



Contents lists available at ScienceDirect

## Learning and Instruction

journal homepage: [www.elsevier.com/locate/learninstruc](http://www.elsevier.com/locate/learninstruc)

## Heterogeneity in students' perceptions of teaching quality: A marker of adaptive teaching?

Sebastian Röhl<sup>a,b,\*</sup> , Kyle Davison<sup>c</sup>, Lisa Bardach<sup>c,d</sup>

<sup>a</sup> University of Education Freiburg, Institute of Education, Kunzenweg 21, 79117, Freiburg, Germany

<sup>b</sup> University of Tübingen, Institute of Education, Münzgasse 30, 72070, Tübingen, Germany

<sup>c</sup> University of Giessen, Department of Psychology, Otto-Behagel-Straße 10, 35394, Giessen, Germany

<sup>d</sup> Kaunas University of Technology, Kaunas, Lithuania

## ARTICLE INFO

## Keywords:

Teaching quality

Student ratings

Heterogeneity

Adaptive teaching

## ABSTRACT

**Background:** Students attending the same class can vary greatly in their perceptions of the teaching quality of the same lessons with the same teacher; however, the current understanding of heterogeneity in students' perceptions of teaching quality, its sources, and implications remains limited.

**Aims:** This study investigated whether heterogeneity in students' perceptions of teaching quality within classes may indicate that the teacher's instruction is not equally adaptive for all students.

**Sample:** We used longitudinal data from the TALIS Global Teaching InSights study ( $N = 19,659$  students,  $N = 679$  mathematics teachers, across eight countries).

**Methods:** We ran multigroup latent change score models to examine the effect of student-reported adaptive teaching on (a) teaching quality ratings (clarity of instruction, autonomy support, student-teacher relationships, cognitive engagement, support for learning, support for competence, and classroom disruptions) and, importantly, (b) heterogeneity in students' perceptions of these dimensions.

**Results:** Our findings revealed that adaptive teaching predicted lower heterogeneity in students' perceptions—particularly regarding autonomy support, clarity of instruction, and, to some extent, support for learning, support for competence, and student-teacher relationships. Adaptive teaching was related to less classroom disruptions and more heterogeneous perceptions of disruptions within classes.

**Conclusions:** Our findings indicate that adaptive teaching can explain the extent of heterogeneity in students' perceptions of several teaching quality dimensions. This underscores the substantive value of considering heterogeneity in student perceptions as a meaningful construct.

Teachers play an important role in promoting students' educational outcomes, such as their motivation or achievement. Hence, teachers' behavior, and specifically the quality of their teaching, is relevant for students' positive development and learning outcomes (e.g., Kuhfeld, 2017; Maulana & Helms-Lorenz, 2016; Seidel & Shavelson, 2007). To assess teaching quality, surveys capturing students' perceptions of teaching quality have become widely used (e.g., Praetorius et al., 2018).

When analyzing student surveys, researchers are faced with their multi-level data structure as students are typically grouped in classes. If data from many individual students within a class is available, researchers can aggregate individual student ratings at the class level, yielding a measure of the “shared perceptions” of teaching quality, (i.e., the mean ratings of the students within each class). Aggregated student

ratings of teaching quality have thus been used to study teaching quality, conceptualized as a feature that varies between classes (e.g., Lüdtke et al., 2006). At the same time, students in the same class who rate the same teacher can vary greatly in their perceptions. While heterogeneity in perceptions of students within a class has often been ignored or treated as “noise” and something to control for, an emerging line of research has started to explore heterogeneity in students' perceptions in more depth, focusing on its substantive meaning, sources, and implications (e.g., Bardach et al., 2019; Bardach et al., 2021; Schenke et al., 2017, 2018; Schweig, 2016; Wittwer, 2008).

The present study aims to contribute to the emerging literature on heterogeneity in student-rated teaching quality by establishing a link between heterogeneity in students' perceptions of teaching quality and

\* Corresponding author. University of Education Freiburg, Institute of Education, Kunzenweg 21, 79117 Freiburg, Germany.

E-mail address: [sebastian.roehl@ph-freiburg.de](mailto:sebastian.roehl@ph-freiburg.de) (S. Röhl).

<https://doi.org/10.1016/j.learninstruc.2025.102245>

Received 8 July 2024; Received in revised form 25 September 2025; Accepted 30 September 2025

Available online 7 October 2025

0959-4752/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

adaptive teaching. Adaptive teaching refers to teachers' attempts to tailor their teaching to the varying needs and characteristics of students (e.g., Corno, 2008; Hardy et al., 2019). Here, we propose that, if students in the same class feel equally supported and perceive that the teacher adapts teaching to their individual needs and learning prerequisites (e.g., Bondie et al., 2019; Corno, 2008; Tetzlaff et al., 2022), then heterogeneity in teaching quality perceptions within classes should be low. Conversely, a lack of tailored teaching should increase heterogeneity in student-rated teaching quality. Hence, we propose that adaptive teaching may predict heterogeneity in students' perceptions of teaching quality, or put differently, that heterogeneity in student-rated teaching quality may serve as a marker of adaptive teaching.

To provide, to the best of our knowledge, the first empirical test of this hypothesis, we leveraged a large, longitudinal multi-country dataset with two measurement points (TALIS GTI survey; OECD, 2020) and examined whether adaptive teaching predicted heterogeneity in students' perceptions of teaching quality (clarity of instruction, cognitive engagement, support for learning, support for competence, autonomy support, student-teacher relationship, and classroom disruptions) within classrooms. In addition, we looked at effects of adaptive teaching on class-mean levels of all teaching quality dimensions (i.e., more positive perceptions of clarity of instruction, cognitive engagement, support for learning, support for competence, autonomy support, student-teacher relationship, and classroom disruptions). Overall, this study advances the current knowledge on heterogeneity in students' teaching quality perceptions, with potential implications for educational practice (e.g., on the role of adaptive teaching).

