




# Prozess- und strukturorientierte Klassenführungsmaßnahmen von Lehrpersonen im Anfangsunterricht – Ergebnisse zur zeitlichen Stabilität von Beobachterratings innerhalb und zwischen 90-minütigen Unterrichtseinheiten

Katrin Gabriel-Busse  · Armin Jentsch  · Frank Lipowsky 

Eingegangen: 1. Mai 2021 / Überarbeitet: 30. September 2021 / Angenommen: 1. Oktober 2021 / Online publiziert: 22. Oktober 2021  
© Der/die Autor(en) 2021

**Zusammenfassung** Der Erfassung des Unterrichtsqualitätsmerkmals Klassenführung durch Beobachterratings wurde in den letzten Jahren vor allem im Rahmen von Studien in der Sekundarstufe viel Aufmerksamkeit geschenkt. Hier liegt ausreichend Evidenz für eine eher geringe bis moderate Variabilität der Klassenführung zwischen mehreren Unterrichtsstunden gemessen im Abstand von wenigen Wochen und Monaten vor. Bislang kaum untersucht wurde die Frage nach der Variabilität der Klassenführung innerhalb von Unterrichtseinheiten (Doppelstunden), d.h. innerhalb von inhaltlich und methodisch aufeinander aufbauenden Unterrichtsstunden, die zeitlich direkt nacheinander beobachtet wurden. Des Weiteren liegen keine Studien vor, in denen die zeitliche Stabilität einzelner Klassenführungsmerkmale auf Itemebene (prozess- versus strukturorientierte Maßnahmen der Klassenführung) für den Anfangsunterricht der Grundschule untersucht wurde. Die hier dargestellte Studie knüpft an diese Desiderate an und untersucht die Stabilität einzelner Klassenführungsmerkmale sowohl innerhalb von 90-minütigen Unterrichtseinheiten als auch über zwei im Abstand von 12 Monaten beobachteten Unterrichtseinheiten. Die Klassenführung wurde von 21 Lehrpersonen im 1. und 2. Schuljahr jeweils zwei Mal in einer Doppelstunde (jeweils nach 45 Minuten) videobasiert erfasst. Zwei geschulte

---

Jun.-Prof.\*in Dr. Katrin Gabriel-Busse (✉)

Institut für Erziehungswissenschaft, Johannes Gutenberg-Universität Mainz,  
Jakob-Welder-Weg 12, 55128 Mainz, Deutschland  
E-Mail: kgabriel@uni-mainz.de

Dr. Armin Jentsch

Institut für Erziehungswissenschaft, Schulpädagogik, Universität Greifswald,  
Ernst-Lohmeyer-Platz 3, 17489 Greifswald, Deutschland  
E-Mail: armin.jentsch@uni-greifswald.de

Prof. Dr. Frank Lipowsky

Fachgebiet Empirische Schul- und Unterrichtsforschung, Universität Kassel,  
Nora-Platiel-Str. 1, 34109 Kassel, Deutschland  
E-Mail: lipowsky@uni-kassel.de

Beobachter/innen schätzten unabhängig voneinander sechs Klassenführungsmerkmale (prozess- vs. strukturorientierte Merkmale) reliabel ein. Die Ergebnisse deuten auf Unterschiede in der Variabilität zwischen prozess- und strukturorientierten Klassenführungsmerkmalen innerhalb der 90-minütigen Unterrichtseinheiten. Eine große Variabilität in den Beobachterratings zur Klassenführung zeigt sich – wie erwartet – zwischen 12 Monate auseinanderliegenden Unterrichtseinheiten. Die Befunde werden im Zusammenhang mit der Wahl der Analyseebene (Item- vs. Skalenebene) und der Konzeptualisierung der Klassenführung im Anfangsunterricht diskutiert.

**Schlüsselwörter** Klassenführung · Beobachterratings · Generalisierbarkeitsstudie · Grundschule

### **Process- and structure-oriented techniques of teachers' classroom management in elementary school—findings about the stability of observer ratings during and between 90-minute lessons**

**Abstract** Recording classroom management, a characteristic of lesson quality, with observer ratings has received great attention within the last years, especially for studies in secondary school, resulting in (sufficient) evidence of the variability that exists between lessons, ranging from rather low to moderate. These lessons were observed over a long period of time, and variability was measured at specific time points during that time (over a few weeks and months). However, so far, only little research exists on the extent to which observer ratings of techniques of classroom management vary during a 90-minute lesson, i.e. during lessons that are designed to build on each other in terms of contents and method; these lessons were observed one after the other. Furthermore, there are only a few studies in which the stability of different characteristics of classroom management at item level (process- and structure-oriented) were examined for the beginning grades in (or in the first school years in) elementary school.

The study presented here departs from these desiderata and investigates the stability of different characteristics of techniques of classroom management, both during as well as between two 90-minute lessons that were observed every 12 months. We measured classroom management in 21 classes during two 90-minute lessons and for two segments per lesson. Two trained observers rated six characteristics of classroom management (process-oriented and structure-oriented) independently. The results indicate acceptable reliabilities for all of the six characteristics of classroom management. In addition, variability differed between process-oriented and structure-oriented characteristics of classroom management during 90-minute lessons. Between two 90-minute lessons that were observed only once every 12 months, there is a high variability for process-oriented as well as structure-oriented characteristics of classroom management. We discuss our findings in the context of the analysis level chosen (item level vs. scale level) and with regard to the conceptualization of classroom management in the first school years in elementary school.

**Keywords** Classroom management · Observer ratings · Generalizability theory · Elementary school

## 1 Theoretischer Hintergrund

Eine effektive Klassenführung hat sich in den letzten Jahrzehnten immer wieder als wichtiger Prädiktor von schulischen Lernleistungen auch in der Grundschule erwiesen (zusf. Gabriel 2014; Helmke und Weinert 1997; Pianta et al. 2008; Cadima et al. 2010). Sie gilt als sogenannte Basisdimension guten Unterrichts (z. B. Klieme 2019; Praetorius et al. 2018) und kann nach Seidel (2020) als ein „Syndrom“ aus verschiedenen Maßnahmen und Unterrichtsstrategien verstanden werden, mit dem Ziel für alle Schüler/innen eine möglichst hohe aktive Lernzeit zu ermöglichen und die für das Lernen erforderliche Aufmerksamkeit sicherzustellen. Die bisherige Befundlage zur zeitlichen Stabilität der Klassenführung (gemessen über Beobachterratings) gibt erste Hinweise darauf, dass Klassenführung, beobachtet z. B. im Abstand von einigen Wochen und Monaten, relativ stabil ausgeprägt ist. Wird Klassenführung jedoch innerhalb einer Doppelstunde mehrfach beobachtet (vgl. Jentsch et al. 2019; Gabriel-Busse und Lipowsky 2020, eingereicht), zeigt sich eine größere zeitliche Variation in den Ausprägungen. Da allgemein von den Beobachterratings zur Klassenführung in einer beobachteten (Doppel-)Stunde auf die Effektivität der Klassenführung in einer untersuchten Klasse geschlossen wird, gilt es langfristig wirksame und demnach zeitlich stabile Unterrichtsmerkmale (z. B. in Bezug auf den Lernerfolg der Schüler/innen) zu erfassen (Jentsch et al. 2019). Dies gilt auch für den Anfangsunterricht der Grundschule, dem eine besondere Bedeutung zugeschrieben wird, da hier erstmals „die prinzipiellen Inszenierungsmuster schulischen Unterrichts“ (Ophardt 2008, S. 246) etabliert werden. Die Situation des Anfangsunterrichts stellt Lehrpersonen im Zusammenhang mit der Klassenführung vor besondere Herausforderungen, da sich die Schüler/innen mit ihren heterogenen Lernvoraussetzungen hier erstmalig zu einer Lerngemeinschaft zusammenfinden, „in der sie Regeln eines in der Klasse organisierten Lehrens und Lernens kennenlernen und verinnerlichen müssen“ (Hörter et al. 2020, S. 259). Im Vergleich zu den älteren Jahrgangsstufen der Grundschule laufen Unterrichtsaktivitäten häufig noch nicht automatisch ab bzw. folgen keinen festen Routinen, mit entsprechenden Konsequenzen für die Messung zeitlicher Stabilität der Klassenführung (Gabriel-Busse und Lipowsky 2020, eingereicht). Inwieweit sich Unterschiede in der zeitlichen Stabilität für verschiedene Merkmale der Klassenführung (prozess- vs. strukturorientierte; Doyle 1986) zeigen, wurde bislang nicht überprüft.

