlaufphase zu beobachten oder es sind einzelne eher geringe Planungsmängel zu bemerken (z. B. Nachschieben von Erklärungen, ungünstige Tafelaufteilung).
- Die „2“ wird dann vergeben, wenn es längere Phasen des Leerlaufs bzw. öfter Phasen (oder eine sehr lange Phase) gibt, in denen sich mit außerfachlichen Themen beschäftigt wird oder einzelne Hinweise auf größere Planungsmängel zu beobachten sind (z. B. Abbruch oder hektische Beendigung einzelner Unterrichtsphasen).
- Die „1“ wird vergeben, wenn offensichtlich Zeit verschwendet wird und wenn es lange Phasen gibt, in denen die Lehrperson außerfachliche Themen mit den Schülern diskutiert. Eine „1“ wird auch dann vergeben, wenn es mehrere Hinweise auf größere Planungsmängel gibt.

*Anmerkungen*

Es geht bei diesem Merkmal nicht darum zu beurteilen, wie aufwändig die Materialien etc. sind, sondern ob Ablauf und Inhalt des Unterrichts so geplant sind, dass möglichst viel Unterrichtszeit sinnvoll genutzt werden kann.

Es ist wichtig, eher auf Negativindikatoren – wie Zeitverschwendung oder Planungsmängel – zu achten. Diese sind vermutlich besser zu beobachten als ein Unterrichtsablauf, der auf eine angemessene Zeitnutzung und eine gute Unterrichtsplanung schließen lässt.

Der pünktliche Beginn einer Unterrichtsstunde kann teilweise nur in den staatlichen Schulen eingeschätzt werden, da es an den BIP-Schulen oftmals keine Schulklingel gibt. Sobald die Schulklingel läutet, sollte der Unterricht anfangen. Es ist darauf zu achten, ob es sich um ein Vorklingeln handelt. Beim Unterrichtsbeginn in den BIP-Klassen ist darauf zu achten, welche Intention von der Lehrperson verfolgt wird. Wenn die Lehrperson z. B. den Unterrichtsbeginn in einer Minute ankündigt, dann sollte sie auch nach ungefähr einer Minute anfangen. Trifft dies nicht zu, sondern dauert es zu Beginn der Stunde noch sehr lange, bis alle organisatorischen Sachen etc. geklärt sind, spricht das für einen unpünktlichen Beginn.

Die Auslagerung von sogenannten *non-instructional activities* ist nicht ersichtlich. Wenn keine organisatorischen Sachen im Unterricht geklärt werden, kann davon ausgegangen werden, dass diese aus dem Unterricht ausgelagert und z. B. vor Lektionsbeginn oder in der Pause geklärt wurden. Falls Pausen im Unterrichtsvideo vorhanden sind, werden diese nicht für die Ein-

schätzung herangezogen (vgl. Kapitel 12.1.4). Auflockerungs- oder Entspannungsübungen, die inhaltlich nichts mit dem Thema des Unterrichts zu tun haben, werden als *non-instructional activities* angesehen.

### 19.1.4 Übergangsmangement (*Managing Transition*) (*Ueb*)

#### *Grundidee*

Ein besonderes Merkmal des Zeitmanagements ist die Organisation der einzelnen Übergänge zwischen den Unterrichtsphasen. Die gute Organisation von Übergängen (z. B. beim Umstellen von Tischen, beim Bilden von Stuhl- bzw. Sitzkreisen oder beim Austeilen von Unterrichtsmaterialien) sorgt für einen reibungslosen Unterrichtsablauf. Das Merkmal „Übergangsmangement“ erfasst demnach die Fähigkeit der Lehrperson, die Übergänge zwischen verschiedenen Unterrichtsphasen so zu organisieren, dass sie kurz und reibungslos erfolgen und keine unnötigen Pausen entstehen. Dabei geht es darum, dass die Lehrperson mit den Kindern wiederkehrende Abläufe zwischen den Unterrichtssequenzen so organisiert hat, dass sie keiner Erklärung mehr bedürfen.

#### *Indikatoren*

- Die Übergänge verlaufen reibungslos (z. B. Übergang vom Stuhlkreis zu anderen Beschäftigungen, der Wechsel von lehrerorientiertem Unterricht zu Freiarbeitsphasen oder beim Wechsel zwischen verschiedenen Unterrichtsgegenständen).
- Die Übergänge verlaufen relativ lärmfrei.
- Die Übergänge zwischen einzelnen Unterrichtsphasen erfolgen durch knappe, eindeutige Überleitungen.
- Die Lehrperson setzt akustische oder visuelle Impulse/Signale (Triangel, Glöckchen, Zeichen mit den Fingern etc.) ein, um die Übergänge zu organisieren.
- Die Anleitungen für die Übergänge sind so logisch von der Lehrperson formuliert, dass die Schüler ihnen folgen können.
- Es scheint Routinen zu geben, die die Übergänge regeln: Die Schüler scheinen zu wissen, was von ihnen verlangt wird (z. B. „*in den Kreis kommen*“, „*in den Gruppen zusammensetzen*“), sodass Wechsel oder Übergänge ohne großen Aufwand geschehen und der Unterricht davon kaum unterbrochen wird.
- Materialien und Unterlagen, die im Laufe der nächsten Unterrichtsphase benötigt werden, liegen/stehen bereit, sodass das Verteilen an die Schüler relativ zügig abläuft. Das Austeilen oder Einsammeln ist gut organisiert und wird rasch durchgeführt, sei es von der Lehrperson selbst oder von einem Schüler/mehreren Schülern.

*Negativindikatoren*

- Die Lehrperson gibt keine eindeutigen Anweisungen zum Übergang in andere Unterrichtsphasen (z. B. vom Unterrichtsgespräch zum Sitzkreis vor der Tafel).
- Es scheint keine Routinen für die Übergänge zu geben.
- Die Lehrperson achtet in den Übergangsphasen nicht auf eine zügige Durchführung.
- Die Übergänge verlaufen nicht reibungslos. Die Lehrperson muss die Schüler häufig zurechtweisen und um eine zügige Durchführung bitten.
- Die Lehrperson achtet nicht auf eine konsequente Durchführung: Die Lehrperson macht nach einem Übergang mit dem Stoff weiter, obwohl noch nicht alle Schüler ihre Hefte aufgeschlagen haben oder im Sitzkreis angekommen sind.

*Antwort: Gesamteindruck*

- Eine „4“ wird vergeben, wenn die Übergänge optimal geplant sind und reibungslos sowie ohne Zeitverluste ablaufen. Es kann davon ausgegangen werden, dass die Lehrperson Routinen, die die Übergänge regulieren, bereits eingeführt hat.
- Bei einer „3“ sind einzelne eher geringe Mängel zu erkennen. Die Übergänge sind mitunter etwas ineffizient.
- Die „2“ wird vergeben, wenn die Übergänge wenig effizient gestaltet sind. Die Lehrperson achtet nicht immer auf eine zügige Durchführung und muss die Schüler während des Übergangs häufig zurechtweisen.
- Um eine „1“ zu vergeben, müssen beim Übergang zwischen verschiedenen Aktivitäten Zeitverluste ersichtlich sein. Die Lehrperson scheint keine klaren Übergangsregeln zu haben und macht häufig mit dem Unterrichtsstoff weiter, obwohl einzelne Schüler noch nicht im Sitzkreis sind oder noch keine Materialien haben. Die Schüler wissen häufig nicht, was sie als Nächstes machen sollen.

*Anmerkungen*

Das Merkmal „Übergangsmanagement“ wird separat zum Zeitmanagement erfasst, da besonders in der Grundschule häufige Wechsel zwischen einzelnen Aktivitäten stattfinden und Zeit in Anspruch nehmen, wenn sie von der Lehrperson nicht gut organisiert sind. Im Verlauf des Videos ist besonders auf die Situationen zu achten, in denen die Sozialform wechselt oder eine „Klassenteilung“ stattfindet (vgl. Lotz, 2013c). In der Basiskodierung sind diese Phasen als „UMBAU“ oder „SFMIX“ kodiert und beinhalten auch jene Phasen, in denen Arbeitsmaterialien ausgeteilt oder eingesammelt werden (vgl. Lotz, 2013b).

### 19.1.5 Regelklarheit bzw. -verwendung (Regl)

#### *Grundidee*

Besonders in der Grundschule werden noch viele Regeln und Verfahrensweisen (Routinen und Rituale) eingeführt und gefestigt (z. B. das Melden, bevor man einen Beitrag abgibt, oder wie man sich in Gesprächen oder im Unterricht verhält). Das Regelsystem sorgt dafür, dass die Aktivitäten in der Klasse ohne besondere Erklärungen, Anweisungen und Begründungen in Gang gesetzt und aufrechterhalten werden bzw. dass Interaktionen im Klassenzimmer reibungslos ablaufen. Dieses Merkmal erfasst, inwieweit in der Klasse Verhaltensregeln klar definiert sind/ein Regelsystem existiert, das allen Schülern bekannt ist, und ob den Schülern bei Fehlverhalten die Sanktionen bewusst sind. Außerdem erfasst dieses Merkmal, wie die Lehrkraft mit den Regelverstößen umgeht.

#### *Indikatoren*

- Die Regeln, die eingehalten werden müssen, sind allen bekannt.
- Im Unterricht ist klar, was die Schüler machen dürfen und was nicht.
- Die Regeln werden kontinuierlich gesichert und gefestigt, das heißt die Lehrerin erinnert die Schüler konsequent an die vereinbarten Regeln (sie zeigt auf die Plakate oder erinnert die Schüler daran, im Satz zu antworten etc.).
- Die Schüler wissen, welche Sanktionen auf Regelverstöße folgen.
- Die Regeln sind sichtbar im Klassenraum angebracht (Plakate, Poster oder Ähnliches).
- Die Schüler werden bei Regelverstößen von der Lehrkraft darauf aufmerksam gemacht.
- Der Regelverstoß wird geahndet, indem die Lehrperson kurz auf die vereinbarten Regeln verweist (z. B. „Paul, wir hatten doch eine Abmachung!“). Es kommt nicht zu größeren Unterbrechungen aufgrund von Regelverstößen.

#### *Negativindikatoren*

- Die Regeln in der Klasse scheinen sich dauernd zu ändern, sie scheinen widersprüchlich.
- Es scheint eine Willkür in der Umsetzung aufseiten der Lehrperson zu geben, z. B. sanktioniert die Lehrperson manchmal Zwischenrufe ohne Meldungen, mal ignoriert sie sie, mal arbeitet sie sogar mit den hineingerufenen Antworten.
- Klare Regeln scheinen zu fehlen. Deswegen kann es häufig zu Störungen und Unterbrechungen kommen, weil Schüler in bestimmten Situationen nicht wissen, was von ihnen erwartet wird.

- Die Lehrperson muss aufgrund von Regelverstößen den Unterricht unterbrechen.

*Antwort: Gesamteindruck*

- Eine „4“ wird vergeben, wenn es ein festes System aus Regeln gibt, das von den Schülern verstanden und akzeptiert wird. Die Lehrperson achtet auf die Einhaltung der Regeln. Bei Störungen und Unterbrechungen reicht es aus, dass die Lehrperson nur kurz an die Regeln erinnert, es kommt nicht zu größeren Unterrichtsunterbrechungen. Die Schüler scheinen genau zu wissen, wie sie sich in bestimmten Situationen zu verhalten haben.
- Eine „3“ wird vergeben, wenn die Regeln den Schülern zwar bekannt sind und akzeptiert werden, die Lehrperson jedoch oft an die Regeln erinnern muss. Es kommt zu kleineren Unterbrechungen aufgrund von Regelverstößen.
- Eine „2“ wird vergeben, wenn zwar Regeln existieren, die Lehrperson jedoch nicht konsequent auf deren Einhaltung achtet und mit Regelverstößen nicht konsequent umgeht. Regelverstöße haben größere Unterbrechungen zur Folge.
- Eine „1“ wird vergeben, wenn klare Regeln fehlen bzw. die Lehrperson nicht auf ihre Einhaltung achtet. Es kommt daher öfter zu Unterbrechungen, wobei die Störungen und Unterbrechungen so stark sind, dass sie eine reibungslose Klassenführung unmöglich machen. Die Regeln und die Umsetzung erscheinen oft relativ willkürlich. Die Schüler scheinen die Regeln noch nicht verinnerlicht zu haben oder akzeptieren sie nicht, es kommt zu Unmutsäußerungen oder Widersprüchen.

