Göllner, Richard; Fauth, Benjamin; Lenske, Gerlinde; Praetorius, Anna-Katharina; Wagner, Wolfgang

Do student ratings of classroom management tell us more about teachers or about classroom composition?

Praetorius, Anna-Katharina [Hrsg.]; Grünkorn, Juliane [Hrsg.]; Klieme, Eckhard [Hrsg.]: Empirische Forschung zu Unterrichtsqualität. Theoretische Grundfragen und quantitative Modellierungen. 1. Auflage. Weinheim; Basel : Beltz Juventa 2020, S. 156-172. - (Zeitschrift für Pädagogik, Beiheft; 66)

66. Beiheft April 2020

ZEITSCHRIFT FÜR PÄDAGOGIK

Empirische Forschung zu Unterrichtsqualität. Theoretische Grundfragen und quantitative Modellierungen

BELTZ JUVENTA

Zeitschrift für Pädagogik · 66. Beiheft

Zeitschrift für Pädagogik · 66. Beiheft

Empirische Forschung zu Unterrichtsqualität

Theoretische Grundfragen und quantitative Modellierungen

Herausgegeben von
Anna-Katharina Praetorius, Juliane Grünkorn
und Eckhard Klieme

BELTZ JUVENTA

ISSN: 0514-2717
ISBN 978-3-7799-3534-6 Print
ISBN 978-3-7799-3535-3 E-Book (PDF)
Bestellnummer: 443534

1. Auflage 2020

in der Verlagsgruppe Beltz · Weinheim Basel
Werderstraße 10, 69469 Weinheim

Herstellung: Hannelore Molitor
Satz: text plus form, Dresden
Druck und Bindung: Beltz Grafische Betriebe, Bad Langensalza
Printed in Germany

Weitere Informationen zu unseren Autoren und Titeln finden Sie unter: www.beltz.de

Inhaltsverzeichnis

Themenblock V: Modellierung der Wirkungen von Unterrichtsqualität (englischsprachig)

Kommentare

Richard Göllner/Benjamin Fauth/Gerlinde Lenske/Anna-Katharina Praetorius/ Wolfgang Wagner

Do Student Ratings of Classroom Management Tell us More About Teachers or About Classroom Composition?

Abstract: The present study investigated whether the varying referents (i. e., teacher or student referent) of student ratings of established classroom management measures differ in their associations with compositional classroom characteristics and students' math achievement. Re-analysis of a large-scale dataset (PISA 2003) showed that classrooms with a higher proportion of male students, as well as those with lower math performance, exhibited lower scores on classroom management factors referring more to students than the teacher. These were in turn related to lower pre-adjusted math achievement of students. There were no associations with a measure referring to the teacher. Our results indicate that varying referents tap into different aspects of the classroom management process.

Keywords: Classroom Management, Student Ratings, Indicator References, Classroom Composition, Math Achievement

1. Introduction

In the educational literature focused on effective teaching, no other aspect of teaching quality receives as much attention as classroom management. Classroom management is a central element in most theories of good teaching, and has been shown to be a consistent predictor of students' learning and development (e. g., Creemers, Kyrikides, & Antoniou, 2013; Hamre & Pianta, 2010; Hattie, 2009).

Oftentimes, student ratings are used to assess classroom management, with students asked about their teachers' ability to provide clear and consistent behavioral expectations, monitor the classroom for potential problems, and spend a minimal amount of time on behavior management issues (e. g., Aldrup, Klusmann, Lüdtke, Göllner, & Trautwein, 2018; Kunter & Baumert, 2006; Wagner et al., 2016). Other indicators of high-quality classroom management that are used in empirical studies are the number of disruptions by students, and the extent to which classroom time is used for learning-related purposes.

All of these constructs of the classroom management domain address quite different aspects of classroom management. Some of the constructs, for instance, address teachers' managing actions (e. g., monitoring, or the establishment of rules and classroom procedures), whereas other constructs seem to tap more into the consequences of teachers' classroom management (e. g., number of disruptions). At the same time, frequently

used constructs differ as to which referent of actions they address. Whereas some constructs of the classroom management domain focus directly on teacher behavior (and as such may be assessed solely by items with a teacher referent), others are more related to students in classrooms, or to the interplay between the teacher and students (see also Fauth, Göllner, Lenske, Praetorius & Wagner, in this issue).

In the present article, we examine whether such differences in referent (i. e., teacher vs. students) have an impact on the assessment of classroom management. We argue that shifting the reference from the teacher to the students makes classroom management measures (more strongly) dependent on classroom student composition. Thus, we measure not only the quality of a certain teacher's classroom management, but also the characteristics of classroom students being taught by this teacher.

We begin by summarizing past research on teachers' classroom management, as well as findings using student ratings of classroom management. We then present the results of an empirical study investigating three sub-dimensions of classroom management (i. e., monitoring, the absence of disturbances, and effective time use), examining to what extent measures with varying referents are associated with student characteristics at the classroom level. On the basis of the assumption that there is a higher reliance on measures referring to students in compositional classroom characteristics, finally we investigate the associations between measures of classroom management and students' math achievement.