### 1.1 Prozess- und strukturorientierte Maßnahmen der Klassenführung

Nach Doyle (1986) gelten Handlungen von Lehrpersonen, die das aktuelle Unterrichtsgeschehen unmittelbar regulieren (z. B. „Allgegenwärtigkeit“, Kounin 2006) oder eine reibungslose Strukturierung des Unterrichts gewährleisten (z. B. „Gruppenmobilisierung“ oder ein effektives Übergangsmanagement als Aspekt des Merkmals „Reibungslosigkeit und Schwung“, Kounin 2006), als sogenannte *prozessorientierte Maßnahmen der Klassenführung* (vgl. auch Hörter et al. 2020). Nach Kounin (2006) muss die Lehrperson allgegenwärtig sein bzw. buchstäblich „Augen im Hinterkopf“ (ebd., S. 90) haben, um Störungen frühzeitig zu erkennen und gegebenenfalls sofort unterbinden zu können. Eine allgegenwärtige Lehrperson richtet ihren

Fokus häufig auf die gesamte Klasse, auch wenn sie mit einzelnen Schülern/innen oder Schüler/innengruppen interagiert. Kounin (2006) bezeichnet diese Fähigkeit als Überlappung, d. h. die Lehrperson kann gleichzeitig mehreren Geschehnissen in der Klasse Aufmerksamkeit zuwenden. Diese kontinuierliche „störungspräventive Überwachung der Schülertätigkeiten“ (Gruehn 2000, S. 129) fördert die Mitarbeit und hemmt Fehlverhalten der Schüler/innen (Kounin 2006). Die Strukturierung des Unterrichts zielt auf die „Koordination der individuellen Lernaktivitäten im Klassenverband ab“ (Hörter et al. 2020, S. 261), womit eine hohe aktive Lernzeit für alle Schüler/innen ermöglicht werden soll (Kounin 2006). Unter Strukturierung des Unterrichts werden verschiedene Maßnahmen zusammengefasst (vgl. auch Hörter et al. 2020): (1) die Gruppenmobilisierung (Kounin 2006), die einhergeht mit einer aktiveren Nutzung der Lernangebote und Beteiligung möglichst aller Schüler/innen im Unterricht, (2) die Gestaltung der Übergänge zwischen einzelnen Unterrichtsphasen bzw. dem effektiven Übergangsmanagements (als Aspekt eines effektiven Zeitmanagements) (Hörter et al. 2020).

Neben prozessorientierten Maßnahmen unterscheidet Doyle (1986) zusätzlich *strukturorientierte Maßnahmen der Klassenführung*, die auf eine langfristige und effektive Etablierung wiederkehrender Abläufe und Unterrichtsaktivitäten ausgerichtet sind. Beispielsweise können hier jene Handlungen der Lehrperson subsumiert werden, die auf die Etablierung und das Nutzen von Regeln, Routinen und Ritualen<sup>1</sup> ausgerichtet sind (Hörter et al. 2020). Die Etablierung eines effektiven Regelsystems bzw. von Routinen zielt ebenso auf eine effektive Nutzung der Unterrichtszeit sowie auf die Strukturierung der Übergangsphasen bzw. des Unterrichts allgemein ab und gilt als entscheidender Faktor erfolgreichen Lehrens und Lernens im Unterricht (Helmke 2006).

## 1.2 Zeitliche Stabilität von Beobachterratings zur Klassenführung

Im Rahmen von Beobachtungs- bzw. Videostudien werden die Merkmale der Klassenführung häufig mit Hilfe von Beobachterratings eingeschätzt und auf ihre Wirkungen hin untersucht (z. B. Seidel und Shavelson 2007; für die Grundschule: zusf. Gabriel 2014). Für den Sekundarschulbereich ist die Frage nach der zeitlichen Stabilität und Generalisierbarkeit von Beobachterratings zur Unterrichtsqualität inzwischen relativ gut untersucht worden (z. B. Jentsch et al. 2019; Meyer et al. 2011; Praetorius et al. 2014; Praetorius 2014). Anhand der Berechnung sogenannter Generalisierbarkeitsstudien (G-Studien, vgl. Cronbach et al. 1972; Shavelson und Webb 1991) zeigen entsprechende Ergebnisse eine eher schwache bzw. moderate zeitliche Variation in den Ausprägungen der Klassenführung zwischen unterschiedlichen Stunden einer Lehrperson in ein und derselben Klasse (gemessen im Abstand von wenigen Wochen oder Monaten) (vgl. Tab. 1). Sekundarlehrpersonen scheinen demnach ausgeprägte Routinen entwickelt zu haben, die sie in jeder Unterrichtsstunde und unabhängig vom Unterrichtsinhalt und Beobachtungszeitpunkt anwenden (Seidel 2020).

<sup>1</sup> Eine detaillierte Definition der Begriffe sowie Gegenüberstellung findet sich in Gabriel (2014).

**Tab. 1** Überblick über G-Studien zur zeitlichen Variation der Klassenführung

Studie	Klassenstufe	Anzahl MZP	Abstand zwischen MZP	Zeitliche Variation zwischen MZP (Effektstärke)
Mashburn et al. (2014)	6.–11.	3 (verteilt auf 1 Jahr)	Ca. 3 Monate	Moderat
Kane & Staiger (2012)	4.–8.	3 (verteilt auf 2 Jahre)	2,5 bis 8 Monate	Moderat bis groß
Meyer et al. (2011)	k. A.	4 (verteilt auf 1 Jahr); 2 pro Halbjahr	Jeweils 2–4 Wochen pro Halbjahr	Klein bis moderat
Casabianca et al. (2015)	6.–7.	2 (verteilt auf 2 Jahre)	8 Monate	Klein
Praetorius (2014; vgl. auch Praetorius et al. 2014)	9.	2 Doppelstunden	Zwischen Doppelstunden: Ø 16 Wochen	Moderat
Jentsch et al. (2019)	7.–10.	2 Doppelstunden	Innerhalb Doppelstunde: 4 MZP (im Abstand von ca. 22,5 Minuten) zwischen Doppelstunden: ca. 14 Tage	Innerhalb Doppelstunde: moderat zwischen Doppelstunden: klein
Gabriel et al. (2015; vgl. auch Gabriel-Busse und Lipowsky 2020)	1.–2.	1 Doppelstunde 1 Doppelstunde	Innerhalb Doppelstunden: 2 MZP (im Abstand von 45 Minuten)	Innerhalb Doppelstunde: groß

Anmerkungen: MZP Messzeitpunkt, k. A. keine Angaben

In Anlehnung an Cohen (1992) wird die praktische Bedeutsamkeit von Varianzanteilen als kleine (<7%), moderate (7–15%) oder große (>15%) Effektstärke definiert. Die Tabelle erhebt keinen Anspruch auf Vollständigkeit

Für den Grundschulbereich mangelt es an entsprechenden Studien bzw. die bisherigen Studien im Grundschulbereich untersuchten die zeitliche Stabilität der Klassenführung häufig in Form von Korrelationen zwischen verschiedenen Messzeitpunkten (im Abstand von mehreren Monaten), wobei vor allem für die ersten beiden Schuljahre (Anfangsunterricht) die untersuchten Zusammenhänge eher schwach ausfallen. Eckerth et al. (2012) vermuten, dass sich Schüler/innen vor allem in den ersten beiden Schuljahren (Anfangsunterricht) mit zunehmender Verweildauer besser im Unterrichtsalltag zurechtfinden (z. B. sich an Regeln halten, Routinen verinnerlichen usw.), aber auch die Lehrpersonen die Bedürfnisse der Schüler/innen besser kennen und demzufolge die Unterrichtsgestaltung und die Klassenführung anpassen, mit entsprechenden Konsequenzen für die zeitliche Stabilität der Klassenführung. Im Vergleich zum ersten und zweiten Schuljahr zeigen Studien für das dritte und vierte Schuljahr (zusf. Gabriel-Busse und Lipowsky eingereicht) eine deutlich höhere zeitliche Stabilität der Klassenführung – gemessen anhand von Korrelationen – sowohl innerhalb eines Schuljahres (Helmke und Weinert 1997) als auch über einen Zeitraum von zwei Unterrichtsstunden hinweg (Pianta et al. 2008). Da im Rahmen von Korrelationen die zeitliche Stabilität mit weiteren Faktoren konfundiert ist (Praetorius 2014; Shavelson und Dempsey-Adwood 1976), sollten weitere Faktoren (z. B. Beobachtungszeitpunkt, Merkmal der Klassenführung/eingesetzte Items)

berücksichtigt und deren Varianzanteil bestimmt werden. Hier bietet sich die Durchführung von G-Studien an, wie sie im Sekundarschulbereich in den letzten Jahren vermehrt durchgeführt wurden (vgl. Tab. 1).