*Anmerkungen*

Ein Hinweis auf das Vorhandensein von Regeln ist, dass Interaktionen im Klassenzimmer reibungslos ablaufen und Aktivitäten in der Klasse ohne besondere Erklärungen, Anweisungen und Begründungen in Gang gesetzt und aufrechterhalten werden. Ein guter Umgang mit eingeführten Regeln zeigt sich z. B., wenn sich die Lehrperson nicht mit halbem Erfolg zufrieden gibt, sondern den Unterricht erst fortsetzt, wenn wirklich alle Schüler still sind und den Blick zur Tafel richten, also nicht, wenn es nur etwas leiser geworden ist (Wahl, Hu-ber & Weinert, 2007).

Beispiele für Regeln sind: „*Erst überlegen und dann melden.*“, „*Flüstern mit eurem Partner ist erlaubt.*“ oder „*Wer die Antwort weiß, meldet sich, ohne Zwischenrufe.*“ Es kann aber auch geregelt sein, wie die „*Stationsarbeit*“ abläuft oder wann die Kinder auf die Toilette gehen dürfen. Zudem gibt es Signale oder Rituale, die z. B. die Unterrichtslautstärke regulieren: Die Lehrperson ersetzt mehrfache und verbale Äußerungen durch Gesten. Beispielsweise

wird ein Glöckchen geläutet, wenn es der Lehrkraft zu laut oder zu unruhig in der Klasse ist, oder der Finger wird auf den Mund gelegt, damit die Klasse wieder ruhiger wird. – Unklare Regeln können sich in häufiger Unruhe und allgemeinem Durcheinander äußern.

### 19.1.6 Störungsfreiheit (Disz)

#### *Grundidee*

Dieses Merkmal soll erfassen, inwieweit der Unterricht störungsfrei abläuft und nicht immer wieder durch größere oder kleinere Störungen unterbrochen oder beeinträchtigt wird, sodass er nicht im geordneten Rahmen durchgeführt werden kann.

#### *Indikatoren*

- Der Unterricht läuft ohne Störungen ab.
- Der Unterricht ist so geplant und organisiert, dass Disziplinstörungen nicht vorkommen.
- Die Lautstärke während der Stunde ist angemessen. Was angemessen ist, hängt von der jeweiligen Unterrichtsphase ab (während Einzelarbeitsphasen und Phasen des öffentlichen Unterrichtsgesprächs sollte sie eher niedrig sein, während Gruppenarbeitsphasen oder „Stationsarbeit“ kann sie unter Umständen relativ hoch sein).

#### *Negativindikatoren*

- Im Unterricht wird häufig Unsinn gemacht (z. B. Umherwerfen von Gegenständen, Schüler rempeln sich gegenseitig an, machen Grimassen hinter dem Rücken der Lehrperson).
- Der Unterricht wird stark gestört (z. B. durch andauerndes lautes Reden der Schüler).
- Die Lehrperson muss häufig geradezu brüllen, um gehört zu werden.
- Im Unterricht ist es laut und alles geht durcheinander.
- Die Lehrperson muss wiederholt zur Ruhe mahnen (z. B. wiederholtes „schsch...“, „Psst...“, Aufrufen von störenden Schülern).
- Die Lehrperson muss die Schüler zur Arbeit auffordern, weil sie sonst anderen störenden Nebenbeschäftigungen nachgehen.

*Antwort: Gesamteindruck von „keine Störungen“ bis „große Disziplinprobleme“*

- Eine „4“ wird vergeben, wenn der Unterricht weitgehend störungsfrei abläuft.
- Eine „3“ wird vergeben, wenn mehrere kleinere Störungen auftreten, die den Unterrichtsablauf aber nicht beeinflussen.

- Eine „2“ wird vergeben, wenn einzelne massive Störungen oder häufigere kleinere Störungen erkennbar sind, die den Unterrichtsablauf beeinflussen.
- Eine „1“ wird vergeben, wenn häufig massive Unterrichtsstörungen bzw. große Disziplinprobleme auftreten. Dies äußert sich besonders in unübersichtlich-chaotischen und lauten Situationen, in denen die Lehrperson Mühe hat sich durchzusetzen.

### *Anmerkungen*

Unterrichtsstörungen können sowohl durch einzelne Schüler als auch durch kleinere Gruppen von Schülern verursacht werden. Es geht bei diesem Merkmal nur um die Einschätzung der Häufigkeit von kleineren und größeren Unterrichtsstörungen.

### *19.1.7 Effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)*

#### *Grundidee*

Bei dem folgenden Merkmal geht es um den Teil der Klassenführung, der sich auf den Umgang mit Disziplin- und Unterrichtsstörungen bezieht. Ziel der Lehrperson sollte es sein, mit Störungen so umzugehen, dass der Unterrichtsfluss nicht unterbrochen und das Unterrichtsklima nicht beeinträchtigt wird.

#### *Indikatoren*

- Die Lehrperson erkennt Störungen frühzeitig und „erstickt sie im Keim“ bzw. unterbricht sie sofort. Die Lehrperson hat die Störungen unter Kontrolle und geht mit ihnen angemessen um.
- Harmlose Aktivitäten, wie z. B. lautes Lachen, Reden, Tagträumen, Nichtbefolgen von Anweisungen etc., die den Unterrichtsfluss stören, versucht die Lehrkraft sofort zu unterbinden bzw. thematisiert sie nicht weiter.
- Kleine Störungen werden von der Lehrperson ganz ignoriert oder nebenbei nicht verbal, sondern durch ein Zeichen, das der Schüler gleich versteht, beendet (z. B. Finger auf den Mund, Handzeichen für „*Mund zu, Ohren auf*“ etc.).
- Die Lehrperson nutzt subtile verbale Aufmerksamkeitslenkungen, wie z. B. das Aufrufen bzw. das Nennen des Namens, den Einbezug in das Unterrichtsgeschehen oder ein kurzes Ermahnen, um Unterrichtsstörungen zu beheben. Das Aufrufen des Namens etc. hat nicht nur kurzfristigen Erfolg.
- Um Störungen durch Leerlaufzeiten zu vermeiden, gibt es Differenzierungsaufgaben, sodass die Schüler wissen, was sie machen können, wenn

sie mit einer Aufgabe fertig sind (z. B. gibt es „Sternchenaufgaben“ für diejenigen Schüler, die schneller fertig sind).

- Die Zurechtweisungen und Ermahnungen der Lehrperson haben nicht nur eine kurzfristige Wirkung.

#### *Negativindikatoren*

- Die Lehrperson nutzt disziplinarische Maßnahmen, um die Störung zu unterbinden.
- Die Lehrperson muss häufig oder massiv eingreifen, um Unterrichtsstörungen zu unterbinden.
- Die Lehrperson greift erst ein, wenn es zu ernsthaften Störungen gekommen ist bzw. sogar zu chaotischen Zuständen.
- Die Lehrperson unterbricht häufig sehr lange den Unterricht, um die Klasse wieder zur Ruhe zu bringen.
- Die Mahnungen oder Versuche der Lehrperson, die Klasse zum ruhigen Arbeiten in Gruppen- oder Einzelarbeitsphasen zu bewegen, haben in den meisten Fällen keine oder nur sehr kurzzeitige Wirkung.

#### *Antwort: Gesamteindruck*

- Die „4“ wird nur dann vergeben, wenn bei Unterrichtsstörungen eine frühe Intervention durch die Lehrperson zu erkennen ist und somit von Vorkehrungsmaßnahmen gegen weitere Disziplinstörungen ausgegangen werden kann. Eine „4“ wird immer auch dann vergeben, wenn der Unterricht störungsfrei abläuft und der Umgang mit Störungen in diesem Fall nicht sichtbar wird.
- Die „3“ wird vergeben, wenn die Lehrperson gelegentlich Störungen nicht sofort erkennt oder ab und zu etwas unangemessen auf Unterrichtsstörungen reagiert. Jedoch zeigt das Eingreifen sofort eine Wirkung.
- Die „2“ muss vergeben werden, wenn die Lehrperson öfter Unterrichtsstörungen zu spät erkennt und ab und zu unangemessen auf die Unterrichtsstörungen reagiert. Das Eingreifen zeigt in manchen Situationen keine sofortige Wirkung.
- Um eine „1“ zu vergeben, sollte die Lehrperson häufig eingreifen und disziplinieren müssen. Ihr Eingreifen hat jedoch häufig keine langfristige Wirkung.

### *19.1.8 Vorbereitung des Klassenraums (Vorb)*

#### *Grundidee*

Das Merkmal erfasst, inwieweit der Klassenraum effektiv für ein erfolgreiches Lernen eingerichtet ist und somit Störungen durch herumstehende Ge-

genstände etc. vermieden werden. Für die Lehrperson muss die Raumgestaltung einen „whole-class-Fokus“ (Kounin, 1976; 2006) ermöglichen. Es geht bei diesem Merkmal hauptsächlich darum einzuschätzen, ob alle Schüler einer Klasse die Möglichkeit haben, am Unterrichtsgeschehen teilzunehmen.

#### *Indikatoren*

- Das Lehrerpult steht für die Schüler ersichtlich im Klassenraum.
- Die Schülertische sind so ausgerichtet, dass jeder Schüler einen guten Blick zur Tafel bzw. zum Overheadprojektor hat (das heißt, dass Schüler sich nicht ständig umdrehen müssen, um einen guten Blick zur Tafel zu haben usw.).
- Der Overheadprojektor hat im Klassenraum einen festen Platz und stört niemanden.
- Jeder Schüler hat für seine Unterrichtsmaterialien, wie z. B. Ordner, Malkasten, Zeichenblöcke oder Bücher, seinen eigenen Platz (Schublade, Fach).
- Die Regale/Schränke im Klassenraum sind so positioniert, dass sie niemanden stören. Sie sind leicht zugänglich.
- Zwischen den Raumobjekten (Tisch, Schränke etc.) ist genügend Platz zum Durchlaufen.

#### *Antwort: Gesamteindruck*

- Eine „4“ wird vergeben, wenn der Klassenraum so gestaltet ist, dass Störungen so weit wie möglich vermieden werden. Alle Schüler sind in der Lage, die Lehrkraft und die Tafel zu sehen bzw. können alles andere, was für den Unterricht benötigt wird, schnell und störungsfrei erreichen.
- Eine „3“ wird vergeben, wenn zwar alle Schüler einen guten Blick zur Tafel haben, jedoch einzelne Raumobjekte nicht optimal für alle sichtbar und schnell erreichbar sind und somit Störungen verursacht werden können.
- Eine „2“ wird vergeben, wenn trotz einer guten Raumaufteilung einzelne Schüler keinen guten Blick zur Tafel bzw. zum Overheadprojektor haben und sich dadurch häufig umdrehen müssen.
- Die „1“ wird vergeben, wenn die Anordnung der Tische nicht so ausgerichtet ist, dass die Tafel oder die Lehrperson im Blick von allen Schülern ist.

#### *Anmerkungen*

Die Einschätzung dieses Merkmals kann relativ frühzeitig getroffen werden. Es wird bereits zu Beginn der Lektion ersichtlich, ob die Indikatoren zutreffen oder nicht. Zu diesem Merkmal zählt nicht die Vorbereitung des Unterrichtsmaterials (z. B. Arbeitsblätter usw.). Es geht nicht darum einzuschät-

zen, ob der Klassenraum groß ist und dadurch mehr Platz zur Verfügung steht, sondern darum, dass auch in kleineren Klassenzimmern eine adäquate Raumaufteilung (ohne viele störende Geräte etc.) möglich ist.