2. Conceptualizing Classroom Management

Several aspects of teaching quality are currently seen as important for students' learning. Existing conceptions contain numerous factors that are relevant in describing the complex nature of teaching quality (Creemers et al., 2013; Hamre & Pianta, 2010; Helmke, 2010; Klieme, Pauli & Reusser, 2009; Kunter & Baumert, 2006). Classroom management is seen as a central element of good teaching, and has an important place in many conceptualizations of teaching quality. Related measures, such as a lack of student misbehavior and effective management of time and classroom routines, have been found to be consistently related to students' learning outcomes (e. g., Aldrup et al., 2018; Kunter et al., 2013; Wagner et al., 2016).

Various conceptualizations and measurement instruments cover a variety of aspects of the construct of classroom management. That is, classroom management is in modern conceptualizations seen as a hierarchical structure, consisting of a broad quality domain encompassing various key aspects, such as the absence of disturbances, effective time use, the existence of classroom rules, and classroom monitoring (e. g., Hamre & Pianta, 2010). A majority of studies have used student ratings to assess teachers' classroom management (e. g., Aldrup et al., 2018; Kunter & Baumert, 2006; Wagner et al., 2016). A quick glance at existing measures, however, shows that student ratings differ not only with respect to the specific quality aspect (i. e., monitoring, time use, disturbance prevention) but also with respect to the reference object. In some instances, operationaliza-

tions clearly refer to the teacher (e. g., "Our math teacher always knows exactly what is happening in class"; Baumert, Gruehn, Heyn, Köller & Schnabel, 1997, p. 85), whereas other measures are vaguer ("Our teacher has to wait a long time before it gets quiet"; Bos, Gröhlich, Dudas, Guill & Scharenberg, 2010, p. 163) or refer more to students and the interplay between the teacher and students ("In math class, the lesson is often disrupted"; Ramm et al., 2006, p. 191).

The use of different referents usually goes along with changes in the construct to be assessed. Indeed, current assessments of teaching quality, and interpretations of study findings, typically take a teacher-oriented perspective. The main question they address is whether teachers are equipped with the abilities and skills they need to manage the classroom in a productive and effective way. However, the organization and management of students' behavior, time, and attention in the classroom can also be seen from an ecological perspective. According to Doyle (2013), classrooms are quite complex systems of individuals. Thus, the assumption that students' behavior in the classroom is almost completely dependent on their teacher's classroom management ability is an oversimplification. Rather, teachers and students both contribute to classroom interactions, and thus, teachers and students are jointly enacted in the classroom management process (Doyle, 2013). It can be argued that this process finds expression the more that indicators refer to students rather than teachers. In addition, one further reason for varying referents in the classroom management indicators used may lie in the distinction between teacher's classroom management actions and their successful realization during a regular class. Whereas indicators with an explicit reference to the teacher focus more on the teacher's management actions aimed to achieve effective classroom management (e. g., monitoring, structure, or rule setting), indicators referring to the students (e. g., disturbances or time use) provide more information as to whether this objective is actually achieved in a classroom. Existing indicators may reflect operationalizations from different theoretical perspectives in a spectrum ranging from the strong focus on behavioral operations on the part of teachers (e. g., Landrum & Kauffman, 2006) to the achievement of an effective classroom management in a class comprising students with specific needs and learning requirements (e. g., Gettinger & Kohler, 2006).

3. Classroom Management and Classroom Student Composition

The idea that classroom management measures vary in the extent to which they refer more to the teacher or to the students in a classroom, raises the question of what classroom characteristics, in terms of student composition, are included in these measures. In general, classroom composition characteristics such as students' SES or academic performance are prominent candidates, and have been shown to be relevant predictors of students' learning over and above their individual learning backgrounds. Consequently, a more favorable classroom composition with higher student SES and higher performance should lead to higher learning achievement because effective teaching is easier to implement and students are equipped with academically-oriented social networks that

facilitate the spread of academic norms and higher academic aspirations (e. g., Harker & Tymms, 2004).

Guided by this perspective, and given the fact that classroom management is revealed to be a consistent predictor of student learning, measures of classroom management that refer to students might represent an important means of understanding the effects of class composition. In other words, using classroom management measures that refer to students makes it possible to test that the repeatedly reported associations between classroom management and students' learning outcomes are at least partly due to classroom composition. In fact, previous research has indicated that ratings of classroom management referring to students make measures of classroom management more dependent on class composition. For instance, a study by Praetorius, Vieluf, Saß, Bernholt, and Klieme (2016) using a measure that mainly refers to discipline problems, showed that student ratings were highly consistent across two subjects (German, and English as a foreign language). This was also the case if subjects were taught by two different teachers. Furthermore, research on observational data has shown that school and classroom composition with respect to achievement, gender, and migration background, is systematically related to teaching quality ratings (Campbell & Ronfeldt, 2018). However, it remains open whether these findings might have arisen from the use of measurement indicators that refer more to students rather than the teacher.

4. The Present Study

In this study, we used data on student ratings of classroom management, including measures with different referents, and examined whether these exhibited different associations with classroom composition and students' math achievement. We addressed two research questions: First, we investigated whether student ratings of multiple aspects of classroom management (monitoring, the absence of disturbances, and effective time use) were associated with class composition in terms of math performance, gender, and socioeconomic background. We expected that the associations with compositional characteristics would be more pronounced for ratings of classroom management referring to students or the interplay between teacher and students, than for ratings that clearly refer to the teacher. Second, we tested whether the measures differed in their prediction of students' math achievement. Assuming that classroom management measures that refer more to the students are associated with class compositional characteristics, which in turn could be shown to be related to students' learning, we hypothesized that the measures referring to students would be more predictive of students' pre-adjusted math achievement.