### 1. Teaching quality: Dimensions and heterogeneity in students' perceptions

The question of which teaching characteristics have a particularly beneficial effect on the educational outcomes (e.g., students' subject-related knowledge, motivation, students' self-concept) has long been of central interest in educational research. Various conceptions of teaching quality have been developed (Darling-Hammond, 2021). These conceptions differ not only in terms of the teaching characteristics included but also in terms of the number of quality dimensions (e.g., Charalambous et al., 2021; Ferguson & Danielson, 2014; Praetorius & Charalambous, 2018). However, certain central characteristics can be identified in almost all conceptions (Praetorius & Charalambous, 2018; Röhl et al., 2025). In the present study, we focused on the following teaching quality dimensions (in italics):

A set of important teaching quality characteristics comprises forms of social and academic support from the teacher; for example, *autonomy support* provided by the teacher and a supportive *student-teacher relationship* figure prominently in several frameworks.

In addition, teaching quality concepts focus on the communication of learning content, such as comprehensibility (e.g., *clarity of instruction*), the stimulation of cognitive learning processes (e.g., students' *cognitive engagement*), and teachers' support for students' learning processes (e.g., *support for learning, support for competence*).

A disruption-free working atmosphere, teachers' effective handling of disruptions, and the establishment of order in the classroom (often referred to as classroom management), are central components of teaching quality. Thereby, researchers often focus on the absence of *classroom disruptions* as an important indicator of such a disruption-free learning atmosphere.

Among the various ways to assess teaching quality, student ratings have become the most commonly used. When using student ratings, all students in a class are usually surveyed using questionnaires, and class-averages of the scales are formed (i.e., the mean of the ratings of all students within a class). However, prior research has documented substantial heterogeneity in the ratings of students attending the same class, with most of the variance in students' perceptions existing within rather than between different classes (e.g., Schenke et al., 2018). This

heterogeneity in students' perceptions of teaching quality is relevant from both a methodological perspective (e.g., a sufficient degree of overlap in students' perceptions of constructs, such as teaching quality, has been regarded as a precondition to the aggregation of individual student ratings at the classroom level, e.g., Lüdtke et al., 2006; Nelson et al., 2016), and a substantive perspective, which is the main focus of our investigation. Prior to further elaborating on heterogeneity from a substantive perspective, we briefly cover how heterogeneity can be conceptualized and operationalized.

### 2. Measuring heterogeneity in students' perceptions of teaching quality

Heterogeneity in students' perceptions of teaching quality can be defined and operationalized in different ways. First, heterogeneity can be understood as the amount of variability in the ratings of the students within a class. Following this, heterogeneity is commonly assessed with variance-based interrater-agreement indices (e.g., Lindell, 1999; Burke et al., 1999). These interrater-agreement indices yield one value for each class that can be used as a (class-level) variable if researchers want to study relations between heterogeneity and other constructs (for an overview see e.g., Lüdtke et al., 2006; for empirical studies see e.g., Bardach et al., 2018; Bardach et al., 2019, 2021; Schweig, 2016; Wittwer, 2008). Depending on the specific index, the resulting value is then often referred to as "agreement" (or "within-class consensus") rather than "heterogeneity"; however, as agreement simply represents the absence of heterogeneity, such indices lend themselves well to the study of heterogeneity.

Second, researchers have adopted profile-based approaches, in which they first clustered students according to their configurations of teaching quality perceptions on multiple dimensions. Next, the extent of profile heterogeneity within classes was calculated. As for agreement indices, the resulting value of profile heterogeneity (one per class) can then be used in further analyses (Schenke et al., 2017).

Third, there are more indirect ways to approach heterogeneity, which build on the assumption that multiple sub-groups exist within a classroom unit (e.g., Seidel, 2006; see also Robinson, 2023). Students within each sub-group have more similar perceptions of teaching quality, whereas the perceptions of members of different subgroups within classes differ. Hence, the existence of different sub-groups within classes—instead of one overall "climate" (i.e., shared perceptions of all students within a class)—can underlie heterogenous student perceptions if the class is looked at as a whole (see e.g., Schweig, 2016 discussing this issue). This approach can be labelled as "indirect", as it does not directly measure heterogeneity within classes and instead pays attention to sub-groups within classes. For example, a recent study employed social network analysis to identify sub-groups of students within classes who were friends with each other and then explored if friends within a class become more similar in their perceptions of teaching (Bardach et al., 2024).

In the present study, we opted for the first approach and thus used an interrater agreement index (see Methods section for methodological details). We did so because the aim of this study was to directly examine heterogeneity in students' perceptions of multiple teaching quality perceptions and to investigate them separately to be able to draw differentiated conclusions.

### 3. Searching for the substantive meaning of heterogeneity in students' perceptions of teaching quality

From a substantive perspective, a better understanding of heterogeneity in students' teaching quality perceptions, its sources and consequences, can help to better understand teaching and learning processes more generally. It has, for example, been argued that classrooms in which students have highly varying perceptions might not be ideal environments for learning, compared with classrooms in which

students do not vary as much in their perceptions (Schenke et al., 2017). Hence, heterogeneity could carry unique information on the classroom learning environment that is distinct to the information captured by class-mean levels of these ratings (i.e., how a class, on average, rates the teaching quality of a lesson or teacher; see also e.g., Bardach et al., 2021; Schenke et al., 2017; Schweig, 2016). In high heterogeneity-classes, friction and less goal-directed learning behavior may occur due to the students' varying perceptions and different interpretations (Bardach et al., 2021; Schenke et al., 2017). In line with these claims, it has been shown that higher levels of heterogeneity in teaching quality predicted lower learning gains (Schenke et al., 2017; Schweig, 2016).