## 19.2 Merkmale des Unterrichtsklimas

In Tabelle 38 sind die Merkmale des Unterrichtsklimas dargestellt. Diese wurden entwickelt, um sowohl Aspekte der Lehrer-Schüler-Beziehungen und der Schüler-Schüler-Beziehungen als auch allgemeine Aspekte der Unterrichts Atmosphäre zu erfassen. In Tabelle 38 ist zudem dokumentiert, auf welche Quellen bei der Entwicklung der Merkmale Bezug genommen wurde. Das hoch inferente Ratingsystem zur Erfassung des Unterrichtsklimas in der Grundschule umfasst insgesamt zehn Merkmale, die anschließend beschrieben werden.

Tabelle 38: Merkmale des hoch inferenten Ratings zum Unterrichtsklima (Quellen)

Merkmale	Quellen
humorvolle Lernatmosphäre (Humo)	– Powell & Anderson (1985), adaptiert – Reißland (2002), adaptiert – Helmke et al. (2007), adaptiert
Anerkennung der Schüler durch die Lehrperson (Anes)	– Rakoczy & Pauli (2006), adaptiert
gegenseitige Anerkennung der Schüler (Aneg)	– Eigenentwicklung in Anlehnung an v. Saldern & Littig (1987, LASSO: „Hilfsbereitschaft von Mitschülern“ und „Zufriedenheit mit den Mitschülern“)
Fürsorglichkeit, Herzlichkeit und Wärme (Fuer)	– v. Saldern & Littig (1987, LASSO: „Fürsorglichkeit des Lehrers/der Lehrerin“), adaptiert – Gruehn (2000, „Sozialorientierung des Lehrers“), adaptiert – Helmke et al. (2007), adaptiert
Einsatz von Lob (Lob)	– Eigenentwicklung in Anlehnung an Hofer (1985)
Ermutigung durch die Lehrperson (Emut)	– Eigenentwicklung in Anlehnung an Brophy (1981) und O’Leary & O’Leary (1977)
sachlich-konstruktive Rückmeldung (Rück)	– Rakoczy & Pauli (2006), adaptiert
positiver Umgang der Lehrperson mit Fehlern (Fehl)	– Rakoczy & Pauli (2006), adaptiert – Helmke et al. (2007), adaptiert – Schulmeiß, Seidel & Meyer (2003), adaptiert – Ramm et al. (2006, PISA 2003), adaptiert
positives Fehlerklima zwischen den Schülern (Fekl)	– Helmke et al. (2007), adaptiert
Wettbewerb und Leistungsdruck (Wett)	– Rakoczy & Pauli (2006), adaptiert – v. Saldern & Littig (1987, LASSO), adaptiert

### 19.2.1 Humorvolle Lernatmosphäre (Humo)

#### Grundidee

Das Merkmal erfasst, inwieweit die Lernatmosphäre locker und entspannt ist und ob ein humorvolles Klima herrscht. Humor gilt als Merkmal positiven Lernklimas, kann aber auch von der Lehrperson als Technik der Verhaltenskontrolle bzw. Führungsstrategie eingesetzt werden (Powell & Anderson, 1985). Humor kann in konkreten Unterrichtssituationen stressvolle Situationen entschärfen, Aufmerksamkeit aktivieren und das Behalten von Unterrichtsinhalten fördern.

#### Indikatoren

- In der Stunde wird häufig gelacht, jedoch nicht übertrieben oft.
- Die Lehrperson lockert den Unterricht mit Humor auf.
- Die Lehrperson nimmt sich nicht immer uneingeschränkt ernst.
- Die Lehrperson macht in angemessenem Maße Späße und Witze.
- Die Lehrperson reagiert auch auf Disziplinprobleme mit Humor, das heißt die Lehrperson nutzt Humor, um unerwünschtes Verhalten zu unterbinden. Das bedeutet, die Missbilligung erfolgt auf humorvolle, aber nicht beschämende Weise.
- Die Lehrperson gibt witzige Beispiele.
- Die Lehrperson reagiert auch auf Witze der Schüler positiv. Sie lacht mit den Schülern.
- Es herrscht eine mittlere Ausprägung zwischen den Extremen: (a) humorfrei, trocken, gespannt, ernst und (b) ausgelassen, Feuerwerk von Witzen, exzessiv humorvoll.

#### Negativindikatoren

- Die Schüler scheinen die Witze der Lehrperson nicht zu verstehen bzw. der Humor der Lehrperson kommt bei den Schülern scheinbar nicht an.
- Die Lehrperson verwendet herabsetzenden Humor, der verletzend wirkt.
- Die Lehrperson ist ausschließlich sachbezogen, nüchtern und ernst. Sie scheint keinen Spaß zu verstehen.

#### Antwort: Gesamteindruck

- Eine „4“ wird vergeben, wenn eine humorvolle Atmosphäre herrscht, in der im angemessenen Maße gelacht wird. Die Lehrperson nimmt sich nicht immer uneingeschränkt ernst und reagiert auch auf lustige Schülerbeiträge mit Humor. Eine „4“ wird nur vergeben, wenn die Lehrperson (zusätzlich zu den eben genannten Indikatoren) den Unterricht mit lustigen Beispielen auflockert bzw. auch unerwünschtes Verhalten mit Humor unterbindet.

- Eine „3“ wird vergeben, wenn die Atmosphäre humorvoll ist, die Lehrperson jedoch keine zusätzlichen humorvollen Beispiele einsetzt oder Disziplinstörungen nicht humorvoll unterbindet.
- Eine „2“ wird vergeben, wenn die Lernatmosphäre eher trocken ist, jedoch die Lehrperson den Unterricht gelegentlich durch Humor auflockert. Die Lehrperson scheint von sich aus nicht humorvoll, nutzt aber Gelegenheiten, die ihr Schüler bieten, um humorvoll zu reagieren.
- Eine „1“ wird vergeben, wenn die Atmosphäre humorfrei und trocken ist, die Lehrperson also ausschließlich sachbezogen, nüchtern und ernst ist. Die Lehrperson nutzt auch Gelegenheiten durch Schülerkommentare usw. nicht, um den Unterricht humorvoll aufzulockern.

#### *Anmerkungen*

Der Humor der Lehrperson muss nicht unbedingt auch für den Rater witzig sein. Daher sollte bei der Beurteilung besonders auf die Indikatoren geachtet werden. Dabei ist Humor nicht identisch mit Freundlichkeit oder persönlichem Interesse, sondern relativ unabhängig davon.

### *19.2.2 Anerkennung der Schüler durch die Lehrperson (Anes)*

#### *Grundidee*

Ein zentraler Aspekt der sozialen Eingebundenheit und des Zugehörigkeitsgefühls von Schülern ist die grundlegende Wertschätzung, die ihnen von der Lehrperson entgegengebracht wird, und der respektvolle Umgang zwischen Lehrperson und Schülern. Aus diesem Grund erfasst das Merkmal, inwiefern die Lehrperson den Schülern mit Wertschätzung begegnet. Die Lehrperson nimmt die Schüler als Personen ernst und behandelt sie respektvoll. Dieses Merkmal findet besonders Beachtung in Situationen, in denen die Lehrperson mit fehlerhaften Äußerungen von Schülern oder auftretenden Schwierigkeiten im Unterricht umgeht.

#### *Indikatoren*

- Die Lehrperson zeigt den Schülern, dass sie ihnen zutraut, die Inhalte zu verstehen und bei ausreichender Anstrengung die Aufgaben zu lösen (Beispiel: „*Ich bin sicher, ihr findet das heraus!*“ oder „*Du schaffst das auf jeden Fall!*“).
- Die Lehrperson zeigt den Kindern durch positiv gestimmtes Nachfragen, dass sie sie ernst nimmt (Beispiel „*Aha, und wie bist du darauf gekommen?*“).
- Die Lehrperson interessiert sich für die Perspektive und Meinung der Schüler, sie greift sie auf.

- Die Lehrperson nimmt die Gefühle der Schüler ernst und akzeptiert auch, wenn sie ihre Gefühle nicht zeigen wollen.
- Die Lehrperson kritisiert angemessen, ohne die Schüler einzuschüchtern.
- Die Lehrperson geht respektvoll, wohlwollend und ohne Sarkasmus / Zynismus mit fehlerhaften Äußerungen oder auftretenden Schwierigkeiten um. Scherzhafte Bemerkungen oder kleine Späße können durchaus vorkommen, wichtig sind aber Ton und Kontext.

#### *Negativindikatoren*

- Die Schüler werden von der Lehrperson im Klassenraum lächerlich gemacht, beschämt oder bloßgestellt.
- Die Lehrperson behandelt die Schüler herabsetzend, ist verletzend oder kränkt Schüler.
- Die Lehrperson zeigt Gesten der Verächtlichmachung (z. B. wegwerfende Handbewegungen).

#### *Antwort: Gesamteindruck*

- Eine „4“ wird vergeben, wenn durchwegs ein respektvoller Umgang zwischen der Lehrperson und den Schülern herrscht. Die Lehrperson ist den Schülern gegenüber besonders wohlwollend eingestellt und interessiert sich sehr für deren Meinungen und Perspektiven.
- Eine „3“ wird vergeben, wenn die Lehrperson die Perspektiven und Meinungen der Schüler im Großen und Ganzen respektiert und ermutigende Worte findet, es jedoch gelegentlich Zeichen geringen Respekts gibt.
- Eine „2“ wird vergeben, wenn die Lehrperson den Schülern nicht alles zutraut und auch manchmal über die Meinungen und Perspektiven der Schüler hinweggeht.
- Eine „1“ wird vergeben, wenn die Lehrperson die Schüler mehr als einmal herablassend und mit mangelndem Respekt behandelt bzw. die Schüler lächerlich macht. Dazu zählen zum Beispiel Gesten der Verächtlichmachung (z. B. wegwerfende Handbewegungen). Die Lehrperson greift die Meinungen und Perspektiven ihrer Schüler häufig nicht auf.

### *19.2.3 Gegenseitige Anerkennung der Schüler (Aneg)*

#### *Grundidee*

Das Merkmal erfasst, inwieweit sich auch die Schüler gegenseitig positiv wertschätzen und soziale Eingebundenheit erleben. Die Schüler nehmen sich gegenseitig als Personen ernst und behandeln sich respektvoll. Auch die Hilfsbereitschaft und Kameradschaftlichkeit zwischen den Schülern drückt sich in diesem Merkmal aus.

*Indikatoren*

- Die Schüler lachen einander nicht aus.
- Die Schüler reagieren auf fehlerhafte Äußerungen nicht mit Abwertung oder Diskriminierung.
- Die Schüler helfen einander.
- Die Schüler trösten sich gegenseitig.
- Es gibt keine Rivalität innerhalb der Klasse.
- Die Schüler loben einander.
- Die Schüler lassen die Anderen ausreden.

*Antwort: Gesamteindruck*

- Eine „4“ wird vergeben, wenn durchwegs ein respektvoller Umgang zwischen den Schülern bzw. ein Klima der Hilfsbereitschaft herrscht.
- Eine „3“ wird vergeben, wenn die Schüler im Großen und Ganzen respektvoll miteinander umgehen und nur vereinzelt Rivalitäten etc. zwischen den Schülern zu beobachten sind.
- Eine „2“ wird vergeben, wenn die Schüler einander nicht immer respektieren, auf fehlerhafte Äußerungen teilweise abfällig reagieren oder mehrmals einander auslachen.
- Eine „1“ wird vergeben, wenn die Schüler während der ganzen Stunde einander nicht respektieren, sich gegenseitig auslachen und eine große Rivalität zwischen den Schülern vorherrscht.

*19.2.4 Fürsorglichkeit, Herzlichkeit und Wärme (Fuer)**Grundidee*

Das Merkmal erfasst, inwieweit die Lehrperson eine persönlich wichtige Person für die Schüler ist. Es soll eingeschätzt werden, inwieweit die Lehrperson fürsorglich ist und ein Ohr für persönliche und private Belange der Schüler hat sowie Interesse an den Problemen ihrer Schüler zeigt. Das Merkmal erfasst zusätzlich, inwieweit die Lehrperson im Umgang mit den Schülern durch Freundlichkeit, Herzlichkeit, Einfühlungsvermögen und Wärme gekennzeichnet ist. Zuwendung, körperliche Nähe und Berührung werden von jungen Kindern als Anreiz und Belohnung erfahren und zählen zu den sozialen Belohnungen (Reimann, 2004). Besonders in der Grundschule spielt dieses Merkmal eine große Rolle.