5. Method

5.1. Sample

We used data from the German extension of the 2003 cycle of the Programme for International Student Assessment (PISA; Organisation for Economic Co-operation and Development 2004).[1] In this national extension, a subsample of 15-year-old PISA students in Grade 9 and their teachers, took part in an additional longitudinal study, with reassessment in Grade 10. Student assessment took place in the second half of the year in Grade 9, and one year later in Grade 10. Participation in the study was voluntary. Students in PISA classes were administered achievement tests as well as questionnaires concerning background data and aspects of teaching quality in math lessons. We used the sample of N = 4,645 students from intermediate and academic track schools (K = 259 classes).[2] On average, 12 students per class provided data on their math teachers' teaching quality. Due to time constraints, PISA used a matrix design, where half of the students in each class were chosen to complete one set of items while the remaining students answered a different set of items. In the present study, teaching quality perception measures from N = 2,508 students were able to be used.

5.2 Instruments

Classroom management. We used three well-known measures of classroom management: teachers' monitoring activity, the absence of disturbances, and efficient time use. All measures were drawn from the Grade 10 measurement point. The indicators for monitoring refer to the teacher, while the indicators for disturbances and ineffective time use refer more to the students (see Table 1 for a complete set of items). All responses were given on a 4-point Likert scale ranging from 1 (completely disagree) to 4 (completely agree). *Monitoring* assesses the extent to which the teacher keeps an eye on students' actions and is alert to any behavioral problems or learning difficulties. The construct was operationalized using four items (e. g., "The teacher always knows exactly what is going on in class"). Scale reliability was α = .71. *The absence of disturbances* was assessed with two items referring to difficulties in maintaining discipline in the classroom (e. g., "In math, the lesson is often disturbed"). The scale was re-

---

1 We thank the German PISA consortium (Prenzel et al., 2007; Prenzel et al., 2013) and the Research Data Centre (FDZ) at the IQB in Berlin for their approval and support in conducting the secondary analysis.

2 A "tripartite" system of lower track schools (Hauptschule), intermediate track schools (Realschule), and academic track schools (Gymnasium) is the most common system in German states; some states offer multitrack schools that serve lower and intermediate track students in joint classes. The present study sample consisted only of students from intermediate track schools and academic track schools, because lower track students finish school at the end of Grade 9.

verse-coded so that higher values indicated higher classroom management. Finally, *effective time use* captured the amount of time lost through disciplinary problems in class via two items (e. g., "In math, it is long after the lesson starts that students become quiet and start working"). This scale was also reverse-coded, so that higher values would indicate higher classroom management. Descriptive statistics of the classroom management indicators are given in Table 1. For all indicators, a substantial degree of variance could be attributed to the classroom level (ICC1) and assure a reliable assessment of classroom management at the classroom level (ICC2).

Students' mathematics achievement. To measure students' mathematics achievement, we used the standardized math achievement test scores that were applied in prior teaching quality research with the PISA dataset (see Kunter et al., 2013 for a detailed description). As we focused on the effect of classroom management on students' learning gains, we also included students' prior math performance at Grade 9, which was measured as part of the international PISA 2003 assessment. Students' mathematics achievement at the end of Grade 10 was assessed with a test covering standard content from the federal states' curricula for Grade 10 mathematics. All tests were scaled using Rasch analysis. Test items had a closed response format, and subsets of test items were administered us-

Quality dimension		*M*	*SD*	*ICC(1)*	*ICC(2)*
Monitoring (Cronbach's α = .73)					
Mo1	My teacher always knows exactly what is going on in the class	2.52	0.97	.26	.81
Mo2	My teacher always checks our homework very accurately	2.23	0.98	.29	.83
Mo3	My teacher makes sure that we pay attention	2.83	0.92	.23	.78
Mo4	My teacher immediately notices when students start doing something else	2.61	0.96	.20	.75
Absence of disturbances (Cronbach's α = .81)					
AD1	In math class, the lesson is often disturbed (recoded)	2.43	0.99	.32	.85
AD2	In math, a lot of nonsense is going on all the time (recoded)	2.63	1.02	.31	.84
Effective time use (Cronbach's α = .74)					
ET1	In math, it is long after the beginning of the hour by the time the students get quiet and start working (recoded)	2.62	1.00	.29	.83
ET2	In math, a lot of time in class is wasted (recoded)	2.66	1.01	.26	.80

Note. The scale for each item was 1 to 4.

Tab. 1: Summary of item indicators for students' ratings of classroom management

ing a multi-matrix design. Item and person parameters for students' math achievement were estimated, and the weighted likelihood estimates were used as person parameters for individuals' math achievement in Grades 9 and 10. Again, students' math achievement scores revealed a substantial amount of variance at the classroom level, both for the Grade 9 scores (ICC1 = .39, ICC2 = .89) and for the Grade 10 scores (ICC1 = .39, ICC2 = .89).