In addition to studying consequences of heterogeneity in students' teaching quality perceptions (e.g., Bardach et al., 2018; Bardach et al., 2019; Schenke et al., 2017; Schweig, 2016), scholars have sought to gain insights into sources of heterogeneity. Given the central role of teachers as facilitators of student learning and important sources of academic and social support, it has been proposed that specific teacher characteristics and specific teacher behavior may give rise to higher or lower levels of heterogeneity in students' perceptions of teaching quality. For instance, Schenke et al. (2018) found that teacher unfairness/unfriendliness, measured via classroom observations carried out by researchers, predicted heterogeneity in students' ratings of teachers' emotional support. A further study on teacher characteristics (Bardach et al., 2021) investigated, among other factors, the effects of teacher-reported teacher characteristics (teachers' emotional exhaustion, teaching-related anxiety and enjoyment) on agreement (i.e., the absence of heterogeneity) in students' perceptions of teaching quality. It was shown that teachers' emotional exhaustion, in particular, and, to a lesser extent, teaching-related anxiety, negatively predicted agreement regarding several teaching quality aspects. No statistically significant effect occurred for teaching-related enjoyment. In explaining their findings for emotional exhaustion, the authors proposed that the teaching strategies of teachers who feel emotionally exhausted may be less consistent over time and across students. This could then increase heterogeneity in students' perceptions of teaching quality (Bardach et al., 2021).

#### 4. Heterogeneity in student-perceived teaching quality as a potential marker of adaptive teaching

The present work expands upon existing research addressing the substantive meaning of heterogeneity in students' perceptions. We propose that a high level of heterogeneity in students' perceptions of teaching quality may indicate that a teacher is not successfully meeting the individual needs of the various students forming the class. Thus, we conceptualize heterogeneity in students' perceptions of teaching quality as a marker (i.e., indicator) of adaptive teaching in that higher levels of adaptive teaching should be predictive of less heterogeneous perceptions of teaching quality. This assumption has already been put forth by several scholars focusing on teaching quality and heterogeneity in students' perceptions (e.g., see Bardach et al., 2021; Lüdtke et al., 2006; Schenke et al., 2017; Schweig, 2016); however, empirical research directly testing this link is, so far, absent.

##### 4.1. Adaptive teaching

Teachers' adjustments to students' individual needs and differences are generally considered to be fundamental components of effective teaching (see Parsons et al., 2018). Various terms and conceptions have been used for the process by which teaching-learning processes can be adapted to students' individual needs and differences, such as adaptive teaching (e.g., Corno, 2008; Vaughn et al., 2021), personalized learning (e.g., Bernacki et al., 2021), instructional adaptations (e.g., Parsons et al., 2018), individualized instruction (e.g., Connor et al., 2018), differentiated instruction (e.g., Bondie et al., 2019), and others (see e.g., Tetzlaff et al., 2022). However, these terms do not necessarily have an identical meaning, can differ in their underlying theoretical

assumptions, terminology, and instructional arrangements (Pozas & Schneider, 2019), and should thus not be used interchangeably. In addition, it has been stressed that adaptation can be delivered by the teacher, via technology (e.g., intelligent tutoring systems, Spitzer et al., 2024), or a combination of both (e.g., when teachers use information provided by adaptive learning technology).

Our study focuses on adaptive teaching as this concept best aligns with the present investigation. Adaptive teaching operates on a continuum ranging from adjusting the pace of instruction, level of difficulty, or explanations for the class or groups of students to individualized learning and the support of self-directed learning. Adaptive teaching describes the process by which teachers align their teaching to the varying needs and characteristics of students. Adaptive teaching generally refers to teaching practices that are non-routine and flexible in order to actively and proactively meet the needs of a particular student or learning situation (see Hardy et al., 2019). For Corno (2008), these adaptations can occur both at the macro level of teaching through instructional design and implementation, and at the micro level of teaching through teacher-student interactions. Teachers may vary in the degree to which their instructional design creates affordances for adaptive teaching during classroom instruction (Hardy et al., 2019).

It should be mentioned that some conceptualizations of teaching quality consider adaptive teaching as a component of teaching quality (e.g., van de Grift, 2007), and adaptive teaching and related terms have also subsumed under specific teaching quality dimensions (e.g., student support in Praetorius & Charalambous, 2018). However, in the current study, we follow recent integrative conceptualizations that consider adaptive teaching (referred to there as adaptation) as forming a grounding layer of teaching quality aspects (e.g., autonomy support, clarity of instruction), acknowledging its importance for all teacher behaviors and interactions with students (Charalambous & Praetorius, 2020).

##### 4.2. Adaptive teaching and its relationship to mean levels and heterogeneity in specific teaching quality dimensions

Overall, it can be assumed that teaching that adapts to the needs and learning requirements of all students (i.e., adaptive teaching) should have positive effects. First, studies indicate that students rate the teaching quality of teachers with higher adaptive teaching competency more positively (Brühwiler & Blatchford, 2011; Brühwiler & Vogt, 2020). Hence, *adaptive teaching should be predictive of higher class-mean ratings of teaching quality*. Second, adaptive teaching aims to make learning more effective, but also serves to ensure that every student participates actively and successfully in lessons. Hence, in line with conceptualizations of heterogeneity in students' perceptions of teaching quality as marker of adaptive teaching, we thus suggest that *adaptive teaching should be predictive of lower levels of heterogeneity in students' ratings of teaching quality* as well: If high levels of adaptive teaching prevail, students with different needs and learning prerequisites should feel equally supported and should share similar teaching quality perceptions. Fig. 1 displays these conceptual considerations. It should be noted that while we propose that the patterns (adaptive teaching predicts more positive teaching quality perceptions and lower heterogeneity in teaching quality perceptions) should in general hold across most teaching quality dimensions, there may also be exceptions (as outlined below for classroom disruptions).