Begründet wird die geringe zeitliche Variation der Klassenführung in der Sekundarstufe häufig damit, dass sich sowohl Verhaltensweisen der Lehrperson in Bezug auf die Klassenführung als auch das Verhalten der Schüler/innen nicht kurzfristig verändern und somit zu jedem Beobachtungszeitpunkt in ähnlicher Ausprägung sichtbar sein sollten (z.B. Praetorius 2014; König und Lebens 2012). Inwieweit einzelne Maßnahmen der Klassenführung jedoch kurzfristigen Schwankungen unterworfen sein können, wurde bislang kaum untersucht. Geht man davon aus, dass sich Unterricht durch eine deutliche Sequenzierung einzelner Phasen und Schritte auszeichnet (Lipowsky 2007), können lehrergesteuerte und schülerorientierte Phasen unterschiedliche Anforderungen in Bezug auf die Klassenführung stellen. Erste Hinweise darauf, dass sich beispielsweise während einer Doppelstunde Unterschiede im Klassenführungsverhalten von Lehrpersonen (Sekundarschule) zeigen, liefert die Studie von Jentsch et al. (2019). In der Studie erwies sich die Variabilität der Beobachterratings für die Klassenführung innerhalb einer Doppelstunde als beträchtlich, wobei die Klassenführung vor allem zu Beginn der Doppelstunde von den Rater/innen positiver eingeschätzt wurde. Jentsch et al. (2019) erklären die Unterschiede in den Ausprägungen der Klassenführung damit, dass vor allem zu Beginn des Unterrichts (in einer Doppelstunde) „Maßnahmen der Lehrperson zur Herstellung von Aufmerksamkeit [...] erfolgen oder zum Ende einer Unterrichtsstunde ausbleiben“ (ebd. S. 16). Dass das Klassenführungsverhalten innerhalb einer Doppelstunde stärker variiert als bislang angenommen, konnten Gabriel et al. (2015; vgl. auch Gabriel-Busse und Lipowsky 2020, eingereicht) auch für den Anfangsunterricht der Grundschule bestätigen. Für verschiedene fachliche Kontexte (Deutsch und Mathematik) zeigte sich eine hohe Variation in den Ausprägungen der Merkmale der Klassenführung innerhalb einer Doppelstunde (vgl. Tab. 1). Das zugrundeliegende Klassenführungsverhalten einer Lehrperson im Unterricht des ersten und zweiten Schuljahres ist somit auch kurzfristig bestimmten Schwankungen unterworfen.

Um Aussagen über die zeitliche Stabilität der Klassenführung machen zu können, wurden in einem Großteil der bisherigen G-Studien (Tab. 1: z. B. Gabriel-Busse und Lipowsky 2020; Jentsch et al. 2019; Praetorius 2014; Meyer et al. 2011) Analysen auf Itemebene durchgeführt und die Items als zusätzliche Facette (*i*) in das Design der G-Studie integriert. Dem liegt die Annahme zugrunde, dass im Rahmen der Unterrichtsqualitätsforschung Interaktionen zwischen einzelnen Unterrichtssequenzen bzw. Lehrpersonen und Items (z. B. in Bezug auf die Klassenführung) erwartbar und demnach nicht als Fehlervarianz anzusehen sind (Praetorius 2014). In der Studie von Jentsch et al. (2019) ergibt sich so beispielsweise für die Klassenführung ein relativ großer itemspezifischer Varianzanteil (*i*) gepaart mit einem moderaten Interaktionseffekt  $k \times i$ , der wiederum bedeutet, dass Klassenunterschiede bzw. Unterschiede zwischen Lehrpersonen in der Klassenführung sogar innerhalb von Doppelstunden itemspezifisch ausfallen und nicht allein auf Unterschiede in den zugrundeliegenden Klassenführungsmerkmalen zurückzuführen sind. Zu ähnlichen Ergebnissen für die Variabilität innerhalb einer Doppelstunde kommen auch Gabriel-Busse und Lipowsky (2020; vgl. auch Gabriel et al. 2015) für den Anfangsunterricht in der

Grundschule. Der itemspezifische Varianzanteil, d. h. die zeitstabilen Unterschiede in den Itemschwierigkeiten innerhalb einer Skala sind hier zwar eher gering, der Interaktionseffekt ( $u: l) x i$  ist mit einem Varianzanteil von 17 % (1. Schuljahr) bzw. 16 % (2. Schuljahr) jedoch moderat. Die Autor/innen interpretieren dies dahingehend, dass beispielsweise in Abhängigkeit von der jeweiligen Phase im Unterricht bzw. der jeweiligen Teilstunde der Doppelstunde Stärken und Schwächen der Lehrpersonen hinsichtlich der einzelnen Klassenführungsmerkmale ersichtlich werden.

## 2 Ableitung der Forschungsfragen

In Ergänzung zu den bisherigen Studien (im Sekundar- und Grundschulbereich) soll im vorliegenden Beitrag der Frage nachgegangen werden, inwieweit Unterschiede in der zeitlichen Stabilität in Abhängigkeit vom untersuchten Klassenführungsmerkmal (prozess- und strukturorientiert) bestehen. Dabei liegt der Fokus auf der zeitlichen Stabilität (1) innerhalb von Doppelstunden (zwei Messzeitpunkte im Abstand von 45 Minuten) und (2) messwiederholt gemessen im Abstand von 12 Monaten.

**Forschungsfrage Nr. 1:** *Inwieweit zeigen sich Unterschiede in der zeitlichen Stabilität für einzelne Klassenführungsmerkmale in Abhängigkeit vom Abstand zwischen den Beobachtungen (innerhalb einer Doppelstunde vs. im Abstand von 12 Monaten)?*

Aufgrund der besonderen Herausforderung in Bezug auf die Klassenführung im Anfangsunterricht und in Anlehnung an bisherige Studien (s. oben), kann angenommen werden, dass das Klassenführungsverhalten der Lehrpersonen sowohl innerhalb einer Doppelstunde als auch beobachtet im Abstand von 12 Monaten zeitlich variiert (vgl. Abschn. 1.2) (*Hypothese 1*).

Zusätzlich wird im Rahmen der vorliegenden Studie untersucht, inwieweit in Anlehnung an Doyle (1986) Unterschiede in der zeitlichen Stabilität bestehen, je nachdem, ob es sich um eine prozess- oder strukturorientierte Maßnahme der Klassenführung handelt.

**Forschungsfrage Nr. 2:** *Zeigen sich Unterschiede in der zeitlichen Stabilität zwischen prozess- und strukturorientierten Maßnahmen der Klassenführung?*

Prozessorientierte Maßnahmen regulieren eher das aktuelle Unterrichtsgeschehen. Sie sind weniger planbar und sollten somit zeitlich variabler sein (*Hypothese 2*). Strukturorientierte Maßnahmen sind hingegen eher auf die langfristige effektive Etablierung wiederkehrender Abläufe und Unterrichtsaktivitäten ausgerichtet (vgl. Abschn. 1.1) und sollten demnach zeitlich stabiler ausfallen (*Hypothese 3*). Inwieweit dies für den Anfangsunterricht der Grundschule zutrifft, soll explorativ überprüft werden.

### 3 Methodisches Vorgehen

#### 3.1 Stichprobe

Die Überprüfung der Forschungsfragen erfolgt anhand der Daten aus der Videostudie Deutsch/Sprache (VSS) und Videostudie Mathematik (VSM) der PERLE-Studie (Persönlichkeits- und Lernentwicklung von Grundschulkindern; Greb et al. 2007; Lipowsky et al. 2013). Die VSS fand im ersten Schuljahr (März 2007), die VSM im zweiten Schuljahr (März 2008) statt (Greb et al. 2009; Lotz et al. 2011; Lotz und Corvacho del Toro 2013). Für insgesamt 21 Lehrpersonen liegen vollständige Daten zur Klassenführung für beide Fächer vor<sup>2</sup>. Diese bilden die Grundlage für die nachfolgenden Analysen. Von jeder dieser Lehrpersonen ( $\bar{x} = 20$ ;  $\sigma = 1$ ) wurden jeweils 90 Minuten (Doppelstunde) Deutsch- (1. Schuljahr) bzw. Mathematikunterricht (2. Schuljahr) in ein und derselben Klasse beobachtet. Vier Lehrpersonen sind unter 35 Jahre, 14 Lehrpersonen sind zwischen 35 und 55 Jahre und eine Lehrperson ist über 55 Jahre alt. Im Mittel haben die teilnehmenden Lehrpersonen 14,5 Jahre Berufserfahrung ( $N = 17$ ;  $Min = 1$ ;  $Max = 35$ ;  $SD = 10,7$ ). Die Angaben zu Alter und Berufserfahrung der Lehrpersonen stammen aus einem Fragebogen vom November 2006. Für zwei Lehrpersonen liegen keine Angaben zum Alter, für vier Lehrpersonen keine Angaben zur Berufserfahrung vor. Durchschnittlich nahmen pro Klasse 18 Schüler/innen sowohl an der VSS als auch an der VSM teil (vgl. Lotz und Corvacho del Toro 2013; Mösko et al. 2013).