Students' background. In addition to student ratings of classroom management and math achievement in Grade 9, we used students' gender (0 = female, 1 = male) and socioeconomic background (SES) as further measures to assess their learning background. The social status of the students' families was operationalized by the International Socio-Economic Index, which was developed by Ganzeboom and Treiman (2003) on the basis of the International Labour Office's occupation classification system. The score from the parent with the highest index ranking was used in the analyses. At the classroom level, we also included the school track in which students were enrolled (non-academic or academic track). Students from the non-academic track served as reference.

5.3 Statistical Analysis

Preliminary analysis. Based on a two-level first-order factor model that contained correlated first-order factors corresponding to the dimensions of monitoring, effective time use, and absence of disturbances at both levels (within and between classes), we began by checking whether student ratings reflected multiple dimensions of classroom management. Even though a model with three factors exhibited a good model fit, $\chi^2(34) = 195.62$, $p < .001$; CFI = .97; TLI = .95; RMSEA = .04; $SRMR_{within} = .04$; $SRMR_{between} = .04$, scaling correction factor = 1.11, a nearly perfect correlation was revealed between absence of disturbances and effective time use, both at the within- ($r = .94$) and the between-classroom levels ($r = .97$). In contrast, correlations between the absence of disturbances and effective time use with monitoring were substantially smaller (within level: $.31 \leq r \leq .32$; between level: $.81 \leq r \leq .85$). For this reason, we tested a second model, combining the absence of disturbances and effective time use, resulting in a two-factor model. This model's fit was similarly good to the three-factor model, $\chi^2(38) = 212.45$, $p < .001$; CFI = .97; TLI = .95; RMSEA = .04; $SRMR_{within} = .04$; $SRMR_{between} = .04$, scaling correction factor = 1.13. Thus, in the interest of parsimony we decided to combine the two factors in all subsequent analyses (see Table 2).

Nested factor model. In order to address our research questions, we then proceeded to examine a nested factor model. A nested factor model (or bifactor model) assumes the existence of a general factor that directly influences all observed measurement indicators and one or more additional components that account for different specific subsets of indicators (Gustafsson & Åberg-Bengtsson, 2010). In order to deal efficiently with the different referents in the measurement indicators, we used all of the different indicators

	Student level				Classroom level			
	3 factors		2 factors		3 factors		2 factors	
Quality dimension	**Est.**	***SE***	**Est.**	***SE***	**Est.**	***SE***	**Est.**	***SE***
Monitoring								
Mo1	0.59	0.02	0.59	0.02	0.86	0.03	0.86	0.03
Mo2	0.43	0.02	0.43	0.02	0.49	0.06	0.49	0.06
Mo3	0.67	0.02	0.67	0.02	0.91	0.03	0.91	0.03
Mo4	0.68	0.02	0.68	0.02	0.96	0.02	0.96	0.02
Absence of disturbances								
AD1	0.74	0.02	0.73	0.02	0.99	0.01	0.99	0.01
AD2	0.76	0.02	0.75	0.02	1.00	0.01	1.00	0.01
Effective time use								
ET1	0.73	0.02	0.71	0.02	0.99	0.01	0.96	0.01
ET2	0.64	0.02	0.63	0.02	0.91	0.02	0.89	0.02

Note. [a] Factor loadings are shown in their standardized form.

Tab. 2: Factor loadings of first-order factor multilevel models with two and three factors

for monitoring to identify the general factor, thus controlling for differences in monitoring between teachers/classes and for specific references to the teacher in all of the remaining classroom indicators (see the measurement model in Fig. 1).

Here, the specific factors represent classroom management aspects that vary between students within classes (student level) and between classes (classroom level) with identical levels of perceived monitoring.

Classroom compositional effects on ratings of classroom management. As we were interested in whether classroom management measures were associated with classroom student composition, we further conducted multilevel regression analyses. The central question in the analysis of composition effects is whether an aggregated classroom characteristic is associated with an outcome measure after controlling for individual differences among students on that characteristic (Kreft, de Leeuw & Aiken, 1995). Based on the nested factor model, we included classroom composition characteristics as predictors of the general and the specific classroom management factors (see the covariate model in Figure 1), using an approach proposed by Koch, Holtmann, Bohn and Eid (2018). Specifically, we used their residual approach, in which either the general or the specific factors are partialed out from the covariates and are then used as an independent variable in an additional model to predict the general or specific factors. This procedure ensures that the implied model assumption of zero correlations between the general and