Here, we outline theoretical assumptions that link adaptive teaching to students' perceptions of teaching quality dimensions, both in terms of class-mean levels of student ratings and heterogeneity in the ratings of students within classes. Thereby, we focus on the specific teaching quality dimensions that are investigated in our study: clarity of instruction, cognitive engagement, autonomy support, and classroom disruptions.

First, adaptive teaching is needed to ensure that all students can understand the structure and content of lessons, as captured in the

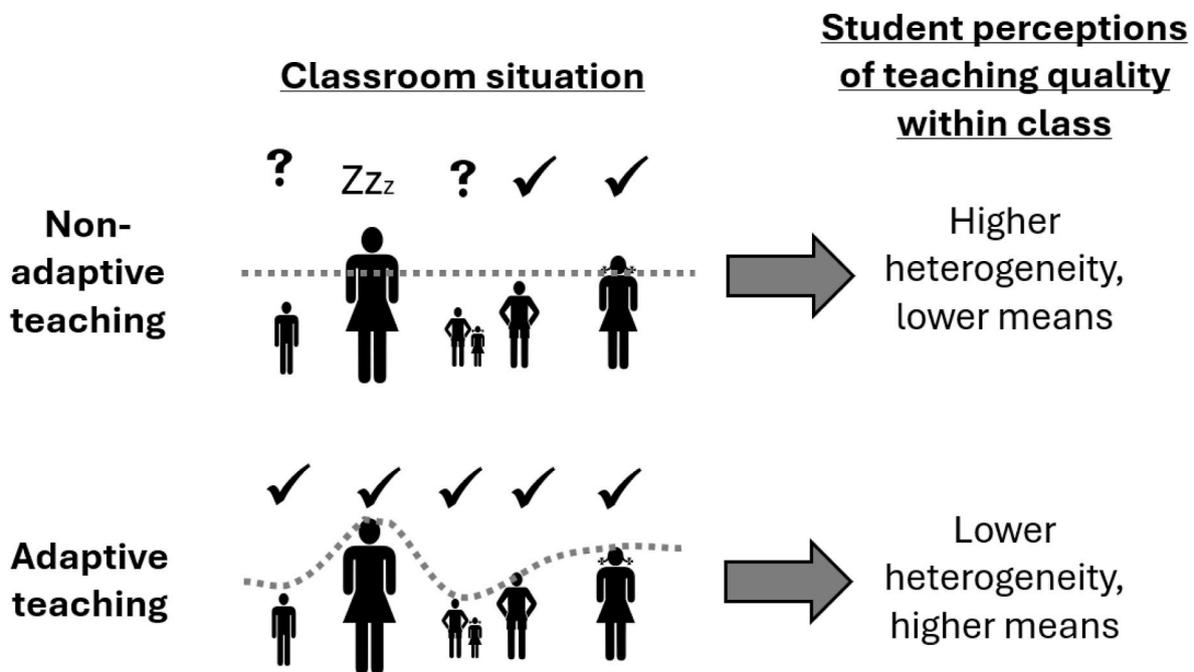


Fig. 1. Conceptual illustration of the possible consequences of adaptive teaching on student perceptions of teaching quality.

teaching quality dimension *clarity of teaching* (Brühwiler & Blatchford, 2011); accordingly, in the presence of high levels of adaptive teaching, students within a class should, on average, perceive the clarity of teaching more positively (i.e., at higher class-mean levels). Moreover, as adaptive teaching caters instruction (e.g., the clarity of teaching) to the potentially differing needs of students within a class, higher rating of adaptive teaching should also be conducive to reducing heterogeneity in students' perceptions of clarity of teaching (i.e., lower heterogeneity).

Similarly, adaptive teaching ensures that all students engage cognitively and actively with the learning content at an appropriate level (e.g., teaching quality dimension *cognitive engagement*; Corno, 2008). Hence, adaptive teaching should be linked to students feeling more cognitively engaged on average (i.e., higher class-mean levels of cognitive engagement). Given that in classes characterized by higher levels of adaptive teaching, teaching practices fostering cognitive engagement take place at the right level for different students or groups of students, it seems conceptually plausible that adaptive teaching should also give rise to lower levels of heterogeneity in students' perceptions of cognitive engagement.

Furthermore, if a teacher successfully adapts their teaching, all students should perceive higher levels of *autonomy support* that fits to their learning needs and, as all students are supported, there should be lower levels of heterogeneity with respect to the provision of autonomy support (McCaslin et al., 2015). More precisely, considering that adaptive teaching seeks to enable students to become more self-directed and independent in their learning (Corno, 2008), adaptive teaching should be linked to more positive perceptions of autonomy support (i.e., higher class-mean levels). As adaptive teaching, per definition, aims to enable all students with their differing needs and learning prerequisites to become more self-directed and independent in their learning and makes them feel equally supported, it should also reduce heterogeneity in students' perceptions of autonomy support. Similarly, adaptive teaching helps each student feel supported and recognized (Corno, 2008; Gasser et al., 2017; Hardy et al., 2019), which may contribute to more positive perceptions of both other types of support for students (i.e., *support for learning* and *support for competence*) as well as student-teacher relationships and less heterogeneity in those perceptions.