#### 3.2 Das hoch inferente Ratingsystem zur Klassenführung

Die Klassenführung wurde mit Hilfe eines hoch inferenten Ratingsystems erfasst. Für jedes der in Tab. 2 aufgeführte Klassenführungsmerkmal wurde die Grundidee beschrieben sowie diverse verhaltensnahe Indikatoren/Negativindikatoren formuliert (Gabriel und Lipowsky 2013). Zwei geschulte Rater/innen (Student/innen des Lehramts, höheres Semester) schätzten unabhängig voneinander insgesamt sechs Merkmale der Klassenführung auf einer vier-stufigen Skala ein. Während einer zweitägigen Schulung wurden die Rater/innen zuerst mit dem Aufbau des Ratingsystems vertraut gemacht. Hierbei wurden den Ratern/innen Ankerbeispiele und Videoausschnitte für jedes zum Rating anstehende Klassenführungsmerkmal vorgestellt. In Anlehnung an Korbarg und Seidel (2003) galt das Training ab dem Zeitpunkt als

<sup>2</sup> Die hier verwendete Stichprobe weicht von der bei Gabriel-Busse und Lipowsky (2020) verwendeten Stichprobe ab. Im Unterschied zu den Analysen von Gabriel et al. (2015) und den hier dargestellten Ergebnissen war es das Ziel der Studie von Gabriel-Busse und Lipowsky (2020) zu überprüfen, ob sich die Einschätzungen der Rater/innen über die drei Basisdimensionen der Unterrichtsqualität in Abhängigkeit von der jeweiligen Unterrichtsstunde (1. Stunde vs. 2. Stunde einer Doppelstunde) im Fach Mathematik (2. Schuljahr) unterscheiden. Aus diesem Grund konnten die Daten von allen 36 Mathematiklehrpersonen der PERLE-Studie genutzt werden. Im vorliegenden Beitrag geht es jedoch um die Frage, inwieweit sich Unterschiede in der Ausprägung der Klassenführung bei ein und derselben Lehrperson in ein und derselben Klasse innerhalb von Doppelstunden und über zwei Unterrichtseinheiten im Abstand von 12 Monaten zeigen. Aus diesem Grund gingen nur die Daten der diejenigen Lehrpersonen in die Analysen ein, für die entsprechende Daten sowohl aus der 1. Schuljahr (Deutsch) als auch aus dem 2. Schuljahr (Mathematik) vorliegen.

**Tab. 2** Mittelwerte, Standardabweichungen, bivariante Korrelationen sowie Inter-Rater-Übereinstimmung auf Item- und Skalenebene getrennt nach prozess- und strukturorientierten Maßnahmen der Klassenführung

	1. Schuljahr (März 2007)				2. Schuljahr (März 2008)				Bivariate Korrelation 1. SJ/ 2. SJ <sup>a</sup>		
	Doppelstunde		Inter-Rater-Übereinstimmung	Bivariate Korrelation		Doppelstunde		Inter-Rater-Übereinstimmung		Bivariate Korrelation	
	1. Stunde MW (SD)	2. Stunde MW (SD)		1. Stunde/ 2. Stunde	2. Stunde MW (SD)	1. Stunde MW (SD)	2. Stunde MW (SD)			1. Stunde/ 2. Stunde	1. Stunde/ 2. Stunde
<b>Prozessorientierte Maßnahmen der Klassenführung</b>											
Allgegenwärtigkeit der Lehrperson	3,11 (0,79)	2,86 (0,79)	0,90	3,24 (0,58)	2,83 (0,78)	0,43*	3,24 (0,58)	2,83 (0,78)	0,15 ns	0,88	0,42 ns
Gruppenmobilisierung	3,29 (0,77)	3,29 (0,64)	0,82	3,40 (0,44)	3,38 (0,42)	0,79***	3,40 (0,44)	3,38 (0,42)	0,49*	0,92	0,04 ns
Übergangsmanagement („Managing Transitions“)	3,10 (0,70)	2,93 (0,66)	0,79	3,05 (0,71)	2,86 (0,48)	0,67***	3,05 (0,71)	2,86 (0,48)	0,69***	0,94	0,35 ns
<b>Subskala prozessorientierte Maßnahmen (3 Items)</b>	<b>3,17 (0,62)</b>	<b>3,02 (0,55)</b>	–	<b>3,23 (0,38)</b>	<b>3,02 (0,41)</b>	<b>0,62**</b>	<b>3,23 (0,38)</b>	<b>3,02 (0,41)</b>	<b>0,64**</b>	–	<b>0,27 ns</b>

Tab. 2 (Fortsetzung)

	1. Schuljahr (März 2007)				2. Schuljahr (März 2008)				Bivariate Korrelation	
	Doppelstunde		Doppelstunde		Doppelstunde		Doppelstunde		1. SJ/	2. SJ <sup>a</sup>
	1. Stunde MW (SD)	2. Stunde MW (SD)	Bivariate Korrelation 1. Stunde/ 2. Stunde	Inter-Rater- Überein- stimmung	1. Stunde MW (SD)	2. Stunde MW (SD)	Bivariate Korrelation 1. Stunde/ 2. Stunde	Inter-Rater- Überein- stimmung		
<b>Strukturorientierte Maßnahmen der Klassenführung</b>										
Regelklarheit bzw. -verwendung (Lehrperson)	3,07 (0,86)	3,12 (0,77)	0,78***	0,68	3,26 (0,54)	3,12 (0,52)	0,59***	0,95		0,22 <sup>ns</sup>
Regeln für soziales Miteinander (Schüler/ innen)	3,43 (0,48)	3,10 (0,44)	0,15 <sup>ns</sup>	0,85	3,43 (0,55)	3,24 (0,74)	0,66***	0,81		0,27 <sup>ns</sup>
Effektiver Umgang mit auftretenden Unter- richtsstörungen	2,95 (0,86)	2,93 (0,62)	0,39 <sup>ns</sup>	0,85	3,24 (0,46)	3,00 (0,72)	0,48*	0,86		0,36 <sup>ns</sup>
<b>Subskala strukturori- entiertere Maßnahmen (3 Items)</b>	<b>3,15 (0,63)</b>	<b>3,05 (0,52)</b>	<b>0,63**</b>	–	<b>3,31 (0,47)</b>	<b>3,11 (0,57)</b>	<b>0,54*</b>	–		<b>0,28<sup>ns</sup></b>
<b>Skala Klassenführung (Gesamt 6 Items)</b>	<b>3,16 (0,57)</b>	<b>3,04 (0,49)</b>	<b>0,64**</b>	–	<b>3,27 (0,37)</b>	<b>3,07 (0,43)</b>	<b>0,68***</b>	–		<b>0,29<sup>ns</sup></b>

Anmerkung: N 21, MW Mittelwert, SD Standardabweichung; Antwortformat vier-stufig; fett gedruckt: Ergebnisse auf (Sub-)Skalenebene

\*\*\*  $p \leq 0,001$ ; \*\*  $p \leq 0,01$ ; \*  $p \leq 0,05$ , <sup>ns</sup> nicht signifikant

<sup>a</sup>Für jedes Schuljahr wurde ein Mittelwert über die 1. und 2. Stunde der Doppelstunde berechnet. Diese Mittelwerte gingen in die Berechnung der bivariaten Korrelationen ein

abgeschlossen, als sich die Rater/innen und die Entwicklerin des Ratingsystems über die Zuordnung beobachtbaren Verhaltens zu einem hoch inferenten Merkmal der Klassenführung einig waren (konsensuelle Validierung, vgl. Bortz und Döring 2006, S. 328).

Die erste Einschätzung bezog sich auf die ersten 45 Minuten der Doppelstunde, die zweite Einschätzung auf die zweiten 45 Minuten der Doppelstunde. Der Unterteilung einer Doppelstunde in zwei Segmente gingen inhaltliche Überlegungen voraus: Da junge Schüler/innen im Gegensatz zu älteren eher kurze Konzentrationszeiten haben<sup>3</sup>, erschien es vielen Lehrpersonen aus der vorliegenden Stichprobe sinnvoll, eine Pause nach den ersten 45 Minuten zu machen. Im Nachgang der Videostudien (Interview) begründeten die Lehrpersonen dies u. a. damit, dass 90 Minuten Unterricht am Stück eher untypisch wären.

Die Inhalte der Doppelstunde wurden vorgegeben (1. SJ Deutsch: Lotz und Corvacho del Toro 2013; 2. SJ Mathematik: Mösko et al. 2013). Die Einschätzungen erfolgten somit für zwei Unterrichtseinheiten in zwei verschiedenen Fächern im Abstand von ca. 12 Monaten, jeweils bestehend aus zwei aufeinander aufbauenden Stunden einer Lehrperson in ein und derselben Klasse. Zur Überprüfung der Qualität der hoch inferenten Daten wurde für jedes der sechs Klassenführungsmerkmale auf Basis aller Videos der PERLE-Studie (1. SJ Deutsch:  $N=48$  Videos, 2. SJ Mathematik:  $N=49$  Videos) der relative Generalisierbarkeitskoeffizient (G-Koeffizient) als Maß für die Übereinstimmung zwischen den zwei Rater/innen (Inter-Rater-Reliabilität) berechnet (Gabriel und Lipowsky 2013). Die Inter-Rater-Übereinstimmungen für die einzelnen Klassenführungsmerkmale (berechnet über die gesamte Doppelstunde) liegen für das 1. Schuljahr (Deutsch) zwischen 0,68 und 0,90 und für das 2. Schuljahr (Mathematik) zwischen 0,81 und 0,95 (vgl. Tab. 2) und können als zufriedenstellend angesehen werden.