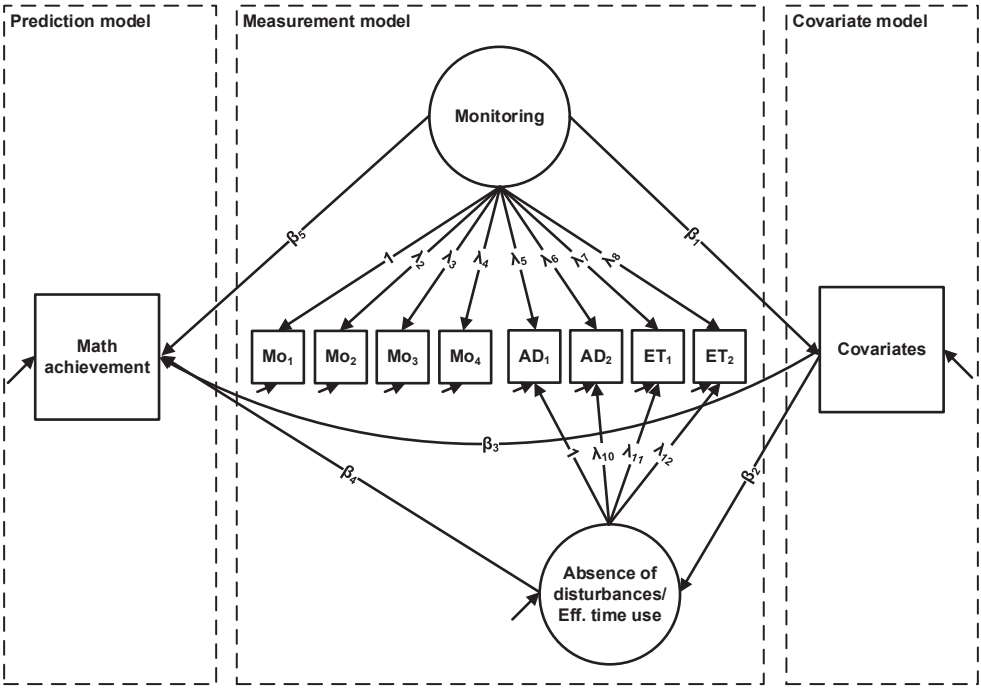

Fig. 1: Nested factor model representing a general (monitoring, Mo) and a specific (absence of disturbances, AD; effective time use, ET) classroom management factor. Covariates of the general and specific factors were included using an approach proposed by Koch et al. (2018). Models were estimated simultaneously at the student and classroom levels. Covariates (math performance, SES, and gender) at the classroom level were modeled by manifest aggregation. School track was used only at the classroom level.

specific factors is compatible with testing these same predictors on the general and the specific factors (see Koch et al., 2018 for more detail).

Effects of classroom management on student achievement. Finally, we extended the analytical model by including student math achievement at Grade 10 in the multilevel analysis (see the prediction model in Figure 1) and tested whether student ratings of classroom management were associated with students' math achievement at Grade 10 after controlling for students' prior math achievement at Grade 9.

One important prerequisite for testing such effects with latent measures is the cross-level invariance of measures. Otherwise, the comparison of level-specific associations may not be meaningful, as the factor variances of the same set of indicators at the two levels – and thus, also the corresponding regression coefficients – are not comparable. Consequently, we constrained the measurement models (i. e., factor loadings) to be equal across levels (see Wagner, Göllner, Helmke, Trautwein & Lüdtke, 2013). For nested factor models, this procedure is somewhat more complicated as the differ-

ences between factor intercorrelations at the different levels in a simple first-order factor model (i. e. the structural model) are represented by different loadings on the general factor (i. e. the measurement model) for indicators that also load on specific factors. In other words, some parts of the measurement model (i. e., loadings on the general factor from indicators that also loaded on specific factors) in the nested factor model reflected differences in the structural model. For this reason, we tested only for partial measurement invariance, by applying equality constraints to the monitoring indicators' loadings on the general factor, and effective time use and absence of disturbance loadings on the specific factors.

We conducted all analyses using the Mplus 7.3 software (Muthén & Muthén, 1998–2012). Robust maximum likelihood estimation (MLR) for continuous data was used to obtain reliable standard errors and fit tests for non-normally distributed data. In addition, we utilized full information maximum likelihood (FIML) estimation, where model variables are used to predict missing data. All continuous variables were z-standardized ($M = 0$, $SD = 1$) before analysis.

6. Results

6.1. Nested Factor Model

To address our research questions, we applied a nested factor model with one general classroom management factor and one specific factor, with factor loadings on absence of disturbances/effective time use (see Figure 1). All descriptive fit indices indicated a good model fit, $\chi^2(33) = 140.80$, $p < .001$; CFI = .98; TLI = .97; RMSEA = .03; $SRMR_{within} = .02$; $SRMR_{between} = .04$, scaling correction factor = 1.07. Constraining the factor loadings for monitoring (general factor) and for absence of disturbances/effective time use (specific factor) led to a highly comparable model fit, $\chi^2(39) = 144.23$, $p < .001$; CFI = .98; TLI = .97; RMSEA = .03; $SRMR_{within} = .02$; $SRMR_{between} = .04$; scaling correction factor = 1.11, indicating measurement equivalence across levels.

6.2 Compositional Classroom Effects

We then examined the association between class composition and classroom management factors (research question 1). Specifically, we included students' gender, math performance at Grade 9, and SES into the multilevel regression model and examined the associations both at the within- and the between- classroom levels, between students' background and their ratings of classroom management. In addition, we controlled for potential school track differences at the classroom level. The results are shown in Table 3 and can be summarized as follows: First, the results at the within level revealed no statistically significant associations between students' background and classroom management. Neither students' math performance, SES, nor gender was associated with class-

	Monitoring				Absence of disturbances/effective time use[c]			
	Student level		Classroom level		Student level		Classroom level	
Variables	**Est.**	***SE***	**Est.**	***SE***	**Est.**	***SE***	**Est.**	***SE***
Math performance	−0.02	0.03	−0.06	0.13	0.02	0.02	0.15	0.06*
SES	−0.03	0.04	−0.03	0.09	0.01	0.02	−0.06	0.08
Gender[a]	−0.04	0.02	−0.05	0.03	0.02	0.03	−0.36	0.18*
School track[b]			−0.13	0.09			0.01	0.08