For *classroom disruptions*, the relationship with adaptive teaching is less clear. On the one hand, it is conceivable that adaptive teaching may

be predictive of fewer student-perceived classroom disruptions. A mechanism underlying this effect could be that adaptive teaching leads to higher learning motivation of students (Corno, 2008; Hardy et al., 2019), which may result in fewer classroom disruptions from non-motivated or under- or over-stretched students. On the other hand, adaptive teaching may result in more disruptions as teachers may provide tailored support to individual students to a greater extent and spend only a few phases of the lesson interacting with the class as a whole (McCaslin et al., 2015). It is thus possible that the interactive nature of adaptive teaching may create noisier environments, and the reduced presence of the teacher at the front of the classroom may feed into higher levels of disruptive student behavior. To summarize, with respect to class-mean levels of classroom disruptions, both positive and negative effects of adaptive teaching seem theoretically plausible.

Regarding heterogeneity in students' perceptions of classroom disruptions, strong conceptual links to adaptive teaching are, from our perspective, lacking. Still, we acknowledge that successfully adapting instructions to the different needs of students within a class may also prompt teachers to select appropriate routines for all students, which may then contribute to less heterogeneous perceptions of classroom disruptions within a class. However, since our study focuses exclusively on the presence or absence of classroom disruptions rather than on classroom management strategies, such as implementing routines for students, a direct connection to the heterogeneity in students' perceptions of classroom disruptions is less apparent.

## 5. The current study

This study provides the first empirical test of the novel hypothesis that heterogeneity in students' perceptions of specific teaching quality aspects serves as a marker of adaptive teaching in that adaptive teaching predicts heterogeneity in students' perceptions of these specific teaching quality aspects. In addition to heterogeneity in students' perceptions, we also considered class-mean levels of teaching quality as further outcomes to contribute to a more comprehensive understanding.

The following research questions were addressed: Does adaptive teaching predict perceived teaching quality (i.e., higher class-mean levels), and importantly, heterogeneity in students' perceptions of teaching quality (i.e., lower heterogeneity)? We hypothesized that

adaptive teaching in terms of a better adjustment to the learning requirements of the class and individual students should result in (a) more positive teaching quality perceptions (i.e., higher class-mean levels) of the dimensions clarity of teaching, cognitive engagement, support for learning, support for competence, autonomy support, and student-teacher relationships), and, most importantly, (b) lower levels of heterogeneity in students' perceptions of these teaching quality dimensions. With respect to mean levels of classroom disruptions, we left it open whether the effect would be positive or negative. Due to a lack of strong conceptual links between adaptive teaching and heterogeneity in students' perceptions of classroom disruptions, we did not specify concrete hypotheses and conducted exploratory analyses.

Our study used data from the large multi-country TALIS Global Teaching InSights study (OECD, 2020). A mathematics teaching unit on the topic of quadratic equations lasting several weeks was examined in eight different countries, with assessments of teaching quality at two measurement points, both before (with a focus on the usual teaching practice so far) and after the unit (with a focus on the teaching practice during the unit).<sup>1</sup> We adopted latent change score models on the class level to analyze the data. Hence, even though we phrase our hypotheses in terms of "higher levels of adaptive teaching predicting less heterogeneity in teaching quality dimensions", the more precise, but lengthy, description would be that we are interested in whether "more positive perceptions of the teaching unit on quadratic equations, as compared to normal mathematics lessons, predicts a lower degree of heterogeneity in students' perceptions of teaching quality in the teaching unit on quadratic equations, as compared to heterogeneity in their perceptions in normal mathematics lessons."

In order to draw conclusions about the results across the different countries of the TALIS GTI study, we conducted a two-step integrative data analysis that has already been used for the international integration of national findings from other large-scale assessments (Brunner et al., 2022; Curran & Hussong, 2009).

## 6. Methods

### 6.1. Sample and procedure

This study uses student surveys (adaptive teaching, teaching quality aspects) from the TALIS Global Teaching InSights study (OECD, 2020), which includes data from Chile, Colombia, England, Germany, Japan, Spain, Mexico, and Shanghai (for more detailed information on the regions surveyed, see OECD, 2020). It surveyed teaching quality perceptions of  $N = 19,659$  secondary school students and their  $N = 679$  math teachers (between 50 and 103 per country) at two measurement time-points before and after a teaching unit on quadratic equations lasting 6 to 13 lessons (depending on national curriculum standards; Klieme & Schweig, 2020). As there were problems in the data in one country (Madrid/Spain) with the exact allocation of students and classes between the measurement points (Herbert et al., 2022; OECD, 2020), we only used the other seven countries with  $N = 16,143$  students and  $N = 594$  math teachers in the study.

The average age of the students (51 % female) was 14.9 ( $SD = 1.1$ ), ranging from 10.1 to 21.4 years. More details on the recruitment process and the characteristics of the sample can be found in the technical reports of the TALIS GTI study (OECD, 2021).

<sup>1</sup> As reported by Opfer et al. (2020), between 20 % and 55 % of participating teachers in individual countries stated that they had prepared their lessons more carefully than usual during the study. Therefore, we assume that teachers focused more strongly during the study on what they considered to be the most effective way of teaching.

### 6.2. Measures

To assess teaching quality dimensions, we used student ratings. Adaptive teaching was assessed from both the students' and their teachers' perspective. All items are listed at the end of the supplementary document. Please note that we attempted to include as many teaching quality dimensions as possible in our analyses; however, we could not include all of those that were assessed for various reasons (e.g., assessed only once, low reliability of some scales in many countries). Tables S1 and S2 in the Online Supplement provides an overview of these issues.