### 3.3 Design der G-Studien

Um die Fragestellungen zu überprüfen, werden die unterschiedlichen Klassenführungssitem – im Gegensatz zum bisherigen Vorgehen im Rahmen von G-Studien (vgl. Abschn. 1.2) – nicht als zusätzliche Facette in das Untersuchungs-Design der G-Studie integriert, sondern es wird für jedes Klassenführungsmerkmal eine separate G-Studie berechnet (vgl. Abschn. 4.2).

Den Analysen liegt ein geschachteltes Zwei-Facetten-Design ( $(s: l) \times j$ -Design) zugrunde, mit dem es möglich ist, insgesamt fünf Varianzquellen für jedes der acht Klassenführungsmerkmale zu unterscheiden. Neben dem zu erfassenden Merkmal (stabile Ausprägung des jeweiligen Merkmals der Klassenführung) einer Lehrperson ( $l$ ) (bzw. Klasse) wurden folgende Facetten (d. h. Varianzquellen) einbezogen:

- die Unterrichtsstunden (geschachtelt in Lehrpersonen;  $s: l$ ) (= Varianz, die auf die Unterschiede einer Lehrperson zwischen der 1 und 2. Stunde innerhalb einer Doppelstunde entfällt),

<sup>3</sup> Thal und Ebert (2004) geben als Richtzeiten für die Aufmerksamkeitsphasen von 5- bis 10-Jährigen 15 bis 20 Minuten an.

- das Schuljahr ( $j$ )<sup>4</sup> (= Varianz, die auf die Unterschiede zwischen den 12 Monaten auseinanderliegenden Unterrichtseinheiten entfällt),
- die Interaktion Lehrperson-Schuljahr ( $l \times j$ ) (= Varianz, die auf Unterschiede einer Lehrperson zwischen den 12 Monaten auseinanderliegenden Unterrichtseinheiten entfällt),
- die Residualvarianz bzw. Interaktion höchster Ordnung ( $(s: l) \times j, e$ ).

Eine Analyse der Rater-Effekte steht nicht im Fokus der Untersuchung. Die beiden Rater-Urteile wurden daher pro Messzeitpunkt durch Mittelwertbildung zusammengefasst. Dieses Vorgehen wird von Pietsch und Tosana (2008) als Maßnahme „zur Sicherung der Urteilsqualität bei Unterrichtsbeobachtungen“ diskutiert (S. 438). Da Studien zeigen, dass das Ausmaß an Varianz in den Beobachtungsratings in Abhängigkeit von der Analyseebene (Item- vs. Skalenebene) variiert (z. B. Praetorius 2014), werden in weiteren Schritt zusätzlich die über die einzelnen Klassenführungsitems gemittelten Skalenwerte für zusätzliche G-Studien herangezogen, wobei neben der Skala Klassenführung auch die zwei Subskalen der Klassenführung (prozess- und strukturorientierte Maßnahmen) überprüft werden (vgl. Tab. 3).

## 4 Ergebnisse

### 4.1 Deskriptive Statistiken

Tab. 2 stellt die deskriptiven Statistiken für die prozess- und strukturorientierten Klassenführungsmerkmale für die 1. und 2. Stunde der Doppelstunde getrennt nach Schuljahren und jeweils auf Item- und (Sub-)Skalenebene dar. Die Klassenführungsmerkmale sind durchgehend positiv ausgeprägt ( $MW_{1. SJ 1. Stunde} \geq 2,95$ ;  $MW_{1. SJ 2. Stunde} \geq 2,86$ ;  $MW_{2. SJ 1. Stunde} \geq 3,05$ ;  $MW_{2. SJ 2. Stunde} \geq 2,83$ ). Während die zwei Rater/innen sowohl die prozess- als auch die strukturorientierten Klassenführungsmerkmale im 2. Schuljahr jeweils in der 2. Stunde der Doppelstunde niedriger einschätzen, kann dieser Trend für das 1. Schuljahr nur auf Skalenebene, nicht aber für alle Klassenführungsmerkmale auf Itemebene bestätigt werden: „Gruppenmobilisierung“ wird im 1. Schuljahr im Mittel von den zwei Rater/innen in beiden Stunden gleich hoch ( $MW=3,29$ ), „Regelklarheit bzw. -verwendung“ in der 1. Stunde hingegen höher eingeschätzt. Die bivariaten Korrelationen zwischen den zwei Stunden liegen je nach Klassenführungsmerkmal im 1. Schuljahr zwischen  $r=0,15$  und  $r=0,79$ , wobei für die strukturorientierten Merkmale „Regeln für das soziale Miteinander (Schüler/innen)“ und „Effektiver Umgang mit Unterrichtsstörungen“ im ersten Schuljahr (noch) kein signifikanter Zusammenhang besteht. Im 2. Schuljahr liegen die bivariaten Korrelationen zwischen  $r=0,15$  und  $r=0,69$  (vgl. Tab. 2). Lediglich für

<sup>4</sup> Im Vergleich zu den G-Studien, die bislang im Rahmen der PERLE -Studie durchgeführt wurden (Gabriel et al. 2015; Gabriel-Busse und Lipowsky 2020) unterscheidet sich das hier zugrunde liegende Design dahingehend, dass die Facetten Rater ( $r$ ) und Items ( $i$ ) nicht berücksichtigt wurden. Die Berechnung eines geschachtelten Drei-Facetten-Designs ( $(s: j: l) \times r$  - Design) führte aufgrund der Komplexität des Modells zu einer größeren Anzahl an negativen Varianzen vor allem für die Interaktion zwischen Lehrperson  $\times$  Rater ( $l \times r$ ).

das Merkmal „Allgegenwärtig der Lehrperson“ lässt sich kein signifikanter Zusammenhang zwischen den beiden Teilstunden der Doppelstunde bestätigen. Über ein Schuljahr hinweg zeigt sich sowohl auf Item- als auch auf (Sub-)Skalenebene kein signifikanter Zusammenhang (vgl. rechte Spalte, Tab. 2).

#### 4.2 Unterschiede in der zeitlichen Stabilität für prozess- und strukturorientierte Maßnahmen der Klassenführung – Ergebnisse der G-Studien

Die Ergebnisse der G-Studien (vgl. Tab. 3) zeigen, dass sich die Unterschiede in den Ausprägungen für alle Klassenführungsmerkmale auf Unterschiede zwischen den Lehrpersonen in den Klassen zurückführen lassen (stabile Varianz:  $I$ ), der Anteil liegt zwischen 5 % und 21 %. Ein deutlich größerer Varianzanteil entfällt auf die Interaktion von Lehrperson und Schuljahr ( $I \times j$ ). Je nach Klassenführungsmerkmal liegt dieser zwischen 21 % (prozessorientiertes Merkmal: Allgegenwärtigkeit der Lehrperson) und 50 % (strukturorientiertes Merkmal: Regelklarheit bzw. -verwendung). Je höher dieser Varianzanteil, desto deutlicher variieren die Ausprägungen der Klassenführungsmerkmale einer Lehrperson in ein und derselben Klasse *zwischen dem ersten und zweiten Schuljahr*. Lehrpersonen zeigen demnach ein unterschiedliches Klassenführungsverhalten in Abhängigkeit vom untersuchten Schuljahr im Anfangsunterricht, wobei keine eindeutige Aussage in Bezug auf Unterschiede zwischen prozess- und strukturorientierten Merkmalen der Klassenführung getroffen werden kann, da bspw. auch das Merkmal „Gruppenmobilisierung“ als eine prozessorientierte Maßnahme einen ähnlich hohen Varianzanteil (45 %) aufweist wie das Merkmal „Regelklarheit bzw. -verwendung“. Auf Ebene der Subskalen zeigt sich ein eindeutigeres Bild: Für die prozessorientierten Maßnahmen der Klassenführung entfällt mit 39 % gegenüber 24 % für die strukturorientierten Maßnahmen ein deutlich größerer Varianzanteil auf Unterschiede zwischen den beiden Schuljahren. Hypothese 2 und 3 können somit bestätigt werden: Strukturorientierte Maßnahmen der Klassenführung sind – wenn einmal etabliert – zeitlich stabiler. Betrachtet man ergänzend die Ergebnisse für die Gesamtskala Klassenführung zeigt sich, dass für prozess- und strukturorientierte Maßnahmen ca. 60 % der Varianz aufgeklärt werden können, wobei ca. ein Drittel auf Unterschiede zwischen Lehrpersonen entfällt und zwei Drittel auf Unterschiede zwischen Fächern/Schuljahren entfallen. Der Varianzanteil, der auf unterschiedliches Klassenführungsverhalten *innerhalb einer Doppelstunde* (1. Stunde vs. 2. Stunde,  $s$ :  $I$ ) zurückgeführt werden kann, variiert sowohl für prozess- als auch für strukturorientierte Maßnahmen der Klassenführung gleichermaßen, wobei sich die Höhe der Ausprägung für das Merkmal „Allgegenwärtigkeit der Lehrperson“ mit 21 % deutlich zwischen den zwei Stunden einer Doppelstunde unterscheidet (vgl. Tab. 2). Zusätzlich entfallen je nach Klassenführungsmerkmal bis zu 55 % der Varianz auf das Residuum.