Note. [a] Gender (male); [b] School track (Gymnasium). [c] Based on the modeling approach proposed by Koch et al. (2018), predictors and outcomes were residualized by monitoring. Regression coefficients are shown in their unstandardized form.
* $p < .05$

Tab. 3: Multilevel regression models predicting monitoring and disturbances with compositional classroom characteristics

room management factors. At the classroom level, two findings are noteworthy. First, all of the classroom student characteristics were unrelated to monitoring ($p > .05$; see Table 3). Second, the results for the specific factor of absence of disturbances/effective time use revealed two statistically significant findings. Classrooms with higher math performance at Grade 9 ($b = 0.15$, SE = 0.06, $p = .015$) and a higher proportion of female students ($b = −0.36$, $SE = 0.18$, $p = .047$) reported greater absence of disturbances and more effective time use. School track was unrelated to the two classroom management factors (monitoring: $b = −0.13$, $SE = 0.09$, $p = .155$; absence of disturbances/effective time use: $b = 0.01$, $SE = 0.08$, $p = .891$). In order to test whether these associations also held after controlling for individual differences among students (i. e., compositional effects), we compared the level-specific coefficients. The results showed that the association between absence of disturbances/effective time use (math performance: $b = 0.13$, $SE = 0.06$, $p = .047$; gender: $b = −0.38$, $SE = 0.18$, $p = .037$) remained statistically significant, even after accounting for individual differences among students.

6.3 Associations with Math Achievement

Finally, for our second research question we examined the association between classroom management factors and student achievement at Grade 10. To do this, we additionally regressed students' math achievement at Grade 10 on the classroom factors and controlled for students' gender, school track differences, SES, and math performance at Grade 9. The complete model including the classroom management factors, predictor variables and achievement outcomes elicited a good model fit, $\chi^2(95) = 265.69$, $p < .001$; CFI = .98; TLI = .97; RMSEA = .02; $SRMR_{within} = .02$; $SRMR_{between} = .05$, scaling cor-

rection factor = 1.01. The associations found between student background variables and classroom factors remained unchanged at both the student and classroom levels. In addition, monitoring and absence of disturbances/effective time use were not associated with students' achievement at the within-classroom level (monitoring: $b = 0.02$, $SE = 0.03$, $p = .436$; absence of disturbances/effective time use: $b = -0.01$, $SE = 0.03$, $p = .821$).

However, the prediction results at the classroom level revealed a significant association between absence of disturbances/effective time use and achievement at Grade 10 ($b = 0.20$, $SE = 0.08$, $p = .008$). Monitoring, on the contrary, was not associated with math achievement ($b = 0.02$, $SE = 0.04$, $p = .643$). The results for absence of disturbances/effective time use also held after controlling for individual differences among students ($b = 0.21$, $SE = 0.08$, $p = .014$).

In sum, these findings demonstrate that classroom management was associated with students' achievement, although only those measures referring to the students showed statistically significant results.

7. Discussion

In the present study, we examined student ratings of classroom management and potential differences between different measures. We tested whether classroom management measures, according to the extent to which they referred more to the teacher or to the students in a classroom, differed in their associations with classroom student composition and students' achievement in mathematics.

Our results show that the way students are specifically asked about classroom management is of direct relevance to which aspect of the classroom management process is being addressed. Classroom management indicators that do not directly address the teacher as the referent of actions, yield information about students as an equally important part of the classroom management process. Specifically, the results revealed associations between classroom student composition and students' ratings of classroom management, but only for measures that referred to the students or to the interplay between teacher and students (absence of disturbances and effective time use). The same pattern of results was also found for the prediction of students' pretest-adjusted math achievement at Grade 10. Whereas the absence of disturbances and effective time use were associated with students' pretest-adjusted math achievement, teachers' monitoring exhibited a non-significant effect.

In sum, our results add to recent literature showing that classroom management indicators referring more to students in the class, correspondingly capture information about the composition of the class, and thus, tap into a conceptually different aspect of the classroom management process (Fauth et al., see in this special issue).

7.1. Classroom Management and Classroom Student Composition

Classroom management is known to be one of the most consistent predictors of students' learning, and is frequently assessed via student ratings. In this study, we compared three measures of classroom management that differed not only in the aspect of classroom management addressed, but also in the extent to which they referred more to the teacher or to the students in a given class. Consistently with prior research, the results of factor analysis showed that students are able to distinguish different aspects of classroom management. However, students' perceptions were much less differentiated when measures referred to students. In fact, the association between the absence of disturbances and effective time use was nearly perfect, indicating that the concordance of measures assessing the same underlying domain also depends on the referent to which the measures refer. This may sound trivial but this point is largely ignored in teaching quality research, where different aspects of classroom management such as the absence of disturbances and monitoring are often used more in the sense of interchangeable domain indicators. In addition, our results showed that these two sets of measures were differently related to classroom student composition. Teachers who taught lower-performing students and a higher proportion of male students received lower ratings on classroom management measures referring to students, but not on a measure referring to the teacher.