### 6.3. Adaptive teaching

The adaptive teaching scale in the student questionnaire consisted of five items (see Supplemental Table S3 for exact item wordings and item-factor loadings). However, a closer inspection revealed that three of the items referred to the teacher's adaptation of the entire lesson to the prerequisites and emerging needs of the *whole class* (e.g., "Our mathematics teacher adapts the lessons to my class's needs and knowledge") and addressed situational adaptations (microadaptations) made by the teacher during the lessons (Corno, 2008; e.g., "Our mathematics teacher changes the structure of the lesson on a topic that most students find difficult to understand"). In contrast, the fourth item referred to the adaptation of tasks to the learning prerequisites of individual students ("Our mathematics teacher gives different work to students of different ability levels"). Since the latter item conceptually differed from the other items (see also Hardy et al., 2019) and had only weak loadings on the overall scale ( $T1: \lambda = .29$ ;  $T2: \lambda = .35$ ), we decided to analyze the two aspects of adaptive teaching separately as *class-related adaptive teaching* and *student-related adaptive teaching*. Additionally, the fifth item of the original scale focused on teachers' use of diagnostic questions as another different aspect of teaching (formative assessment) and showed only weak loadings on the original scale ( $T1: \lambda = .52$ ;  $T2: \lambda = .59$ ). Therefore, we decided to exclude this item from further analyses. All items referring to adaptive teaching used a 4-point agreement answering scale, ranging from 1 = "strongly disagree" to 4 = "strongly agree." The scale assessing class-related adaptive teaching showed sufficient reliability in each of the seven countries ( $.71 \leq \text{McDonald's } \omega \leq .94$ ; Mihaly et al., 2021).

### 6.4. Teaching quality dimensions

For indicators of students' perceptions of teaching quality, we used the student-reported scales *clarity of instruction* (4 items,  $.69 \leq \omega \leq .87$ , example "Our mathematics teacher presents a summary of recently learned content"), *cognitive engagement* (3 items,  $.63 \leq \omega \leq .86$ , "I think intensively about the mathematical content"), *support for learning* (3 items,  $.75 \leq \omega \leq .91$ , "Our mathematics teacher continues teaching until we understand"), *support for competence* (4 items,  $.86 \leq \omega \leq .94$ , "Our mathematics teacher makes me feel confident in my ability to learn the material"), *autonomy support* (4 items,  $.66 \leq \omega \leq .87$ , "Our mathematics teacher encourages me to find the best way to proceed by myself"), *student-teacher relationship* (5 items,  $.83 \leq \omega \leq .93$ , "My mathematics teacher makes me feel she/he really cares about me") and *classroom disruptions* (3 items,  $.81 \leq \omega \leq .91$ , "There is much disruptive noise in this classroom"). All measures use a 4-point agreement answering scale, ranging from 1 = "strongly disagree" to 4 = "strongly agree," except for clarity (4-point frequency scale).

In the latent change score models set up to address our research questions, only the class-mean values and heterogeneity measures (one index per class, see below) were used. Nonetheless, we also provide information on construct validity (confirmatory factor analyses for each country, following the approach of Kline (2016) in the Online Supplement for interested readers. The confirmatory factor analyses showed acceptable model fit indices at both measurement time points (Supplemental Table S4). In addition, we tested each scale in

longitudinal analyses (Supplemental Table S5;  $2.8 \leq \chi^2 \leq 202.8$ ;  $5 \leq df \leq 17$ ;  $.000 \leq p \leq .731$ ;  $.934 \leq CFI \leq 1.000$ ;  $.000 \leq RMSEA \leq .108$ ;  $.006 \leq SRMR \leq .135$ ).<sup>2</sup> In order to compare statistical effects from different countries, at least metric measurement invariance should be present. The analysis based on the criteria proposed by Rutkowski and Svetina (2014; changes of fit values between configural and metric model specification  $\Delta CFI < .02$  and  $\Delta RMSEA < .01$ ) pointed out that this requirement is met for the constructs used (Supplemental Table S6).

6.5. Heterogeneity in students' perceptions of teaching quality

For each student-rated teaching quality scale, heterogeneity of students' perceptions within classes was calculated using Lindell's  $r^*_{wg(J)}$  (Lindell et al., 1999), an interrater agreement index, using the following formula:

$$r^*_{wg(J)} = 1 - \frac{\bar{S}^2_{x_j}}{\sigma^2_{EU}}$$

Where  $\bar{S}^2_{x_j}$  is the average item variability (observed group variance),  $\sigma^2_{EU}$  is the expected random variance of a rectangular distribution of the item answers. Higher values of  $r^*_{wg(J)}$  indicate a higher agreement of students within classes (i.e., lower levels of heterogeneity).  $r^*_{wg(J)}$  can range from  $-0.8$  to  $1$  for a four-point scale and from  $-1$  to  $1$  for five-point scales, with lower values indicating higher heterogeneity. Compared to many other agreement indices,  $r^*_{wg(J)}$  has the advantage that it is independent of the number of items on a scale and that different assumptions regarding the expected random variance lead to linearly dependent values (O'Neill, 2017). The standardized regression coefficients reported in this study are therefore independent of the assumed random variance and number of items. In return, we accept some disadvantages, such as the possibility of a relatively higher correlation with the scale mean, which is less relevant in our study.

Missing values only occurred at the student level. Two percent of the students did not answer any of the items used here at either the first or second measurement time point. A further 6.1 % only took part in the first measurement time point and a further 3.8 % only in the second. Single missing values were only present for around 2.4 % of students and can be considered negligible (Graham, 2009). As we calculated mean values and heterogeneity measures at class level for each time point separately, the complete items available within a scale at the respective time point were therefore used. On average, the number of valid student responses per class was 25.2 ( $SD = 7.3$ ).