*Indikatoren*

- Die Lehrperson ist nicht nur Vermittler fachlichen Wissens, sondern sie hat auch ein offenes Ohr für außerschulische / nicht fachliche (persönliche) Probleme der Schüler.

- Die Lehrperson nimmt sich die Zeit, wenn Schüler etwas mit ihr bereden wollen.
- Die Lehrperson ist allen Schülern gegenüber freundlich, sie lächelt viel.
- Die Lehrperson geht auf die Gefühle ihrer Schüler ein.
- Die Lehrperson zeigt Einfühlungsvermögen und Wärme. Dies ist besonders bei Einzelzuwendungen ersichtlich, in denen die Schüler die Nähe der Lehrperson suchen.
- Der Umgangston der Lehrperson ist freundlich.

*Antwort: Gesamteindruck*

- Eine „4“ wird vergeben, wenn in der Klasse stets ein freundlicher Umgang zwischen der Lehrperson und den Schülern herrscht. Die Lehrperson zeigt persönliches Interesse an ihren Schülern und ist durchwegs durch Wärme und Herzlichkeit gekennzeichnet.
- Eine „3“ wird vergeben, wenn die Lehrperson überwiegend freundlich und interessiert ist.
- Bei einer „2“ wirkt die Lehrperson manchmal distanziert und abwesend, den Belangen der Schüler gegenüber nicht aufgeschlossen.
- Eine „1“ wird vergeben, wenn die Lehrperson sehr distanziert und abwesend scheint. Die Lehrperson beschränkt den Umgang mit den Schülern ganz nüchtern auf das rein Fachliche, zeigt keinerlei Wärme und Herzlichkeit.

*Anmerkungen*

Beim Rating dieses Merkmals ist auch auf nonverbale Verhaltensweisen zu achten, z. B. Lächeln, ermutigendes und zustimmendes Auf-die-Schulter-Klopfen etc. Dieses Verhalten ist häufig während Einzelzuwendungen der Lehrperson zu beobachten.

### 19.2.5 Einsatz von Lob (Lob)

*Grundidee*

Dieses Merkmal erfasst, ob die Lehrperson ihre Schüler lobt. Nach Hofer (1985) wird unter Lob im Allgemeinen der Ausdruck von Zustimmung und Billigung des Verhaltens eines anderen verstanden. Dabei geht Lob „über die schiere Rückmeldung der Richtigkeit hinaus und enthält eine positive gefühlsmäßige Komponente der Freude“ (Hofer, 1985, S. 416). Besonders in der Grundschule wird das Lob nicht nur eingesetzt, um die Leistungen der Schüler, sondern auch um soziale Aktionen bzw. positives Verhalten der Schüler zu würdigen.

*Indikatoren*

- Die Lehrperson lobt die Leistungen bzw. das Verhalten einzelner Schüler bzw. der ganzen Klasse öffentlich.
- Die Lehrperson äußert Freude über die Leistungen bzw. das Verhalten der Klasse bzw. einzelner Schüler („*Ja richtig. Gib mir 5!*“, „*Super!*“, „*Toll gemacht*“ etc.).
- Die Lehrperson lobt die ganze Klasse am Ende der Stunde für die gute Mitarbeit etc.
- Die Lehrperson lobt, wenn die Schüler zu Stundenbeginn schnell auf ihren Plätzen sind und die Stunde pünktlich beginnen kann.
- Die Lehrperson lobt die Schüler, wenn sie sich an vereinbarte Regeln halten etc.
- Die Lehrperson lobt die Schüler, wenn sie mit dem Aufräumen ihres Platzes etc. als erstes fertig sind.

*Antwort*

- „4“ = Die Lehrperson lobt sehr häufig.
- „3“ = Die Lehrperson lobt häufig.
- „2“ = Die Lehrperson lobt weniger häufig.
- „1“ = Die Lehrperson lobt nicht häufig bis gar nicht.

*Anmerkungen*

Hierzu zählen keine materiellen Belohnungen, wie das Austeilen von Süßigkeiten am Ende der Stunde oder als Belohnung auf eine richtige Antwort.

Bei diesem Merkmal geht es noch nicht darum, sachlich-konstruktive Rückmeldungen (also Rückmeldung mit Bezug auf die erbrachte Leistung) einzuschätzen. Es soll lediglich erfasst werden, ob Lob sehr häufig, häufig, weniger häufig oder nicht häufig vorkommt.

*19.2.6 Ermutigung durch die Lehrperson (Emut)**Grundidee*

Damit sich ein Lob positiv auf die Schüler auswirkt, sollte dem Lob eine inhaltliche Erläuterung folgen (z. B. was an der Leistung besonders gut ist oder warum die Leistung als lobenswert betrachtet werden kann). Der einzelne Schüler bzw. die Klasse sollte das Lob auf das eigene Verhalten oder die eigene Leistung zurückführen können. Damit ein Lob ermutigt, muss die Lehrperson Einzelheiten des Verhaltens ansprechen. Die Schüler sollten das Gefühl haben, dass die Lehrperson das Lob aufrichtig und ernst meint und nicht bei jedem Schüler gleich einsetzt.

*Indikatoren*

- Gegenstand des Lobes kann sowohl eine erbrachte Leistung als auch ein gezeigtes Verhalten sein.
- Die Lehrperson lobt nicht nur, sondern ermutigt die Schüler bzw. die Klasse auch (direkt oder indirekt), beim nächsten Mal ein ähnliches Verhalten zu zeigen bzw. so weiter zu machen.
- Dem Lob folgt eine Ermutigung (z. B. *„Diese Aufgabe hast du aber toll gemacht, die nächste schaffst du auch noch.“*).

*Negativindikatoren*

- Die Lehrperson differenziert in ihrem Lob nicht bzw. wenig. Die Lehrperson sagt zu jedem Kind das Gleiche: *„Super!“*, *„Klasse!“*, *„Prima!“*, *„Toll gemacht!“*
- Das Lob ist nicht an eine Leistung oder ein Verhalten gebunden.
- Das Lob drückt nur Freude der Lehrperson aus, ohne Bezug zu Leistungen oder dem Verhalten einzelner Schüler.
- Die Lehrperson ermutigt die Schüler nicht zusätzlich.

*Antwort: Gesamteindruck*

- Eine „4“ wird vergeben, wenn das eingesetzte Lob oft einen inhaltlichen Bezug hat und den Schüler oder die Klasse zum Weiterarbeiten usw. ermutigt. Es muss eine eindeutige Freude über das gezeigte Verhalten oder die erbrachte Leistung bei der Lehrperson erkennbar sein.
- Eine „3“ wird vergeben, wenn das Lob zwar im Großen und Ganzen einen inhaltlichen Bezug hat und die Lehrperson erfreut ist, jedoch selten bis nie eine weitere „Ermutigung durch die Lehrperson“ ersichtlich wird.
- Eine „2“ wird vergeben, wenn die Lehrperson das Lob ausschließlich dazu benutzt, ihre Freude auszudrücken. Das Lob ist selten an einen Inhalt gebunden bzw. ermutigt die Schüler nicht zusätzlich.
- Eine „1“ wird vergeben, wenn das Lob undifferenziert wirkt und keinen inhaltlichen Bezug hat bzw. keine „Ermutigung durch die Lehrperson“ stattfindet.

*Anmerkungen*

Bei diesem Merkmal geht es nicht um die Einschätzung sachlich-konstruktiver Rückmeldungen, sondern darum, ob das Lob der Lehrperson (egal wie häufig es vorkommt) eher undifferenziert eingesetzt wird (*„Super!“* oder *„Gut gemacht!“*) oder einen inhaltlichen Bezug haben. Zusätzlich wird mit diesem Merkmal erfasst, inwieweit das inhaltsbezogene Lob an Worte des positiven Zutrauens gebunden ist (bzw. ermutigende Worte). Lehrpersonen, die häufig inhaltsbezogenes Lob einsetzen und gleichzeitig die Schüler er-

mutigen, sollten im Gesamteindruck mindestens eine „3“ erreichen, bei sehr häufiger Ermutigung nach einem inhaltsbezogenen Lob sogar eine „4“.

Beim Ansehen der Videos ist nicht nur auf die Aussprache eines inhaltsbezogenen Lobs zu achten (beispielsweise: „3 mal 3 = 9, die Aufgabe ist richtig. Super gemacht!“). Der inhaltliche Bezug wird auch dann deutlich, wenn die Lehrperson z. B. auf das Arbeitsblatt eines Schülers zeigt und „Toll gemacht!“ sagt. Der Schüler weiß in diesem Moment, worauf sich das Lob bezieht, auch wenn es für den Beobachter nicht sichtbar wird.

### 19.2.7 Sachlich-konstruktive Rückmeldung (Rück)

#### Grundidee

Um Lernende zu motivieren und ihnen gleichzeitig die Möglichkeit zu geben, ihr Vorgehen zu verbessern, müssen Rückmeldungen sachlich und konstruktiv formuliert sein. Konstruktive Rückmeldungen zeichnen sich dadurch aus, dass sie korrigierende und zukunftsgerichtete Hilfestellungen geben. Eine sachliche Formulierung der Hilfestellung bedeutet, dass sich die Rückmeldung auf die Lerninhalte an sich und nicht auf die Person bezieht. Diese sachlich-konstruktiven Rückmeldungen sollten in einem wohlwollenden Ton formuliert sein.

#### Indikatoren

- Die Lehrperson gibt nur sachliche Rückmeldungen und Kommentare.
- Auch wenn auf Fehler oder Mängel hingewiesen wird, sind die Rückmeldungen wohlwollend formuliert.
- Die Rückmeldungen der Lehrperson zeigen dem Schüler, wo sein Fehler liegt und geben Hilfestellungen, wie das Vorgehen verbessert werden kann.
- Die Lehrperson zeigt eine lobende Förderhaltung (z. B. durch lobende Äußerungen, das Äußern von Freude über die Leistungen der Klasse/der einzelnen Schüler).

#### Negativindikatoren

- Die Lehrperson gibt unsachliche Rückmeldungen oder Kommentare (Beispiele: „Na, auch schon wach?“ oder „Du hast wohl die letzte halbe Stunde geschlafen?“).
- Es erfolgen auch Rückmeldungen, ohne dass von der Lehrperson weitere Unterstützung geleistet wird. Die Lehrperson sagt nur, was falsch / richtig ist oder welcher Teil der Aufgabe etc. falsch / richtig ist. Die Art und der Inhalt des Fehlers werden von der Lehrperson nicht angesprochen. Sie leistet bei falschen Antworten keine weitere Unterstützung.

*Antwort: Gesamteindruck*

- Eine „4“ wird vergeben, wenn im gesamten Unterrichtsverlauf eine wohlwollende Förderhaltung dominiert.
- Eine „3“ wird vergeben, wenn die Lehrperson zwar durchgehend „sachlich-konstruktive Rückmeldung“ gibt, jedoch im Gesamteindruck keine lobende Förderhaltung dominiert.
- Eine „2“ wird vergeben, wenn sich sachlich-konstruktive und unsachliche Rückmeldungen die Waage halten (Begründung: Unsachliche Rückmeldungen wiegen schwerer, weil die Kinder in solchen Situationen nicht wissen, woran sie sind.). Es liegt keine lobende Förderhaltung vor.
- Eine „1“ wird vergeben, wenn die Rückmeldungen der Lehrperson weder sachlich noch konstruktiv sind. Rückmeldungen wirken pauschalisiert („Richtig“ oder „Falsch“) und es ist kein inhaltlicher und damit konstruktiver Bezug erkennbar.

*Anmerkungen*

Sachlich-konstruktive Rückmeldungen enthalten korrigierende und zukunftsgerichtete Hilfestellungen. Daher werden indirekte Formen der Rückmeldung wie „Bist du sicher, dass das hier richtig ist?“ oder „Kann das sein?“ nicht als Positivindikatoren herangezogen. Sie stellen für den Schüler keine Rückmeldung mit weiterer Unterstützung dar, sondern geben dem Schüler lediglich eine Rückmeldung, dass er diese Aufgabe nochmals überdenken muss. Auch das Vorsagen von Antworten durch die Lehrperson wird nicht als Positivindikator herangezogen, führen aber auch nicht zu einer Abwertung.