These differences are of high practical and theoretical importance, as they show that compositional effects cannot be explained purely by a general bias in students' perceptions, as compositional effects depend on which measures of classroom management are used. Rather, we believe that measures referring more to students than to the teacher provide information on classroom management from a perspective that combines both teachers' abilities and class characteristics. That is, lower or higher ratings on indicators referring to the students cannot simply be equated with the teacher's ability, but need to be considered from an interactionist perspective that takes into account both teachers and the students taught.

7.2 Classroom Management and Student Achievement

Furthermore, our results showed that the examined measures of classroom management were differently related to students' later achievement. Whereas factors referring to students were related to students' pretest-adjusted class achievement, teachers' monitoring did not exhibit a statistically significant effect. Thus, the findings of the present study provide important insights into the frequently observed association between classroom student composition and students' learning. Over the last two decades, empirical studies have shown that a favorable classroom composition provides important benefits above and beyond students' individual learning backgrounds (e. g., Harker & Tymms, 2004). This research suggests that being part of higher-achieving classrooms leads to greater learning, even after controlling for students' personal characteristics (e. g., learning ca-

pabilities and parents' educational background). The present study makes a strong contribution to this research. At the same time, however, these findings also raise the question to what extent the much-reported associations between classroom management and student learning are merely due to classroom composition effects.

7.3 Limitations and Future Research

In sum, our results raise questions about the differences between frequently-used classroom management measures assessed via student ratings. However, the study has important limitations that should be addressed in future research. First, the main aim of our study was to compare classroom management measures with varying referents. Thus, we used classroom management measures that differed in respect of whether or not they referred to the teacher. Even though the study was based on frequently used and well-known aspects of classroom management, the measures used do not allow us to separate the referent from the content of the classroom management constructs examined. For instance, the reason that classroom student composition was related to the absence of disturbances/effective time use but not to monitoring, may be due to content-related differences between measures, instead of the varying referents between measures.

A second limitation refers to the conceptual differences underpinning classroom management measures with varying referents. The present findings indicate that existing indicators tap into theoretically distinct aspects of the classroom management process, ranging from the assessment of concrete behavioral operations and diagnostic aspects on the side of the teacher, to the effectiveness of these operations, given the specific characteristics of the class being taught. Future research should systematically examine to what extent the referent of measures reflects conceptually different aspects of the classroom management process, including teachers' classroom management-related diagnostic abilities, their behaviors, and students' responses in a specific class. Particularly productive would be a longitudinal study using measures with varying referents and applying an interactional perspective, to explain how well teachers' management behavior meets the specific requirements of the class.

Third, considering a larger number of classroom management measures is also important for another reason. Although monitoring reflects a prototypical measure of a teacher's classroom management, other teacher-directed measures of classroom management were not able to be examined in the present study. For this reason, the inclusion of additional measures with a teacher referent (e. g., rule setting) would further clarify potentially confounding effects of teacher-directed measures on measures that are not explicitly teacher-referenced.

Finally, our results relied solely on student ratings of classroom management. The findings showed that students' views on classroom management reflected theoretically-assumed differences, in terms of the underlying management processes, but we were not able to include alternative methods of assessment. Thus, it remains an open question as to whether similar findings would also result from observation ratings or from teacher

self-reports (see Campbell & Ronfeldt, 2018). In addition, it has to be borne in mind that the sole reliance on student ratings has potential to impede the accessibility of teachers' monitoring, leading to non-statistically significant results. As pointed out by Fauth and colleagues (submitted), judgments about teacher-directed behavior place higher cognitive demands on the judgment process of students than do judgments about their own behavior. In the same vein, the use of alternative methods of assessment would also help to clarify whether the high correlations between measures referring to students are a result of using the same referent or rather, of students' lower ability to differentiate between theoretically distinct dimensions of classroom management.

In conclusion, the present study has shown that the way students are asked about classroom management can make a difference to the extent to which such measures provide information about teachers and students in the classroom. The results point to the complex nature of the classroom management process, which involves both teachers and students within a particular class.

References

Aldrup, K., Klusmann, U., Lüdtke, O., Göllner, R., & Trautwein, U. (2018). Social support and classroom management are related to secondary students' general school adjustment: A multilevel structural equation model using student and teacher ratings. *Journal of Educational Psychology, 110*(8), 1066–1083.

Baumert, J., Gruehn, S., Heyn, S. Köller, O., & Schnabel, K. U. (1997). *Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU) – Dokumentation (Band 1) [Educational trajectories and psychosocial development during adolescents (BIJU) – Documentation (Issue 1)]*. Berlin: Max-Planck-Institut für Bildungsforschung.

Bos, W., Gröhlich, C., Dudas, D.-F., Guill, K., & Scharenberg, K. (2010). *KESS 8 – Skalenhandbuch zur Dokumentation der Erhebungsinstrumente [KESS 8 – Codebook to documenting study measures]*. Münster: Waxmann.

Campbell, S. L., & Ronfeldt, M. (2018). Observational evaluation of teachers: Measuring more than we bargained for? *American Educational Research Journal, 55*(6), 1233–1267.

Creemers, B. P. M., Kyriakides, L., & Antoniou, P. (2013). Establishing theoretical frameworks to describe teacher effectiveness. In B. P. M. Creemers, L. Kyriakides & P. Antoniou (Eds.), *Teacher professional development for improving quality of teaching* (pp. 101–135). Dordrecht, Netherlands: Springer.