6.6. Statistical analyses

To test our hypotheses, we specified dual latent change score models on the class level (McArdle, 2009, Fig. 2). Separate models were specified for adaptive teaching and heterogeneity in teaching quality perceptions, and for adaptive teaching and mean levels of teaching quality. Fig. 2 provides an example of a dual latent change score model for heterogeneity. In this model, the latent change scores for the adaptive teaching variable (in Fig. 2:  $\Delta\eta_{adapt}$ ) and the heterogeneity index ( $\Delta\eta_{rwg}$ ) of the included teaching quality aspect are partialized between the two measurement points using latent regression analysis. Hence, we examined the effect of adaptive teaching during the specific lesson unit, while

<sup>2</sup> For the adaptive teaching scales and the teaching quality scales clarity, cognitive engagement, autonomy support, RMSEA values lay below .08 and the SRMR value below .07, with the exception of the support for autonomy scale in Japan (RMSEA = .108). For the scale classroom-management - disruptions, RMSEA and SRMR values in Shanghai, Germany, Japan, and England lay in a slightly higher range (.064  $\leq$  RMSEA  $\leq$  .100; .094  $\leq$  SRMR  $\leq$  .135). According to Kenny et al. (2015), however, these values are acceptable due to the small number of degrees of freedom in the model ( $df = 17$ ).

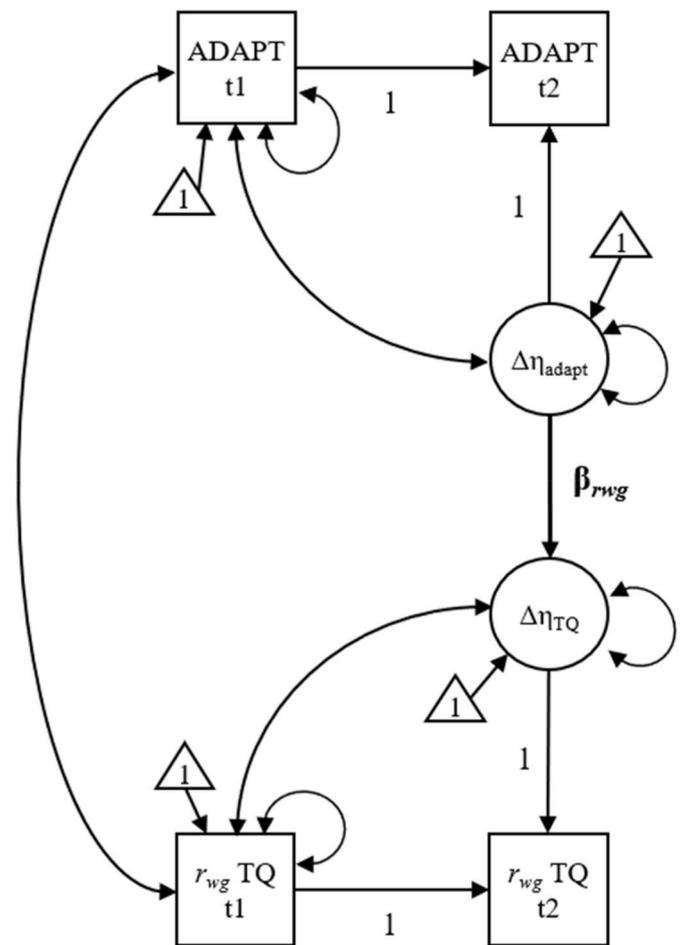


Fig. 2. Dual latent change score model estimating the effect of the change in adaptive teaching on heterogeneity in student perceptions of teaching quality.

controlling for general adaptive teaching in mathematics lessons, on the heterogeneity in students' perceptions of teaching quality during that unit—also controlling for their general perception heterogeneity in mathematics lessons.

We accounted for the different origin countries of the data by using multigroup structural equation modeling, in which the same dual latent change score model was estimated for every country without restrictions between the groups. Analyses were conducted using maximum likelihood estimation with robust (Huber-White) standard errors and a scaled Yuan-Bentler test statistic.

We examined the fit of the latent change score models using model  $\chi^2$  statistics, Steiger-Lind root mean square error of approximation (RMSEA), Bentler comparative fit index (CFI) and standardized root mean square residuals (SRMR), whereby an acceptable model fit is assumed for  $CFI \geq .95$ , and  $SRMR \leq .08$  and as suggested by Kline (2016). As the relatively small number of 16 degrees of freedom of the LCSM combined with a relatively small sample size per country can lead to an increase in the RMSEA index (Kenny et al., 2015), slightly higher values up to a maximum of  $RMSEA \leq .15$  are also tolerated.

To summarize the individual effects of the seven countries, we conducted a two-step integrative data analysis (Brunner et al., 2022; Curran & Hussong, 2009) that has recently been increasingly used in studies with data from international large-scale assessments in order to integrate effects from different countries and survey waves as used for other international large-scale assessments. In this analysis method, the coefficients for the individual subgroups are estimated and then combined using meta-regressive random-effect models. Statistical effects from different countries can be compared if - as in the present analyses

(Supplemental Table S6) – at least metric measurement invariance is given (Putnick & Bornstein, 2016). As with other meta-analytic methods, the integrative data analysis can be used to draw inferences about average effects across different countries, but it is not suitable for addressing country-specific differences.

All analyses were conducted in R using the packages *multilevel* (Bliese, 2022), *lavaan* (Rosseel, 2012) and *metafor* (Viechtbauer, 2010).