**Tab. 3** Ergebnisse der G-Studien für prozess- und strukturorientierte Maßnahmen der Klassenführung

	Lehrperson ( <i>l</i> )	Lehrperson × Schuljahr ( <i>l x j</i> )	Stunde: Lehrperson ( <i>s: l</i> )	Residuum ( <i>(s: l) x j, e</i> )	Total	Relative Fehler	Relative G-Koeffizient <sup>b</sup>
<b>Prozessorientierte Maßnahmen der Klassenführung</b>							
Allgegenwärtigkeit der Lehrperson	VK 0,07	0,14	0,14	0,31	0,66	0,22	0,26
	% 11	21	21	47	100	–	–
Gruppenmobilisierung	VK 0,02	0,22	0 <sup>a</sup>	0,25	0,49	0,17	0,12
	% 5	45	0	50	100	–	–
Übergangmanagement („Managing Transition“)	VK 0,12	0,16	0,01	0,27	0,55	0,15	0,44
	% 21	29	2	49	100	–	–
<b>Subskala prozessorientierte Maßnahmen (3 Items)</b>	<b>VK 0,05</b>	<b>0,10</b>	<b>0,01</b>	<b>0,10</b>	<b>0,26</b>	<b>0,08</b>	<b>0,38</b>
	% 19	39	2	40	100	–	–

Tab. 3 (Fortsetzung)

	Lehrperson (l)	Lehrperson x Schuljahr (l, x, j)	Stunde: Lehrperson (s: l)	Residuum (s: l, x, j, e)	Total	Relative Fehler	Relative G-Koeffizient <sup>b</sup>
<b>Strukturorientierte Maßnahmen der Klassenführung</b>							
Regelklarheit bzw. -verwendung (Lehrperson)	VK 0,08 % 14	0,28 50	0,01 1	0,20 35	0,56 100	0,19 -	0,29 -
Regeln für soziales Miteinander (Schüler/innen)	VK 0,03 % 8	0,14 35	0,05 12	0,17 45	0,39 100	0,13 -	0,19 -
Effektiver Umgang mit auftretenden Unterrichtsstörungen	VK 0,10 % 17	0,13 22	0,04 6	0,32 55	0,58 100	0,16 -	0,38 -
<b>Subskala strukturorientierte Maßnahmen (3 Items)</b>	<b>VK 0,08</b> % 24	<b>0,09</b> 28	<b>0<sup>a</sup></b> 0	<b>0,16</b> 48	<b>0,33</b> 100	<b>-</b> <b>0,09</b>	<b>-</b> <b>0,48</b>
<b>Klassenführung Gesamt (6 Items)</b>	<b>VK 0,05</b> % 23	<b>0,09</b> 38	<b>0<sup>1</sup></b> 0	<b>0,09</b> 39	<b>0,23</b> 100	<b>-</b> <b>0,07</b>	<b>-</b> <b>0,44</b>

Anmerkung: N 21 Lehrpersonen, VK Varianzkomponente, % prozentualer Varianzanteil (gerundete Werte). Die Varianz, die auf den jeweiligen Messzeitpunkt bzw. das Schuljahr entfällt, liegt für alle Klassenführungsmerkmale durchgängig bei Null (keine zusätzliche Spalte in der Tabelle erforderlich). fett gedruckt: Ergebnisse auf (Sub-)Skalenebene

<sup>a</sup> Negative Varianzkomponentenschätzung mit prozentualem Anteil < 5 % wurden auf Null gesetzt

<sup>b</sup> Die G-Koeffizienten weichen von bisher im Rahmen der PERLE-Studie publizierten Studienergebnissen ab, da diese bislang separat für jedes Schuljahr berechnet wurden (z. B. Gabriel et al. 2015). Unterschiede in der Klassenführung zwischen Schuljahren wurden demnach als feste Effekte modelliert, die im Rahmen der G-Theorie keinen Einfluss auf die Zuverlässigkeit einer Messung haben (Shavelson und Webb 1991). In den hier berichteten Modellen werden diese Unterschiede dagegen als Zufallseffekte berücksichtigt. Damit kann die Fragestellung, inwieweit die Generalisierbarkeit über 12 Monate hinweg gegeben ist, explizit untersucht werden

## 5 Zusammenfassung und Diskussion

### 5.1 Die zeitliche Variation prozess- und strukturorientierter Maßnahmen der Klassenführung innerhalb einer Doppelstunde

Im Rahmen der G-Studie erwies sich die Variabilität der Klassenführung auf Ebene der *Gesamtskala* bzw. der beiden *Subskalen* als eher gering (Varianzanteil  $s: l=0$  bis 2%), während der Varianzanteil ( $s: l$ ) *auf Itemebene* für einzelne Klassenführungsmerkmale teilweise deutlich höher ausfällt (z. B. Allgegenwärtigkeit der Lehrperson). Die Annahme, dass sich die zeitliche Stabilität in Abhängigkeit davon unterscheidet, ob es sich bei dem untersuchten Merkmal um eine prozess- bzw. strukturorientierte Maßnahme der Lehrperson handelt, kann jedoch nur teilweise bestätigt werden. Im Fall des Merkmals Allgegenwärtigkeit handelt es sich um ein prozessorientiertes Merkmal, womit von einer zeitlich variierenden Maßnahme ausgegangen werden kann (vgl. Doyle 1986), die eher das aktuelle Unterrichtsgeschehen reguliert und von der Lehrperson adaptiv innerhalb einer Doppelstunde eingesetzt wird. Bereits Kiel et al. (2013) weisen darauf hin, dass das Klassenführungsverhalten von Lehrpersonen situationsgerecht variieren kann bzw. an aktuelle Lehr-Lern-Situationen angepasst werden sollte (Beck et al. 2008; Neuenschwander 2006). Auch für das Merkmal „Regeln für das soziale Miteinander“ zwischen den Schüler/innen entfällt ein größerer Varianzanteil auf den Effekt  $s: l$  (12%). Dieser bedeutet inhaltlich, dass sich die Ausprägungen in Bezug auf die Einhaltung der Regeln des sozialen Miteinanders im Unterricht zwischen den beiden Stunden einer Doppelstunde unterscheiden. Die Rater/innen schätzten das Merkmal in beiden Schuljahren für die 2. Stunde der Doppelstunde niedriger ein (vgl. Tab. 2). Gabriel-Busse und Lipowsky (eingereicht) diskutieren dieses Ergebnis vor dem Hintergrund der unterschiedlichen Unterrichtsverläufe bzw. Inszenierungsformen der jeweiligen Doppelstunde. Diese führten dazu, dass in beiden Videostudien (1. und 2. SJ) in der 2. Stunde eher schülerorientierter Unterricht in Form von Übungen (zu zweit oder in der Gruppe) stattgefunden hat. Während in dem eher lehrerorientierten Unterricht im ersten Teil der Doppelstunde kaum Möglichkeiten des sozialen Miteinanders zu beobachten waren bzw. das Ausbleiben negativen Schüler/innenverhaltens positiv von den Rater/innen gewertet wurde (vgl. Gabriel und Lipowsky 2013), sind soziale Interaktionen und soziale Beziehungen, die maßgeblich durch im Vorfeld etablierte Regeln reguliert sind, erst im zweiten Teil der Doppelstunde z. B. während des Stationenlernens in Mathematik sichtbar bzw. nehmen Einfluss auf die Umgangsformen bzw. die soziale Kommunikation (vgl. Neuenschwander 2005). Für die übrigen zwei strukturorientierten Maßnahmen der Klassenführung bestätigt sich die Annahme, dass strukturorientierte Maßnahmen kaum kurzfristigen Schwankungen unterworfen sind (Varianzanteil  $s: l$  liegt bei 1 und 6%). Für weitere Beobachtungsstudien bedeutet dieses Ergebnis, dass je nach Merkmal im Vorfeld überlegt werden sollte, ob Unterschiede in der Ausprägung in Abhängigkeit von verschiedenen Phasen des Unterrichts erwartbar sein können. Weitere Studien sollten sich hier anschließen.