*19.2.8 Positiver Umgang der Lehrperson mit Fehlern (Fehl)**Grundidee*

Fehler werden häufig als unerwünschte Ergebnisse angesehen, die es zu vermeiden gilt. Sie können jedoch auch positive Funktionen haben. Sie können von der Lehrperson genutzt werden, um mangelndes oder unzutreffendes Verständnis zu erkennen und zu revidieren. Das Merkmal erfasst, wie die Lehrperson mit fehlerhaften Äußerungen und auftretenden Schwierigkeiten umgeht. Außerdem soll erfasst werden, inwiefern auftretenden Fehlern im Unterricht eine positive Bedeutung beigemessen wird, indem sie Gelegenheiten bieten, Denkweisen aufzudecken und zu korrigieren. Dabei sollte die Lehrperson den jeweiligen Schüler nicht bloßstellen bzw. ihm das Gefühl vermitteln, dass Fehler zu machen nichts Schlimmes ist.

*Indikatoren*

- Schwache Leistungen der Schüler werden von der Lehrperson nicht zusätzlich betont.

- Die Lehrperson stellt die Schüler, wenn es zu Fehlern kommt, nicht bloß.
- Die Lehrperson geht auf die Fehler / Schwierigkeiten der Schüler ein, indem sie den Schülern gegenüber Verständnis zeigt und sie unterstützt. Die Lehrperson gibt auf Fehler ein ausführliches, motivierendes Feedback. Sie thematisiert Fehler in konstruktiver Weise.
- Die Lehrperson reagiert geduldig bei Fehlern.
- Die Lehrperson greift Fehler auf und nutzt sie, um das Verständnis der Schüler zu revidieren bzw. zu vertiefen. Sie nutzt Fehler, um Zusammenhänge zu verdeutlichen.
- Bei häufig vorkommenden Fehlern bezieht die Lehrperson bei der Korrektur die ganze Klasse mit ein, nicht einzelne Schüler. Sie erklärt auftretende Probleme noch einmal vor der ganzen Klasse.
- Bei der Lehrperson scheint Fehler zu machen nichts Schlimmes zu sein.
- Die Lehrperson achtet darauf, dass in der Klasse niemand ausgelacht wird, der einen Fehler macht.

#### *Negativindikatoren*

- Die Lehrperson beschämt die Schüler, z. B. lässt sie einzelne Schüler bei Übungen die Anzahl der gemachten Fehler öffentlich demonstrieren.
- Der Tonfall der Lehrperson ist herabsetzend.
- Die Lehrperson geht über Fehler hinweg und gibt keine Fehlerrückmeldung.
- Die Lehrperson nimmt andere Schüler dran, um die Fehler zu korrigieren.

#### *Antwort: Gesamteindruck*

- Eine „4“ wird vergeben, wenn die Lehrperson mit den Fehlern der Schüler konstruktiv (das heißt den Lernprozess anregend und unterstützend) und motivierend (d. h. die Lernbereitschaft anregend) umgeht.
- Eine „3“ wird vergeben, wenn die Lehrperson zwar das Vorkommen von Fehlern akzeptiert, mit den Fehlern jedoch nicht konstruktiv arbeitet.
- Eine „2“ wird vergeben, wenn im Unterricht gelegentlich Situationen erkennbar sind, in denen die Lehrperson eher negativ auf Fehler einzelner Schüler reagiert.
- Eine „1“ wird nur dann vergeben, wenn bei der Lehrperson eine klar abschätzigere Behandlung von Fehlern erkennbar ist. Für die Vergabe der „1“ muss die Lehrperson auf die Fehler der Schüler deutlich negativ und abfällig reagieren.

*Anmerkungen*

Es geht nicht um Fehler, die als Folge mangelnder und ungenauer Instruktion oder schlechter Vorbereitung durch die Lehrperson auftreten. Diese haben nichts mit einer Fehlerkultur zu tun. Es geht ausschließlich um die Fehler, aus denen Schüler lernen können und mit denen konstruktiv umgegangen wird.

Für eine positive Beurteilung dieses Merkmals reicht es nicht aus, dass kein negativer Umgang zu sehen ist, sondern es muss ein positiver Umgang im Sinne der Grundidee und Indikatoren sichtbar sein.

*19.2.9 Positives Fehlerklima zwischen den Schülern (Fekl)**Grundidee*

Nicht nur die Lehrperson, sondern auch die Mitschüler sollten bei auftretenden Fehlern nicht negativ reagieren. Das Merkmal erfasst, wie die Schüler einer Klasse mit den Fehlern ihrer Mitschüler umgehen.

*Indikatoren*

- Die Klasse hat eine positive Einstellung zu Fehlern.
- Die Klasse reagiert bei Fehlern einzelner Mitschüler nicht negativ.
- Die Klasse zeigt sich motivierend und unterstützend bei auftretenden Fehlern von Mitschülern. Die Schüler bieten einander Hilfestellungen an.

*Antwort: Gesamteindruck*

- Eine „4“ wird vergeben, wenn klar erkennbar ist, dass die Schüler Fehlern gegenüber positiv eingestellt sind und ein positiver Umgang mit Fehlern herrscht. Mitschüler werden bei Fehlern deutlich unterstützt (Hilfestellungen).
- Eine „3“ wird vergeben, wenn Schüler eher neutral auf die Fehler ihrer Mitschüler reagieren und ihnen keine Hilfestellungen bieten.
- Eine „2“ wird vergeben, wenn gelegentlich negative Reaktionen aufseiten der Schüler der Klasse erkennbar sind.
- Eine „1“ wird vergeben, wenn in der Klasse eine negative Einstellung zu Fehlern herrscht, das heißt Mitschüler häufig auf Fehler negativ (bis verächtlich) reagieren. Die Schüler lachen einander aus.

*Anmerkungen*

Hilfestellung bedeutet nicht, dass Schüler versuchen, dem Mitschüler das Ergebnis vorzusagen.

### 19.2.10 Wettbewerb und Leistungsdruck (Wett)

#### Grundidee

Das Merkmal soll den im Unterricht erlebten Wettbewerbs- bzw. Leistungsdruck erfassen. Bei diesem Merkmal soll eingeschätzt werden, inwiefern Noten oder andere Formen der expliziten Beurteilung oder Bewertung von Leistungen (z. B. Vergabe von Punkten) bereits in der Grundschule existieren und somit bereits im frühen Grundschulalter die Schüler unter Druck setzen. Leistungsdruck zeigt sich jedoch nicht nur in der Transparenz von Noten, sondern äußert sich beispielsweise darin, dass wenig Zeit für selbständiges Arbeiten zur Verfügung steht, eine geringe Anzahl von Wiederholungen erfolgt oder die Lerninhalte zu umfangreich sind und dadurch in rascher Folge abgearbeitet werden (v. Saldern & Littig, 1987).

#### Negativindikatoren

- Die Lehrperson schafft Transparenz in der Leistungsbewertung, indem sie Noten bzw. Leistungen öffentlich ansagt.
- Die Lehrperson fördert das Konkurrenz- / Wettbewerbsdenken der Schüler durch Worte wie beispielsweise „*Wer als erstes fertig ist...*“
- Die Lehrperson macht den Leistungsstand der Schüler öffentlich, z. B. lässt sie bei den täglichen Übungen die Kinder ihre Anzahl von Fehlern öffentlich ansagen bzw. mit Melden anzeigen.
- Der Unterricht geht schnell voran, sodass einzelne Schüler oft nicht mitkommen. Der Unterricht scheint in einem Tempo abzulaufen, bei dem nicht alle Schüler folgen können.

#### Antwort: Gesamteindruck

- Eine „4“ wird vergeben, wenn ein hoher Wettbewerbs- und Leistungsdruck in der Klasse vorherrscht. Die Schüler müssen mehrmals in der Stunde ihre Leistungen öffentlich machen. Die Lehrperson fördert mit ihrem Verhalten sehr häufig die Wettbewerbsgedanken der Schüler. Die Schüler machen teilweise den Eindruck, dass sie unter Druck stehen.
- Eine „3“ wird vergeben, wenn die Lehrperson zwar relativ häufig die Leistungen der Schüler öffentlich macht bzw. den Wettbewerbsgedanken fördert, die Schüler jedoch nicht den Eindruck machen, dass sie dies unter Druck setzt.
- Eine „2“ wird vergeben, wenn die Lehrperson selten die Leistungen der Schüler öffentlich macht bzw. nur gelegentlich mit ihrem Verhalten den Wettbewerbsgedanken fördert.
- Eine „1“ wird vergeben, wenn kaum bis gar kein Wettbewerbs- bzw. Leistungsdruck für den Beobachter ersichtlich wird.

*Anmerkungen*

Das Merkmal „Wettbewerb und Leistungsdruck“ muss für weiterführende Analysen rekodiert werden, da eine „4“ auf diesem Merkmal hohen „Wettbewerb und Leistungsdruck“ bedeutet, die „1“ hingegen vergeben wurde, wenn kaum bis gar kein „Wettbewerb und Leistungsdruck“ in der Klasse zu beobachten war.

## 20 Auswertungsbogen für die Rater

IDLehrer: \_\_\_\_\_

VSM

VSS

Merkmal	Code SPSS	1. Kodierung (Lek-Beginn bis ca. 45. Minute oder PAUSE)				2. Kodierung (45. Minute bzw. nach der PAUSE bis Lek-Ende)				Mittelwert der beiden Kodierungen
		1	2	3	4	1	2	3	4	
<b>Klassenführung</b>										
1	Allgegenwärtigkeit der Lehrperson	ALLG								
2	Gruppenfokus	GRUP								
3	Zeitmanagement / effektive Zeitznutzung	ZEIT								
4	Übergangsmangement (Managing Transition)	UEB								
5	Regelklarheit bzw. -verwendung	REGL								
6	Störungsfreiheit	DISZ								
7	Effektiver Umgang mit auftretenden Unterrichtsstörungen	STÖR								
8	Vorbereitung des Klassenraums	VORB								
<b>Unterrichtsklima</b>										
1	humorvolle Lernatmosphäre	HU-MO								
2	Anerkennung der Schüler durch die Lehrperson	ANES								
3	Gegenseitige Anerkennung der Schüler	ANEG								
4	Fürsorglichkeit, Herzlichkeit und Wärme	FUER								
5	Der Einsatz von Lob	LOB								
6	Ermütigung durch die Lehrperson	EMUT								
7	sachlich-konstruktive Rückmeldung	RÜCK								
8	positiver Umgang der Lehrperson mit Fehlern	FEHL								
9	positives Fehlerklima zwischen den Schülern	FEKL								
10	Wettbewerb und Leistungsdruck	WETT								

Anmerkungen: \_\_\_\_\_

## 21 Deskriptive Statistiken

Für die Berechnung der Mittelwerte wurde jeder Lehrperson für jede Videostudie ein Wert zugeordnet, indem der Mittelwert der beiden Ratings pro Dimension (2 x 45 Minuten) und pro Rater gebildet wurde (vgl. Kapitel 12.1.3). Anschließend sind die deskriptiven Statistiken und die Häufigkeitsverteilungen der Mittelwerte der hoch inferenten Ratings getrennt für beide Videostudien und Unterrichtsqualitätsmerkmale dargestellt. Dabei ist zu bedenken, dass für das Fach Deutsch (1. Schuljahr) die Dimension „positiver Umgang der Lehrperson mit Fehlern (Fehl)“ nicht reliabel von den zwei Ratern eingeschätzt werden konnte. Für das Fach Mathematik (2. Schuljahr) kamen die Rater für die folgenden vier Merkmale zu keinem übereinstimmenden Urteil: „Vorbereitung des Klassenraums (Vorb)“, „positiver Umgang der Lehrperson mit Fehlern (Fehl)“, „positives Fehlerklima zwischen den Schülern (Fekl)“ und „Wettbewerb und Leistungsdruck (Wett)“ (vgl. Kapitel 13.2). Die Variable „Wettbewerb und Leistungsdruck (Wett)“ wurde in beiden Videostudien im Vorfeld der Analysen recodiert (Wett\_rec). Eine „1“ bedeutet demnach, dass kaum bis gar kein Wettbewerbs- bzw. Leistungsdruck für den Beobachter ersichtlich wird. Eine „4“ bedeutet, dass ein hoher Wettbewerbs- und Leistungsdruck in der Klasse vorherrscht. Die Schüler müssen mehrmals in der Stunde ihre Leistungen öffentlich machen. Die Lehrperson fördert mit ihrem Verhalten sehr häufig die Wettbewerbsgedanken der Schüler. Die Schüler machen teilweise den Eindruck, dass sie unter Druck stehen.