Doyle, W. (2013). Ecological approaches to classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management* (pp. 107–136). Mahwah, NJ: Erlbaum.

Ganzeboom, H. B. G., & Treiman, D. J. (2003). Three internationally standardized measures for comparative research on occupational status. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Eds.), *Advances in cross-national comparison: A European working book for demographic and socio-economic variables* (pp. 159–193). New York: Kluwer Academic/Plenum.

Gettinger, M., & Kohler, K. (2006). Process-outcome approaches to classroom management and effective teaching. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 73–96). Mahwah, NJ: Erlbaum.

Gustafsson, J.-E., & Åberg-Bengtsson, L. (2010). Unidimensionality and interpretability of psychological instruments. In S. E. Embretson (Ed.), *Measuring psychological constructs: Advances in model-based approaches* (pp. 97–121). Washington, DC: American Psychological Association.

Harker, R., & Tymms, P. (2004). The effects of student composition on school outcomes. *School Effectiveness and School Improvement, 15,* 177–199.

Hattie, J. (2009). *Visible learning: A synthesis of meta-analyses relating to achievement.* London: Routledge.

Helmke, A. (2010). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts [Instructional quality and teacher professionalism. Diagnosis, evaluation, and improvement of instruction].* Seelze, Germany: Klett-Kallmeyer.

Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research, 30,* 1–21.

Hamre, B. K., & Pianta, R. (2010). Classroom environments and developmental processes: Conceptualization, measurement, & improvement. In J. L. Meece & J. S. Eccles (Eds.), *Handbook of research on schools, schooling and human development* (pp. 25–41). New York, NY: Routledge.

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras Study: Investigating the effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Münster: Waxmann.

Koch, T., Holtmann, J., Bohn, J., & Eid, M. (2018). Explaining general and specific factors in longitudinal, multimethod, and bifactor models: Some caveats and recommendations. *Psychological Methods, 23,* 505–523.

Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research,* 9, 231–251.

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teacher: Effects on instructional quality and student development. *Journal of Educational Psychology, 105,* 805–820.

Landrum, T. J., & Kauffman, J. M. (2006). Behavioural approaches to classroom management. In C. M. Everton & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 47–72). Mahwah, NJ: Erlbaum.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide. Seventh Edition.* Los Angeles, CA: Muthén & Muthén.

Organisation for Economic Co-operation and Development. (2004). *Learning for tomorrow's world: First results from PISA 2003.* Paris: OECD.

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,* 191–209.

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rost, J., & Schiefele, U. (2007). *PISA 2003 [Version 1, Code book and dataset].* Berlin: Institute for Educational Quality Improvement.

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rost, J., & Schiefele, U. (2013). *PISA-I-Plus 2003, 2004 [Version 1, Code book and dataset].* Berlin: Institute for Educational Quality Improvement.

Ramm, G., Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rolff, H.-G., Rost, J., & Schiefele, U. (2006). *PISA 2003. Dokumentation der Erhebungsinstrumente [PISA 2003. Documentation of the study instruments].* Münster: Waxmann.

Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B., & Trautwein, U. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology, 108,* 705–721.

Zusammenfassung: In der vorliegenden Studie wurden die Konsequenzen variierender Referentenbezüge (Lehrkraft vs. Schülerinnen und Schüler) bei etablierten Skalen zur Erfassung der Klassenführung aus Schülersicht untersucht. Die Ergebnisse einer Reanalyse der PISA 2003-Daten zeigten, dass Klassen mit höherem Jungenanteil und niedrigerem mittleren Leistungsniveau niedrigere Werte auf Skalen mit stärkerem Schülerbezug aufwiesen. Diese waren wiederum mit einer geringeren Leistungsentwicklung von Schülerinnen und Schülern im Fach Mathematik assoziiert. Für eine Skala mit Lehrkraftbezug fanden sich hingegen keine Zusammenhänge. Die Ergebnisse legen nahe, dass der Referentenbezug von Skalen für die erfassten Aspekte der Klassenführung entscheidend ist.

Schlagworte: Klassenführung, Schülerurteile, Itemreferenten, Klassenkomposition, Mathematikleistung

Contact

Prof. Dr. Richard Göllner, University of Tübingen,
Hector Research Institute of Education Sciences and Psychology,
Europastraße 6, 72072 Tübingen, Germany
E-Mail: richard.goellner@uni-tuebingen.de

Prof. Dr. Benjamin Fauth, Institut für Bildungsanalysen Baden-Württemberg (IBBW),
Heilbronner Str. 172, 70191 Stuttgart, Germany
E-Mail: benjamin.fauth@ibbw.kv.bwl.de

Prof. Dr. Gerlinde Lenske,
Universitätsallee 1, 21335 Lüneburg, Germany
E-Mail: gerlinde.lenske@leuphana.de

Prof. Dr. Anna-Katharina Praetorius, University of Zurich,
Institute of Education,
Freiestrasse 36, 8032 Zurich, Switzerland
E-Mail: anna.praetorius@ife.uzh.ch

Dr. Wolfgang Wagner, University of Tübingen,
Hector Research Institute of Education Sciences and Psychology,
Europastraße 6, 72072 Tübingen, Germany
E-Mail: wolfgang.wagner@uni-tuebingen.de