## 5.2 Die zeitliche Variation prozess- und strukturorientierter Maßnahmen der Klassenführung innerhalb eines Schuljahres (zwischen 12 Monate auseinanderliegenden Unterrichtseinheiten)

In Bezug auf die zeitliche Variabilität der Beobachterratings zur Klassenführung *im Abstand von 12 Monaten bzw. innerhalb eines Schuljahres* zeigt sich, dass ein beträchtlicher Varianzanteil auf den Interaktionseffekt Lehrperson  $\times$  Schuljahr ( $l \times j$ ) entfällt. Dieser liegt *auf Skalenebene* bei 38 % und variiert *auf Itemebene* zwischen 21 und 50 %. Das bedeutet, dass in Abhängigkeit vom untersuchten Klassenführungsmerkmal ein großer Varianzanteil auf Unterschiede zwischen den 12 Monaten auseinanderliegenden Unterrichtseinheiten einer Lehrperson entfällt. Sowohl prozess- als auch strukturorientierte Maßnahmen zeigen hier hohe Effektstärken, wobei der Varianzanteil für die prozessorientierten Maßnahmen der Klassenführung (Subskala) höher ist als für die strukturorientierten Maßnahmen. Die hohe Variabilität innerhalb eines Schuljahres kann dadurch erklärt werden, dass es sich bei der vorliegenden Studie um eine Untersuchung im Anfangsunterricht der Grundschule handelt. Im Vergleich zur Sekundarstufe geht es im Primarbereich darum, Regeln (z. B. auch zum sozialen Miteinander oder der Kommunikation) und Routinen zu etablieren. Vor allem in den ersten Monaten der Schulzeit ist es die Aufgabe der Lehrperson, die technischen Regeln und Verfahrensweisen wie beispielsweise das Ablegen der Hefte einzuführen, diese immer wieder zu erläutern und zu erklären und mit den Schülern/innen kontinuierlich einzuüben (Boostrom 1991). Gleiches gilt für die Übergänge zwischen einzelnen Unterrichtsaktivitäten, die auch strukturiert sein bzw. bestimmten Regeln und Ritualen folgen sollten, damit langfristig wenig Zeit verschwendet wird bzw. Unterrichtsstörung aufgrund von Off-Task-Verhalten minimiert werden (Arlin 1979). Aufgrund der hier berichteten Analyseergebnisse muss die Annahme zeitlicher Stabilität für den Anfangsunterricht bzw. für zwei im Abstand von 12 Monaten durchgeführte Beobachtungen zumindest für einzelne Klassenführungsmerkmale verworfen werden, weil aufgrund eines Anpassungsprozesses der beobachteten Lehrpersonen an die Bedürfnisse der Schüler/innen Unterschiede im beobachteten Klassenführungsverhalten zu erwarten sind. Der Anfangsunterricht befindet sich im Verlauf der Schuleingangsphase noch in einem Entwicklungsprozess (vgl. Eckerth et al. 2012). Im Zusammenhang mit der zeitlichen Variation der hier untersuchten Klassenführungsmerkmale muss an dieser Stelle jedoch berücksichtigt werden, dass die beiden untersuchten Doppelstunden (beobachtet im Abstand von 12 Monaten) teilweise nicht so vergleichbar waren (z. B. in Bezug auf Inhalt, Anteil Sozialformen etc.), wie dies für die vorliegende Fragestellung wünschenswert gewesen wäre (s. unten).

Als Konsequenz für weitere Beobachtungsstudien kann geschlussfolgert werden, dass es für den Anfangsunterricht von Bedeutung ist, wann die Merkmale der Klassenführung erfasst werden. Strukturorientierte Maßnahmen variieren im ersten Schuljahr auf Grund der Besonderheit des Anfangsunterrichts. Im zweiten Schuljahr sollten Regeln und Routinen z. B. für ein effektives Übergangsmanagement jedoch etabliert sein, mit entsprechenden Konsequenzen für die Messung der zeitlichen Stabilität. Zudem kann aus den hier berichteten Ergebnissen geschlussfolgert werden, dass Beobachtungen in Bezug auf die Klassenführung im Anfangsunterricht mehr-

mals im Jahr stattfinden sollten. Es bleibt offen, was ein „optimaler“ Zeitrahmen wäre, um Klassenführung im Anfangsunterricht mit einer hohen Zuverlässigkeit zu erfassen. Wenn die Klassenführung im Anfangsunterricht innerhalb eines Schuljahres so stark variiert, stellt sich zudem die Frage, inwieweit durch ein einmalig erfasstes Merkmal („Momentaufnahme“) Schülerleistungen erklärt werden können, die innerhalb eines Schuljahres zustande kommen.

In der vorliegenden Studie entfiel zusätzlich ein nicht unerheblicher Varianzanteil auf die jeweiligen Residuen. In weiteren G-Studien sollten demnach zusätzliche potenzielle Fehlerquellen (z. B. Rater/innen, vgl. Gabriel et al. 2015; Gabriel-Busse und Lipowsky 2020) berücksichtigt werden. Die zusätzliche Aufnahme weiterer Varianzquellen führte in dieser Studie dazu, dass negative Varianzkomponenten auftraten, die angesichts der geringen Stichprobengröße für eine zu hohe Modellkomplexität sprechen. Für die Beantwortung der hier zugrunde liegenden Forschungsfrage spielte der Einfluss der Rater/innen jedoch keine Rolle, sodass diese Varianzquelle im Rahmen der G-Studien bewusst unberücksichtigt blieb.

## 6 Limitationen der Studie

In dieser Studie wurden G-Koeffizienten ermittelt, die im Sinne eines Reliabilitätskoeffizienten als niedrig bis moderat gelten können. Hier können neben der starken zeitlichen Variabilität innerhalb von 12 Monaten zwei weitere mögliche Gründe angeführt werden: 1) Eine eher homogene Stichprobe, die zu einer geringen Variabilität in der Klassenführung zwischen verschiedenen Lehrpersonen geführt haben könnte (der Anteil stabiler Varianz liegt lediglich zwischen 5 % und 20 %, je nach Klassenführungsmerkmal) und 2) ein durchweg hoher Anteil an Residualvarianz, der auf unbeobachtete Einflussfaktoren der Klassenführung hindeutet (z. B. Ratereffekte, Unterrichtsgestaltung). Einschränkend muss für die vorliegende Studie angemerkt werden, dass die Unterrichtsstunden nicht zufällig ausgewählt wurden, sondern dass für jede der 21 Lehrpersonen jeweils eine Doppelstunde im 1. Schuljahr (Fach Deutsch) und 2. Schuljahr (Fach Mathematik) videografiert wurde. Da für beide Videostudien das Thema vorgegeben wurde (s. oben), bauen die videografierten Einzelstunden der jeweiligen Doppelstunde größtenteils inhaltlich und methodisch aufeinander auf. Demnach liegt keine stochastische Unabhängigkeit zwischen den Einzelstunden vor. Hier stellt sich die Frage, inwieweit für Einzelstunden mit unterschiedlichen Fächern (Themen), die in einem größeren zeitlichen Abstand beobachtet werden, eine größere Variabilität der Klassenführung grundsätzlich erwartet werden kann.

Für weitere Analysen wäre es sinnvoll, anstatt als Analyseeinheiten jeweils 45 Minuten einer Doppelstunde zu verwenden, die Doppelstunde in mehrere Segmente (ähnlich Mashburn et al., 2014 bzw. Jentsch et al. 2019) zu unterteilen und die Klassenführung für verschiedene Oberflächenmerkmale (Lehrmethoden sowie konkrete Organisations- und Sozialformen) des Unterrichts zu erfassen (vgl. Gabriel-Busse und Lipowsky 2020). Im Unterschied zur Sekundarstufe zeichnet sich der Unterricht in der Grundschule durch eine höhere Methodenvielfalt bzw. durch eine Verbindung von eher lehrergelenkten mit offenen, selbstgesteuerten

Unterrichtsphasen (Gabriel-Busse und Lipowsky 2020) aus. Es kann angenommen werden, dass in Phasen der Stationen-, Projekt- oder Gruppenarbeit ein wesentlicher Teil der Klassenführung von der Lehrperson auf die Schüler/innen übertragen wird (Bohl und Kucharz 2010). Übergänge zwischen einzelnen Stationen würden dann beispielsweise von den Schüler/innen selbst initiiert werden und müssten im Vorfeld geregelt werden. Im Vergleich zu traditionellem Unterricht, der weniger anspruchsvolle Lehrmethoden einsetzt, müssen für Formen des selbstständigen oder kooperativen Lernens außerdem deutlich mehr Regeln eingeführt werden bzw. es wird ein flexibleres Regelsystem benötigt (zusf. Doyle 1986). Allgemein könnten durch die Erfassung von Zusammenhängen mit verschiedenen Oberflächenmerkmalen zusätzliche Varianzanteile von Beobachterratings zur Klassenführung erklärt werden (Jentsch et al. 2019; Patrick und Montzicopoulos 2016).