### *Test auf Normalverteilung (fachspezifisch)*

Zusätzlich zu der deskriptiven Analyse erfolgt für jedes der hoch inferenten Merkmale die Prüfung auf Vorliegen einer Normalverteilung mithilfe des Kolmogorov-Smirnov-Tests. Die Ergebnisse aus der Videostudie Deutsch (1. Schuljahr, VSS) sind in Tabelle 43 und die der Videostudie Mathematik (2. Schuljahr, VSM) in Tabelle 44 dargestellt.

Welche der Merkmale normalverteilt sind, geht aus Irrtumswahrscheinlichkeit  $p$  hervor, die im vorliegenden Fall als „asymptotische Signifikanz“ bezeichnet wird. Ist dieser Wert signifikant ( $p \leq .05$ ), bedeutet dies eine signifikante Abweichung der Verteilung einer Dimension von der Normalverteilung. Die Annahme der Normalverteilung muss in diesem Fall zurück gewiesen werden. Sobald der Wert  $p \geq .05$  ist, kann angenommen werden, dass die Werte der getesteten Dimension hinreichend normalverteilt sind.

Tabelle 39: Häufigkeitsverteilung der Klassenführungsmerkmale in der Videostudie Deutsch (VSS, 1. Schuljahr,  $N = 50$ )

MW	Allg	Grup	Zeit	Ueb	Regl	Disz	Stör	Vorb
1.00	0	1	0	0	0	0	0	0
1.25	0	0	0	1	0	0	0	0
1.50	1	0	0	0	0	1	1	0
1.75	0	0	1	0	1	1	1	0
2.00	4	0	0	2	4	3	3	1
2.25	5	1	0	3	1	0	1	0
2.50	5	1	4	2	4	9	8	3
2.75	7	0	2	4	4	7	6	0
3.00	8	10	11	16	8	8	10	2
3.25	3	3	4	8	2	6	6	0
3.50	9	11	12	5	10	5	4	3
3.75	2	12	8	4	5	4	3	0
4.00	6	9	8	5	10	6	7	41
fehrend	0	2	0	0	1	0	0	0
<i>MW</i>	2.99	3.44	3.35	3.09	3.21	3.02	3.02	3.80
<i>SD</i>	.64	.55	.50	.57	.66	.63	.64	.48
<i>Min</i>	1.50	1.00	1.75	1.25	1.75	1.50	1.50	2.00
<i>Max</i>	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00

Tabelle 40: Häufigkeitsverteilung der Unterrichtsklimamerkmale in der Videostudie Deutsch (VSS, 1. Schuljahr,  $N = 50$ )

MW	Humo	Anes	Aneg	Fuer	Lob	Emut	Rück	Fehl	Fekl	Wett_rec
1.00	11	0	0	1	5	0	0	0	0	45
1.25	8	0	0	0	3	0	1	0	0	1
1.50	8	1	0	0	5	3	0	0	0	1
1.75	4	0	0	2	5	3	2	0	0	1
2.00	13	1	1	6	6	4	7	1	4	1
2.25	2	0	0	2	3	0	3	0	0	0
2.50	2	2	3	5	7	4	10	2	6	0
2.75	1	0	5	3	3	3	10	2	3	0
3.00	0	9	12	27	9	33	17	28	37	0
3.25	1	2	9	2	1	0	0	8	0	0
3.50	0	6	11	1	1	0	0	5	0	0
3.75	0	6	2	1	0	0	0	2	0	0
4.00	0	23	7	0	2	0	0	1	0	0
fehrend	0	0	0	0	0	0	0	0	0	1
<i>MW</i>	1.63	3.55	3.25	2.73	2.23	2.70	2.58	3.09	2.85	1.05
<i>SD</i>	.53	.58	.46	.51	.78	.50	.43	.33	.30	.19
<i>Min</i>	1.00	1.50	2.00	1.00	1.00	1.50	1.25	2.00	2.00	1.00
<i>Max</i>	3.25	4.00	4.00	3.75	4.00	3.00	3.00	4.00	3.00	2.00

Tabelle 41: Häufigkeitsverteilung der Klassenführungsmerkmale in der Videostudie Mathematik (VSM, 2. Schuljahr,  $N = 51$ )

MW	Allg	Grup	Zeit	Ueb	Regl	Disz	Stör	Vorb
1.00	0	0	0	0	0	0	0	0
1.25	0	0	0	0	0	0	0	0
1.50	0	0	0	0	0	1	1	0
1.75	0	0	1	1	1	1	0	0
2.00	3	1	0	2	0	0	1	0
2.25	1	1	1	3	0	1	1	0
2.50	5	1	2	2	2	6	4	0
2.75	4	0	1	5	6	5	5	0
3.00	9	14	5	10	11	6	6	0
3.25	7	2	2	2	4	9	10	0
3.50	11	13	7	10	4	7	7	4
3.75	4	4	9	8	6	2	2	2
4.00	7	12	19	5	13	11	10	45
fehrend	0	3	4	3	4	2	4	0
<i>MW</i>	3.20	3.42	3.56	3.18	3.36	3.21	3.23	3.95
<i>SD</i>	.56	.48	.55	.59	.55	.62	.59	.14
<i>Min</i>	2.00	2.00	1.75	1.75	1.75	1.50	1.50	3.50
<i>Max</i>	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00

Tabelle 42: Häufigkeitsverteilung der Unterrichtsklimamerkmale in der Videostudie Mathematik (VSM, 2. Schuljahr,  $N = 51$ )

MW	Humo	Anes	Aneg	Fuer	Lob	Emut	Rück	Fehl	Fekl	Wett_rec
1.00	14	1	0	1	4	0	0	0	0	38
1.25	3	0	0	0	0	0	0	0	0	7
1.50	12	0	0	0	3	0	1	0	0	5
1.75	5	0	0	0	4	1	1	0	0	0
2.00	9	0	1	14	4	4	3	1	2	0
2.25	1	1	0	3	1	1	2	1	0	1
2.50	4	0	5	8	7	4	4	4	3	0
2.75	1	0	3	4	4	3	6	4	11	0
3.00	1	4	8	19	14	38	32	19	30	0
3.25	0	5	5	0	5	0	2	8	3	0
3.50	0	12	9	1	3	0	0	13	1	0
3.75	0	5	5	0	0	0	0	0	0	0
4.00	1	23	15	1	2	0	0	1	0	0
fehrend	0	0	0	0	0	0	0	0	1	0
<i>MW</i>	1.66	3.61	3.40	2.57	2.57	2.83	2.80	3.09	2.90	1.12
<i>SD</i>	.63	.54	.54	.53	.76	.34	.38	.38	.26	.23
<i>Min</i>	1.00	1.00	2.00	1.00	1.00	1.75	1.50	2.00	2.00	1.00
<i>Max</i>	4.00	4.00	4.00	4.00	4.00	3.00	3.25	4.00	3.50	2.25

Tabelle 43: Prüfung auf Normalverteilung der Merkmale in VSS (Kolmogorov-Smirnov-Test)

	<b>Merkmale</b>	<b>Schiefe</b>	<b>Exzess (Kurtosis)</b>	<b>K-S-Test asymptotische Signifikanz</b>
<b>Klassenführung</b>	Allgegenwärtigkeit der Lehrperson (Allg)	-1.03	- .762	.384
	Gruppenfokus (Grup)	-1.988	5.215	.018
	Zeitmanagement / effektive Zeitznutzung (Zeit)	-.748	.577	.077
	Übergangsmanagement (Managing Transition) (Ueb)	-.664	1.171	.040
	Regelklarheit bzw. -verwendung (Regl)	-.500	-.821	.077
	Störungsfreiheit (Disz)	-.198	-.408	.562
<b>Unterrichtsklima</b>	effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)	-.155	-.417	.588
	humorvolle Lernatmosphäre (Humo)	.709	.386	.250
	Anerkennung der Schüler durch die Lehrperson (Anes)	-1.513	2.349	.006
	gegenseitige Anerkennung der Schüler (Aneg)	-.143	-.037	.400
	Fürsorglichkeit, Herzlichkeit und Wärme (Fuer)	-1.145	1.525	.000
	Einsatz von Lob (Lob)	.205	-.587	.718
	Ermutigung durch die Lehrperson (Emut)	-1.428	.536	.000
	sachlich-konstruktive Rückmeldung (Rück)	-.961	.385	.049

Anmerkung: Datengrundlage  $N_{VSS} = 50$ . Fett gedruckt: Die Merkmale sind hinreichend normalverteilt ( $p \geq .05$ ).

Tabelle 44: Prüfung auf Normalverteilung der Merkmale in VSM (Kolmogorov-Smirnov-Test)

	<b>Merkmale</b>	<b>Schiefe</b>	<b>Exzess (Kurtosis)</b>	<b>K-S-Test asymptotische Signifikanz</b>
<b>Klassenführung</b>	Allgegenwärtigkeit der Lehrperson (Allg)	- .416	- .462	.281
	Gruppenfokus (Grup)	- .637	.357	.114
	Zeitmanagement / effektive Zeitznutzung (Zeit)	-1.495	1.882	.014
	Übergangsmanagement (Managing Transition) (Ueb)	- .529	-.480	.081
	Regelklarheit bzw. -verwendung (Regl)	- .471	-.306	.151
	Störungsfreiheit (Disz)	- .525	-.120	.455
<b>Unterrichtsklima</b>	effektiver Umgang mit auftretenden Unterrichtsstörungen (Stör)	- .574	.098	.427
	humorvolle Lernatmosphäre (Humo)	1.235	1.998	.114
	Anerkennung der Schüler durch die Lehrperson (Anes)	-2.681	10.472	.007
	gegenseitige Anerkennung der Schüler (Aneg)	- .527	-.617	.145
	Fürsorglichkeit, Herzlichkeit und Wärme (Fuer)	-.205	.653	.031
	Einsatz von Lob (Lob)	- .516	-.333	.062
	Ermutigung durch die Lehrperson (Emut)	-1.930	2.534	.000
	sachlich-konstruktive Rückmeldung (Rück)	-1.789	2.643	.000

Anmerkung: Datengrundlage  $N_{VSM} = 51$ . Fett gedruckt: Die Merkmale sind hinreichend normalverteilt ( $p \geq .05$ ).

## 22 Ergebnis des Modellvergleichs – Darstellung der Modelle

Die folgenden Abbildungen stellen die Ergebnisse der konfirmatorischen Faktorenanalysen (vgl. Kapitel 15.1.3) graphisch dar.