Aufgrund der Konfundierung von Schuljahr (1. vs. 2. Schuljahr) und Fach (Deutsch vs. Mathematik) können keine allgemeinen Schlussfolgerungen für die ersten beiden Schuljahre im Anfangsunterricht gezogen werden, ohne das jeweilige Fach zu berücksichtigen. Obwohl in der vorliegenden Studie das Sample so gewählt wurde, dass die Lehrpersonen beide Fächer gleichermaßen unterrichten, können Unterschiede in der Klassenführung auch vom fachlichen Lerninhalt abhängen. Wie Praetorius et al. (2016) ausführen, hat jedes Fach seine eigene Kultur, weswegen auch die Lehrkräftebildung in Fachdidaktiken organisiert ist. Zudem beeinflusst der jeweilige fachspezifische Inhalt die Wahl der Methoden (Prange 2011) und damit auch die Klassenführung. Aus diesem Grund sollte in weiteren Studien überprüft werden, zu welchen Anteilen die zeitstabile Varianz in den Klassenführungsmerkmalen auf das Schuljahr oder den fachlichen Kontext zurückzuführen ist. Studien für die Sekundarstufe zeigen, dass die Klassenführung, die eine Lehrperson in verschiedenen Fächern zeigt, ähnlich hoch ausgeprägt ist (Praetorius et al. 2016). Lehrpersonen scheinen demnach einmal ausgebildete Routinen in Bezug auf die Klassenführung über verschiedene fachliche Kontexte anzuwenden. Für die Grundschule stehen entsprechende Analysen aus.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

## Literatur

- Arlin, M. (1979). Teacher transition can disrupt time flow in classroom. *American Educational Research Journal*, 16(1), 42–56.
- Beck, E., Baer, M., Guldemann, T., Bischoff, S., Brühwiler, C., Müller, P., Niedermann, R., Rogalla, M., & Vogt, F. (2008). *Adaptive Lehrerkompetenz. Analyse und Struktur, Veränderbarkeit und Wirkung handlungssteuernden Lehrerwissens*. 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.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- 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.
- Casabianca, J. M., Lockwood, J. R., & McCaffrey, D. F. (2015). Trends in classroom observation scores. *Educational and Psychological Measurement*, 75(2), 311–337.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159.
- Cronbach, L. J., Gleser, G., Nanda, H., & Rajaratman, N. (1972). *The dependability of behavioral measurements: theory of generalizability for scores and profiles*. New York: Wiley.
- Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Hrsg.), *Handbook of research on teaching* (S. 392–431). New York: Macmillan.
- 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.
- Gabriel, K. (2014). *Videobasierte Erfassung von Unterrichtsqualität im Anfangsunterricht der Grundschule – Klassenführung und Unterrichtsklima in Deutsch und Mathematik*. Dissertation, Disputation am 19.07.2013. Kassel: University Press.
- Gabriel, K., & Lipowsky, F. (2013). Hoch inferentes Rating: Klassenführung in drei Fächern. 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., Praetorius, A.-K., & Lipowsky, F. (2015). Wie in der einen, so auch in der anderen Stunde? – Analysen zur Stabilität der Klassenführung im Anfangsunterricht. *Jahrbuch für Allgemeine Didaktik. Themenheft: Klassenmanagement/Klassenführung. Perspektiven. Befunde. Kontroversen*, 1, 85–95.
- Gabriel-Busse, K., & Lipowsky, F. (2020). 90 min Mathematikunterricht bei gleichbleibender Unterrichtsqualität? Analysen zur zeitlichen Stabilität der Basisdimensionen der Unterrichtsqualität im 2. Schuljahr. *Unterrichtswissenschaft*, 49, 137–163. <https://doi.org/10.1007/s42010-020-00086-4>.
- Gabriel-Busse, K., & Lipowsky, F. (eingereicht). Reicht eine Doppelstunde im Anfangsunterricht aus? – Ergebnisse aus den PERLE-Videostudien Deutsch und Mathematik zur zeitlichen Stabilität und Generalisierbarkeit von Beobachterratings zur Klassenführung. In C. Jacobi-Theurer, M. Hess, A.-K. Denn & F. Lipowsky (Hrsg.), *Determinanten und Effekte der Persönlichkeits- und Lernentwicklung in der Grundschule – Ergebnisse der PERLE-Studie*. Münster: Waxmann.
- 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. (2009). Nina und Michael. Miró und ein Nussknacker! Persönlichkeits- und Lernentwicklung von Grundschulkindern. *Die Grundschulzeitschrift*, 23, 18–21.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen. Schüler als Quellen der Unterrichtsbeschreibung*. Münster: Waxmann.
- 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., & Weinert, F. E. (1997). Unterrichtsqualität und Leistungsentwicklung: Ergebnisse aus dem SCHOLASTIK-Projekt. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 241–251). Weinheim: Beltz.
- Hörter, P., Gippert, C., Holodnyński, M., & Stein, M. (2020). Klassenführung und Fachdidaktik im (Anfangs-)Unterricht Mathematik erfolgreich integrieren – Konzeption einer videobasierten Lehrveranstaltung zur Förderung der professionellen Unterrichtswahrnehmung. *Herausforderung Lehrer\_innenbildung – Zeitschrift zur Konzeption, Gestaltung und Diskussion (HLZ)*, 3(1), 256–282.

- Jentsch, A., Casale, G., Schlesinger, L., Kaiser, G., König, J., & Blömeke, S. (2019). Variabilität und Generalisierbarkeit von Ratings zur Qualität von Mathematikunterricht zwischen und innerhalb von Unterrichtsstunden. *Unterrichtswissenschaft*, 48, 179–197.
- Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: combining high-quality observations with student surveys and achievement gains*. MET Project Research Paper. Seattle: Bill & Melinda Gates Foundation.
- Kiel, E., Frey, A., & Weiß, S. (2013). *Trainingsbuch Klassenführung*. Bad Heilbrunn: Klinkhardt.
- Klieme, E. (2019). Unterrichtsqualität. In M. Harring, C. Rohlf, Carsten, M. Gläser-Zikuda (Hrsg.), *Handbuch Schulpädagogik* (S. 393–408). Münster: Waxmann.
- 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.
- König, J., & Lebens, M. (2012). Classroom Management Expertise (CME) von Lehrkräften messen: Überlegungen zur Testung mithilfe von Videovignetten und erste empirische Befunde. *Lehrerbildung auf dem Prüfstand*, 5(1), 3–28.
- Kounin, J.S. (2006). *Techniken der Klassenführung* (2. Aufl.). Münster: Waxmann.
- Lipowsky, F. (2007). Was wissen wir über guten Unterricht? *Friedrich-Jahresheft*, 25, 26–30.
- 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.
- Lotz, M., & Corvacho del Toro, I. (2013). Die Videostudie im Fach Deutsch: „Lucy rettet Mama Kroko“. 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., Berner, N.E., Gabriel, K., Post, S., Faust, G., & Lipowsky, F. (2011). Unterrichtsbeobachtung im Projekt PERLE. In D. Kucharz, T. Irion & B. Reinthaler (Hrsg.), *Grundlegende Bildung ohne Brüche* (S. 183–194). Wiesbaden: VS.
- Mashburn, A.J., Meyer, J.P., Allen, J.P., & Pianta, R.C. (2014). The effect of observation length and presentation order on the reliability and validity of an observational measure of teaching quality. *Educational and Psychological Measurement*, 74(3), 400–422.
- Meyer, J.P., Cash, A.H., & Mashburn, A. (2011). Occasions and the reliability of classroom observations: alternative conceptualizations and methods of analysis. *Educational Assessment*, 16, 227–243.
- Mösko, E., Gabriel, K., & Lipowsky, F. (2013). Die Videostudie im Fach Mathematik: „Einführung in die Multiplikation“. In M. Lotz, F. Lipowsky & G. Faust (Hrsg.), *Technischer Bericht zu den PERLE-Videostudien*. Materialien zur Bildungsforschung, (Bd. 23/3, S. 45–49). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF).
- 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.
- 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.
- Patrick, H., & Montzicopoulos, P. (2016). Is effective teaching stable? *The Journal of Experimental Education*, 84(1), 23–47.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system. Manual*. Baltimore: Brookes.
- Pietsch, M., & Tosana, S. (2008). Beurteilereffekte bei der Messung von Unterrichtsqualität. Das Multifacetten-Rasch-Modell und die Generalisierbarkeitstheorie als Methoden der Qualitätssicherung in der externen Evaluation von Schulen. *Zeitschrift für Erziehungswissenschaft*, 11(3), 430–452.
- Praetorius, A.-K. (2014). *Messung von Unterrichtsqualität durch Ratings*. Münster: Waxmann.
- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *ZDM Mathematics Education*, 50, 407–426.
- Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2–12.
- Praetorius, A.-K., Vieluf, S., Saß, S., Bernholt, A., & Klieme, E. (2016). The same in German as in English? Investigating the subject-specificity of teaching quality. *Zeitschrift für Erziehungswissenschaft*, 19(1), 191–209.

- Prange, K. (2011). Didaktik und Methodik. In J. Kade, W. Helsper, C. Lüders, B. Egloff, F.-O. Radtke & W. Thole (Hrsg.), *Pädagogisches Wissen. Erziehungswissenschaft in Grundbegriffen* (S. 183–190). Stuttgart: Kohlhammer.
- Seidel, T. (2020). Klassenführung. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 119–131). Heidelberg: Springer.
- 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, 454–499.
- Shavelson, R., & Dempsey-Atwood, N. (1976). Generalizability of measures of teaching behavior. *Review of Educational Research*, 46, 553–611.
- Shavelson, R., & Webb, N. (1991). *Generalizability theory: A primer*. Thousand Oaks: SAGE.
- Thal, J., & Ebert, U. (2004). *Methodenvielfalt im Unterricht. Mit Lust stressarm und effektiv lernen*. München: Luchterhand.