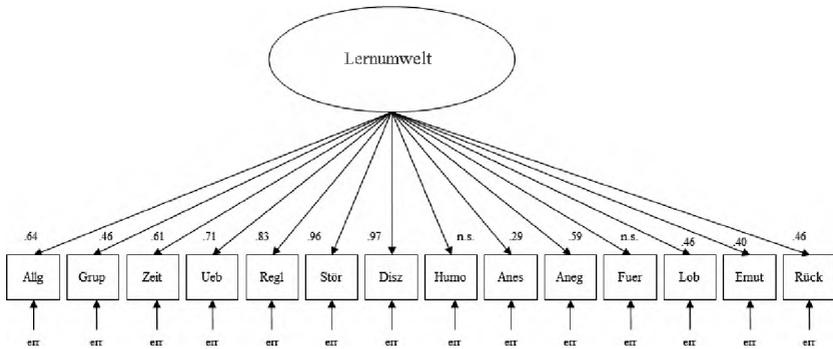


Abbildung 34: Ergebnis der konfirmatorischen Faktorenanalyse – Das Generalfaktor-Modell

Anmerkung:  $N = 101$ ,  $\chi^2 = 268.862$ ,  $df = 77$ ,  $\chi^2/df = 3.492$ ,  $CFI = .751$ ,  $TLI = .706$ ,  $AIC = 1799.159$ ,  $BIC = 1908.994$ ,  $RMSEA = .157$

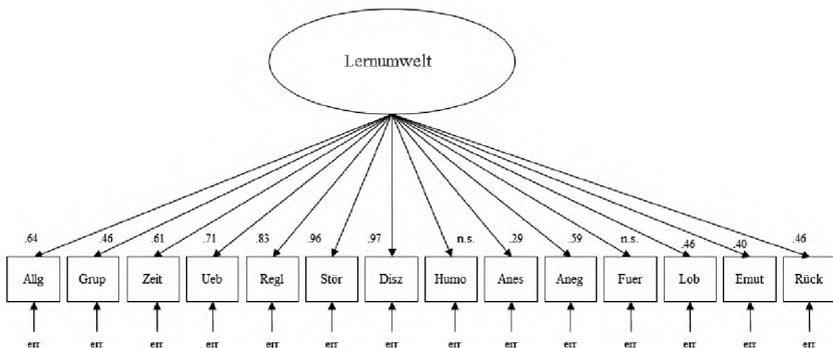


Abbildung 35: Ergebnis der konfirmatorischen Faktorenanalyse – Das Modell 2

Anmerkung:  $N = 101$ ,  $\chi^2 = 255.903$ ,  $df = 76$ ,  $\chi^2/df = 3.367$ ,  $CFI = .767$ ,  $TLI = .721$ ,  $AIC = 1788.200$ ,  $BIC = 1900.650$ ,  $RMSEA = .153$

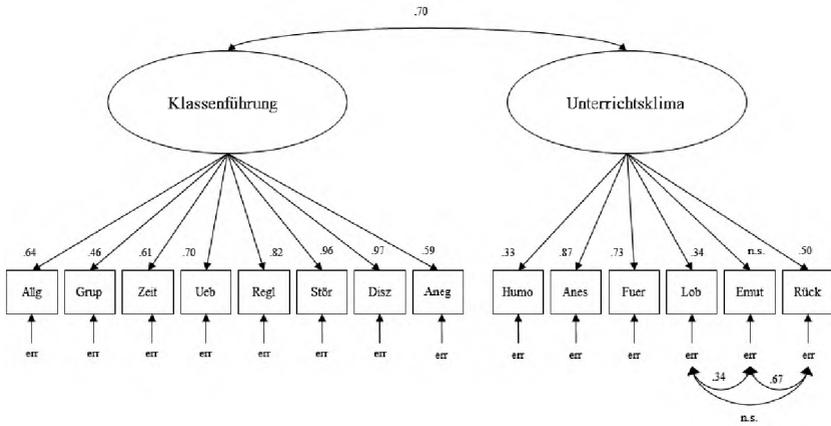


Abbildung 36: Ergebnis der konfirmatorischen Faktorenanalyse – Das Modell 2b

Anmerkung:  $N = 101$ ,  $\chi^2 = 169.895$ ,  $df = 73$ ,  $\chi^2/df = 2.327$ , CFI = .874, TLI = .843, AIC = 1708.191, BIC = 1828.487, RMSEA = .115

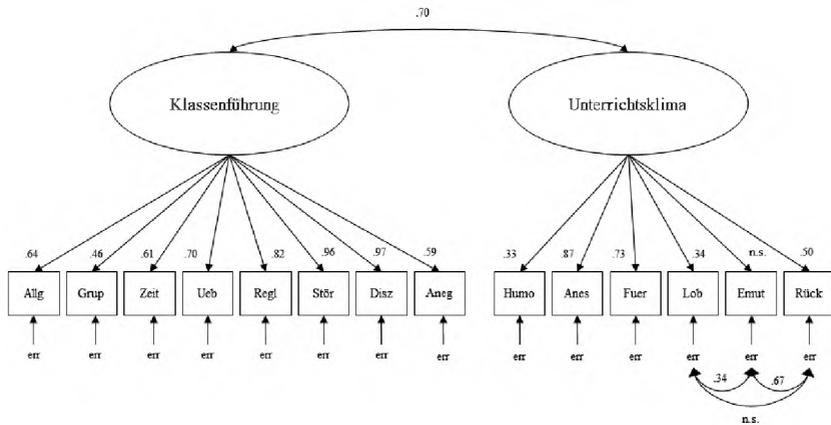


Abbildung 37: Ergebnis der konfirmatorischen Faktorenanalyse – Das Modell 3

Anmerkung:  $N = 101$ ,  $\chi^2 = 176.424$ ,  $df = 74$ ,  $\chi^2/df = 2.384$ , CFI = .867, TLI = .837, AIC = 1712.721, BIC = 1830.401, RMSEA = .117

## 23 Dokumentation der hoch inferenten Skalen

Die folgende Skalendokumentation umfasst die verwendeten Skalen der Videostudie Deutsch (VSS, 1. Schuljahr, vgl. Kapitel 23.1) und der Videostudie Mathematik (VSM, 2. Schuljahr, vgl. Kapitel 23.2). Für jede sich ergebene Skala sind die Skalenkennwerte und Reliabilitäten (Cronbachs Alpha) angegeben. Resultierend aus der Reliabilitätsanalyse werden neben dem Cronbachs Alpha für jedes hoch inferente Merkmal die Trennschärfen ( $r_{ij}$ ), Mittelwerte ( $MW$ ), Standardabweichungen ( $SD$ ) und Höhe der Ladung auf der entsprechenden Skala ( $a_j$ ) angegeben. Die Itemtrennschärfe gibt die Korrelation der Dimension mit dem Summenwert aus allen anderen in die Analyse einbezogenen Merkmalen an (vgl. Bortz & Döring, 2005).

Grundlage für die mehrfaktoriellen Lösungen bildet die Mustermatrix der rotierten Faktorlösung. Es kam teilweise vor, dass die Ladungen aller Merkmale eines Faktors negativ waren. Aus Gründen der besseren Lesbarkeit werden diese invertiert dargestellt. Dies wird an entsprechender Stelle durch eine Fußnote ersichtlich.

### 23.1 Videostudie Deutsch (VSS, 1. Schuljahr)

Für die Faktorenanalyse im Fach Deutsch (VSS, 1. Schuljahr) wurden nur diejenigen hoch inferenten Merkmale berücksichtigt, die dem Kriterium  $g \geq .65$  bei der Berechnung der fachspezifischen Beobachterübereinstimmung standhielten. Folgende Merkmale erfüllten im Fach Deutsch (VSS, 1. Schuljahr) das gesetzte Kriterium nicht: „positiver Umgang der Lehrperson mit Fehlern (Fehl)“ (vgl. Kapitel 13.2).

Um eine Vergleichbarkeit mit der Videostudie Mathematik zu gewährleisten, wurde zusätzlich auf die folgenden Merkmale verzichtet: „Vorbereitung des Klassenraums (Vorb)“, „positives Fehlerklima zwischen den Schülern (Fekl)“ und „Wettbewerb und Leistungsdruck (Wett)“. Eine erste explorative Faktorenanalyse mit allen übrigen 14 Merkmalen (Hauptkomponentenanalyse, Faktorenextraktion nach Screeplot und Kaiserkriterium, Rotation: Oblimin) ergab drei differenzierbare und reliable Faktoren: Klassenführung, Unterrichtsklima und Feedback. Für die vorliegenden Analysen wurde auf die Skala Feedback verzichtet. Aus diesem Grund werden anschließend nur die übrigen zwei Skalen dokumentiert.

*I. Skala VSS: Klassenführung (Gesamtskala)*

Anzahl der Items: 8  
 Antwortformat / Kategorien: 4-stufig  
 VSS\_Allg, VSS\_Grup, VSS\_Zeit, VSS\_Ueb, VSS\_Regl, VSS\_Disz, VSS\_Stör, VSS\_Aneg  
 Skalenbildung: Mittelwert  
 Konstrukte in der Faktorenlösung: Unterrichtsklima VSS u. Feedback VSS  
 Angewandte Verfahren / Vorgehen: Hauptkomponentenanalyse, Faktorextraktion nach Screplot und Kaiserkriterium, Rotation: Oblimin, Reliabilitätsanalyse

Variable		Skalenkennwerte			
	G-Koeffizient (relativ)	MW	SD	$r_{tt}$	$a_r$
VSS_Allg	.90	2.98	.65	.62	.63
VSS_Grup	.82	3.40	.61	.42	.52
VSS_Zeit	.66	3.34	.51	.70	.59*
VSS_Ueb	.79	3.09	.58	.81	.74
VSS_Regl	.68	3.18	.68	.84	.85
VSS_Disz	.86	3.03	.63	.91	.88
VSS_Stör	.85	3.02	.64	.88	.93
VSS_Aneg	.86	3.24	.46	.53	.84**
Skala: VSS_CM		Cronbachs $\alpha$ = .91 MW = 3.16 SD = .47 N = 49			

\* Nebenladung auf dem Faktor Feedback (.35).

\*\* Nebenladung auf dem vierten Faktor zusammen mit VSS\_Fekl (.48).

Variable	
VSS_Allg	VSS Allgegenwärtigkeit (Mittelwert aus ALLG_MM und ALLG_AD, gesamte Lektion); G-Koeff fachspezifisch = .90 Datengrundlage N = 48 (ohne Schulungsvideos)
VSS_Grup	VSS Gruppenfokus (Mittelwert aus GRUP_MM und GRUP_AD, gesamte Lektion); G-Koeff fachspezifisch = .82 Datengrundlage N = 48 (ohne Schulungsvideos)
VSS_Zeit	VSS Zeitmanagement / effiziente Zeitznutzung (Mittelwert aus ZEIT_MM und ZEIT_AD, gesamte Lektion); G-Koeff fachspezifisch = .66 Datengrundlage N = 48 (ohne Schulungsvideos)
VSS_Ueb	VSS Übergangsmanagement (Mittelwert aus UEB_MM und UEB_AD, gesamte Lektion); G-Koeff fachspezifisch = .79 Datengrundlage N = 48 (ohne Schulungsvideos)
VSS_Regl	VSS Regelklarheit bzw. -verwendung (Mittelwert aus REGL_MM und REGL_AD, gesamte Lektion); G-Koeff fachspezifisch = .68 Datengrundlage N = 48 (ohne Schulungsvideos)
VSS_Disz	VSS Störungsfreiheit (Mittelwert aus DISZ_MM und DISZ_AD, gesamte Lektion); G-Koeff fachspezifisch = .86 Datengrundlage N = 48 (ohne Schulungsvideos)

Variable	
VSS_Stör	VSS Effektiver Umgang mit auftretenden Unterrichtsstörungen (Mittelwert aus STÖR_MM und STÖR_AD, gesamte Lektion); G-Koeff fachspezifisch = .85 Datengrundlage N = 48 (ohne Schulungsvideos)
VSS_Aneg	VSS Gegenseitige Anerkennung der Schüler (Mittelwert aus ANEG_MM und ANEG_AD, gesamte Lektion); G-Koeff fachspezifisch = .86 Datengrundlage N = 48 (ohne Schulungsvideos)

2. Skala VSS: Unterrichtsklima

Anzahl der Items: 3  
 Antwortformat / Kategorien: 4-stufig  
 VSS\_Humo, VSS\_Anes, VSS\_Fuer  
 Skalenbildung: Mittelwert  
 Konstrukte in der Faktorenlösung: Klassenführung (Gesamtskala) VSS und Feedback VSS  
 Angewandte Verfahren / Vorgehen: Hauptkomponentenanalyse, Faktorextraktion nach Screeplot und Kaiserkriterium, Rotation: Oblimin, Reliabilitätsanalyse

Variable	Skalenkennwerte				
	G-Koeffizient (relativ)	MW	SD	$r_{tt}$	$a_2$
VSS_Humo	.80	1.63	.53	.39	.66
VSS_Anes	.90	3.55	.58	.65	.84
VSS_Fuer	.84	2.73	.51	.70	.84
Skala: VSS_UKlima		Cronbachs $\alpha = .76$ MW = 2.64 SD = .44 N = 50			

Variable	
VSS_Humo	VSS humorvolle Lernatmosphäre (Mittelwert aus HUMOR_MM und HUMOR_AD, gesamte Lektion); G-Koeff fachspezifisch = .80 Datengrundlage N = 48 (ohne Schulungsvideos)
VSS_Anes	VSS Anerkennung der Schüler durch die Lehrperson (Mittelwert aus ANES_MM und ANES_AD, gesamte Lektion); G-Koeff fachspezifisch = .90 Datengrundlage N = 48 (ohne Schulungsvideos)
VSS_Fuer	VSS Fürsorglichkeit, Herzlichkeit und Wärme (Mittelwert aus FUER_MM und FUER_AD, gesamte Lektion); G-Koeff fachspezifisch = .84 Datengrundlage N = 48 (ohne Schulungsvideos)
<b>Ausgeschlossene Variable</b>	
VSS_Lob*	VSS Einsatz von Lob (Mittelwert aus LOB_MM und LOB_AD, gesamte Lektion); G-Koeff fachspezifisch = .90 Datengrundlage N = 48 (ohne Schulungsvideos)

\* Bei der Berechnung der Reliabilität der Skala „Unterrichtsklima“ zeigte sich, dass das Cronbachs Alpha ohne das Merkmal „VSS\_Lob“ von  $\alpha = .73$  auf  $\alpha = .76$  steigt. Außerdem

lud „VSS\_Lob“ annähernd gleich hoch auf dem Faktor „Klassenführung“. Im Sinne einer Vereinheitlichung mit der Videostudie Mathematik (VSS) wurde das Merkmal „VSS\_Lob“ bei der Skalenbildung nicht berücksichtigt. „VSS\_Lob“ kann jedoch weiterhin als Einzelitem verwendet werden.

## 23.2 Videostudie Mathematik (VSM, 2. Schuljahr)

Für die Faktorenanalysen im Rahmen der Videostudie Mathematik (2. Schuljahr) wurden nur diejenigen hoch inferenten Merkmale berücksichtigt, die dem Kriterium  $g \geq .65$  standhielten. Folgende Merkmale wurden nicht berücksichtigt: „Vorbereitung des Klassenraums (Vorb)“, „positiver Umgang der Lehrperson mit Fehlern (Fehl)“, „positives Fehlerklima zwischen den Schülern (Fekl)“ und „Wettbewerb und Leistungsdruck (Wett)“ (vgl. Kapitel 13.2).

Eine erste explorative Faktorenanalyse mit allen übrigen 14 Merkmalen (Hauptkomponentenanalyse, Faktorenextraktion nach Screeplot und Kaiserkriterium, Rotation: Oblimin) ergab vier differenzierbare und reliable Faktoren: effizientes Zeitmanagement, effektiver Umgang mit Störungen, Unterrichtsklima und Feedback. Für die vorliegenden Analysen wurde auf die Skala Feedback verzichtet. Aus diesem Grund werden anschließend nur die übrigen drei Skalen dokumentiert.

### *1. Skala VSM: Effizientes Zeitmanagement (Teilskala Klassenführung)*

Anzahl der Items:	4
Antwortformat / Kategorien:	4-stufig VSM_Allg, VSM_Grup, VSM_Zeit, VSM_Ueb
Skalenbildung:	Mittelwert
Konstrukte in der Faktorenlösung:	effektiver Umgang mit Unterrichtsstörungen VSM, Unterrichtsklima VSM und Feedback VSM
Angewandte Verfahren / Vorgehen:	Hauptkomponentenanalyse, Faktorenextraktion nach Screeplot und Kaiserkriterium, Rotation: Oblimin, Reliabilitätsanalyse

Variable	Skalenkennwerte				
	G-Koeffizient (relativ)	MW	SD	$r_{it}$	$a_i^*$
VSM_Allg	.88	3.20	.56	.62	.70
VSM_Grup	.92	3.42	.48	.65	.74
VSM_Zeit	.98	3.56	.55	.55	.80
VSM_Ueb	.94	3.18	.59	.73	.73
Skala: VSM_CM_effZeit		Cronbachs $\alpha$ = .82 MW = 3.33 SD = .44 N = 48			

\* Aus Gründen der besseren Lesbarkeit wurden die Faktorladungen invertiert.

Variable	
VSM_Allg	VSM Allgegenwärtigkeit (Mittelwert aus ALLG_MM und ALLG_AD, gesamte Lektion); G-Koeff fachspezifisch = .88 Datengrundlage N = 49 (ohne Schulungsvideos)
VSM_Grup	VSM Gruppenfokus (Mittelwert aus GRUP_MM und GRUP_AD, gesamte Lektion); G-Koeff fachspezifisch = .92 Datengrundlage N = 49 (ohne Schulungsvideos)
VSM_Zeit	VSM Zeitmanagement / effiziente Zeitznutzung (Mittelwert aus ZEIT_MM und ZEIT_AD, gesamte Lektion); G-Koeff fachspezifisch = .98 Datengrundlage N = 49 (ohne Schulungsvideos)
VSM_Ueb	VSM Übergangsmanagement (Mittelwert aus UEB_MM und UEB_AD, gesamte Lektion); G-Koeff fachspezifisch = .94 Datengrundlage N = 49 (ohne Schulungsvideos)

2. Skala VSM: *Effektiver Umgang mit Störungen (Teilskala Klassenführung)*

- Anzahl der Items: 4
- Antwortformat / Kategorien: 4-stufig  
VSM\_Regl, VSM\_Disz, VSM\_Stör, VSM\_Aneg
- Skalenbildung: Mittelwert
- Konstrukte in der Faktorenlösung: effizientes Zeitmanagement VSM, Unterrichtsklima VSM und Feedback VSM
- Angewandte Verfahren / Vorgehen: Hauptkomponentenanalyse, Faktorextraktion nach Screeplot und Kaiserkriterium, Rotation: Oblimin, Reliabilitätsanalyse

Variable	Skalenkennwerte				
	G-Koeffizient (relativ)	MW	SD	$r_{it}$	$a_i$
VSM_Regl	.95	3.36	.55	.78	.55
VSM_Disz	.77	3.21	.62	.84	.80
VSM_Stör	.86	3.23	.59	.88	.84
VSM_Aneg	.81	3.40	.54	.57	.74
Skala: VSM_CM_effStör		Cronbachs $\alpha$ = .89 MW = 3.29 SD = .50 N = 47			

Variable	
VSM_Regl	VSM Regelklarheit bzw. -verwendung (Mittelwert aus REGL_MM und REGL_AD, gesamte Lektion); G-Koeff fachspezifisch = .95 Datengrundlage N = 49 (ohne Schulungsvideos)
VSM_Disz	VSM Störungsfreiheit (Mittelwert aus DISZ_MM und DISZ_AD, gesamte Lektion); G-Koeff fachspezifisch = .77 Datengrundlage N = 49 (ohne Schulungsvideos)
VSM_Stör	VSM Effektiver Umgang mit auftretenden Unterrichtsstörungen (Mittelwert aus STÖR_MM und STÖR_AD, gesamte Lektion); G-Koeff fachspezifisch = .86 Datengrundlage N = 49 (ohne Schulungsvideos)
VSM_Aneg	VSM Gegenseitige Anerkennung der Schüler (Mittelwert aus ANEG_MM und ANEG_AD, gesamte Lektion); G-Koeff fachspezifisch = .81 Datengrundlage N = 49 (ohne Schulungsvideos)

3. Skala VSM: Unterrichtsklima

- Anzahl der Items: 3
- Antwortformat / Kategorien: 4-stufig  
VSM\_Humo, VSM\_Anes, VSM\_Fuer
- Skalenbildung: Mittelwert
- Konstrukte in der Faktorenlösung: effizientes Zeitmanagement, effektiver Umgang mit Unterrichtsstörungen und Feedback
- Angewandte Verfahren / Vorgehen: Hauptkomponentenanalyse, Faktorextraktion nach Screeplot und Kaiserkriterium, Rotation: Oblimin, Reliabilitätsanalyse

Variable		Skalenkennwerte			
	G-Koeffizient (relativ)	MW	SD	$r_{it}$	$a_2$
VSM_Humo*	.81	1.66	.63	.32	.50**
VSM_Anes	.81	3.61	.54	.38	.81
VSM_Fuer	.91	2.57	.53	.61	.89
Skala: VSM_UKlima		Cronbachs $\alpha$ = .62 MW = 2.61 SD = .43 N = 51			

\* Würde man dieses Merkmal aus der Skala herausnehmen, stiege das Cronbachs  $\alpha$  auf .69.

\*\* Ladung auf Faktor *Feedback* (.52).

Variable	
VSM_Humo	VSM humorvolle Lernatmosphäre (Mittelwert aus HUMO_MM und HUMO_AD, gesamte Lektion); G-Koeff fachspezifisch = .81 Datengrundlage N = 49 (ohne Schulungsvideos)
VSM_Anes	VSM Anerkennung der Schüler durch die Lehrperson (Mittelwert aus ANES_MM und ANES_AD, gesamte Lektion); G-Koeff fachspezifisch = .81 Datengrundlage N = 49 (ohne Schulungsvideos)
VSM_Fuer	VSM Fürsorglichkeit, Herzlichkeit und Wärme (Mittelwert aus FUER_MM und FUER_AD, gesamte Lektion); G-Koeff fachspezifisch = .91 Datengrundlage N = 49 (ohne Schulungsvideos)
<b>Ausgeschlossene Variable</b>	
VSM_Lob*	VSM Einsatz von Lob (Mittelwert aus LOB_MM und LOB_AD, gesamte Lektion); G-Koeff fachspezifisch = .89 Datengrundlage N = 49 (ohne Schulungsvideos)

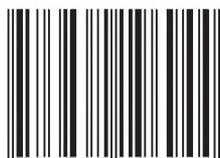
\* Bei der Berechnung der Reliabilität der Skala „Feedback“ zeigte sich, dass das Cronbachs Alpha ohne das Merkmal „VSM\_Lob“ von  $\alpha = .52$  auf  $\alpha = .72$  steigt. Aus diesem Grund wurde das Merkmal „VSM\_Lob“ bei der Skalenbildung nicht berücksichtigt. Die Skala „Feedback“ besteht demnach nur aus den zwei Merkmalen „VSM\_Emut“ und „VSM\_Rück“. „VSM\_Lob“ kann jedoch weiterhin als Einzelitem verwendet werden.

Eine effiziente Klassenführung und ein unterstützendes Unterrichtsklima gelten als Basisdimensionen guten Unterrichts. Ihnen wird eine hohe Bedeutung sowohl für die Leistungs- als auch für die Persönlichkeitsentwicklung von Schülerinnen und Schülern zugeschrieben. Dass es sich bei den beiden Merkmalen um fachübergreifend gültige Merkmale handelt, wird vielfach behauptet, jedoch bislang kaum empirisch geprüft.

Unter Berücksichtigung bisheriger videobasierter Studien und aktueller Forschungsergebnisse zur Unterrichtsqualität in der Grundschule zielt die vorliegende Studie im Projekt PERLE (*P*ersönlichkeits- und *L*ernentwicklung von Grundschulkindern) darauf ab, die Klassenführung und das Unterrichtsklima im Anfangsunterricht reliabel, mithilfe eines neu entwickelten hoch inferenten Ratingsystems, zu erfassen und hinsichtlich ihrer Ausprägungen in den Fächern Deutsch (1. Schuljahr) und Mathematik (2. Schuljahr) zu beschreiben. Es können keine Unterschiede in der Höhe der Ausprägungen zwischen den Fächern/Schuljahren bestätigt werden.

Im Rahmen der Studie werden für das jeweilige Fach zusätzlich die Wirkungen der beiden Unterrichtsqualitätsmerkmale auf die Leistungs- und Selbstkonzeptentwicklung untersucht.

ISBN 978-3-86219-694-4



9 783862 196944 >