7. Results

7.1. Descriptive analyses

The mean values, standard deviations, and pre-post changes of all manifest variables in the overall sample are shown in Table 1. In this study,  $r^*_{wg(J)}$  values (agreement index), our measure of heterogeneity in teaching quality perceptions, ranged between 0.47 and 0.66 for the teaching quality scales at T1 and between 0.50 and 0.63 for the teaching quality scales at T2. According to common interpretation guidelines,  $r^*_{wg(J)}$  values of above 0.30 indicate weak agreement, values above 0.50 indicate moderate agreement, values above 0.70 indicate strong agreement, and values above 0.90 indicate very strong agreement (LeBreton & Senter, 2008). Thus, values observed in this study reflect a generally moderate amount of agreement.

A significant increase in student-related adaptive teaching between T1 and T2 ( $p < .001$ , effect size Hedges'  $g = .037$ ) emerged, whereas class-related adaptive teaching decreased ( $p < .001$ ,  $g = -0.35$ ). Additionally, there were significant changes between T1 and T2 in means and heterogeneity for the teaching quality measures across the entire sample, except for mean clarity of instruction and heterogeneity of cognitive engagement and disruptions. There were significant decreases for the scale mean values of the teaching quality dimensions ( $p < .001$ ,  $-0.12 < g < -0.40$ ), except for an increase in the mean values of disruptions ( $p < .001$ ,  $g = 0.50$ , i.e., more disruptions).

Values on the agreement index used to assess heterogeneity in students' perceptions of teaching quality decreased significantly ( $p < .001$ ) for autonomy support ( $g = -0.26$ ), support for learning ( $g = -0.40$ ), support for competence ( $g = -0.26$ ), and for student-teacher relationship ( $g = -0.33$ ), which indicated an increase in the heterogeneity of perceptions within the classes at T2. Conversely, perceived clarity of instruction showed higher agreement (i.e., lower heterogeneity) at T2 ( $p < .001$ ,  $g = 0.20$ ). Correlations of all variables used in the latent change-

score analyses can be found in Supplemental Table S7.

7.2. Latent change-score analyses

All fit statistics for the latent change-score models lay in an acceptable range ( $10.3 < \chi^2(14) < 32.2$ ;  $.009 < p < .740$ ;  $.983 < CFI < 1.00$ ;  $.000 < RMSEA < .124$ ;  $.022 < SRMR < .065$ ; see Supplemental Table S8 for details). The weighted-mean effects of adaptive teaching change between T1 and T2 on student perceptions of teaching quality are presented in Table 2. We focus on the results across countries in our

Table 2

Weighted-mean standardized regression coefficients of the effects of the student and teacher perceived adaptive teaching change on the heterogeneity and mean change of different dimensions of teaching quality.

	Teaching Quality within Classes					
	Heterogeneity $r^*_{wg(J)}$			Mean		
	$\beta_{rwg}$	SE	p	$\beta_M$	SE	p
<b>Student-Related Adaptive Teaching</b>						
Clarity of Instruction	.08	.04	.044	.18	.04	<.001
Cognitive Engagement	-.09	.04	.018	.12	.04	.001
Support for Learning	.02	.04	.541	.16	.04	<.001
Support for Competence	.05	.04	.175	.25	.04	<.001
Autonomy Support	.09	.04	.014	.39	.03	<.001
Student-Teacher Relationship	.07	.04	.068	.28	.04	<.001
Classroom Management – Disruptions	-.10	.04	.005	-.19	.04	<.001
<b>Class-Related Adaptive Teaching</b>						
Clarity of Instruction	.24	.03	<.001	.50	.03	<.001
Cognitive Engagement	-.03	.04	.468	.17	.04	<.001
Support for Learning	.24	.04	<.001	.60	.03	<.001
Support for Competence	.19	.04	<.001	.56	.03	<.001
Autonomy Support	.19	.04	<.001	.52	.03	<.001
Student-Teacher Relationship	.20	.04	<.001	.55	.03	<.001
Classroom Management – Disruptions	.04	.04	.268	.02	.04	.546

Note. Heterogeneity was assessed using  $r^*_{wg(J)}$ , which is an agreement index (i.e., higher values reflect higher within-class agreement and thus, lower within-class heterogeneity),  $\beta$  = standardized regression coefficient of the change in adaptive teaching on students' teaching quality perceptions. SE = standard error. p = two-tailed significance value of regression coefficient.

Table 1

Overall means, standard deviations and pre-post changes of the variables used in the latent change-score models.

		T 1		T2		T(591)	p	ES g
		M	SD	M	SD			
Adaptive Teaching	Student-related	2.09	0.43	2.20	0.42	-8.31	<.001	0.34
	Class-related	3.07	0.30	3.01	0.31	8.37	<.001	-0.34
Clarity of Instruction	$r^*_{wg(J)}$	0.47	0.15	0.50	0.16	-4.90	<.001	0.20
	Mean	3.09	0.35	3.10	0.37	-1.04	.149	0.04
Cognitive Engagement	$r^*_{wg(J)}$	0.57	0.11	0.56	0.13	0.96	.168	-0.04
	Mean	2.80	0.25	2.78	0.28	2.93	.002	-0.12
Autonomy Support	$r^*_{wg(J)}$	0.57	0.13	0.53	0.15	6.91	<.001	-0.28
	Mean	3.03	0.30	2.99	0.33	6.41	<.001	-0.26
Classroom Disruptions	$r^*_{wg(J)}$	0.55	0.15	0.55	0.17	0.67	.252	-0.03
	Mean	2.78	0.51	2.88	0.50	-12.26	<.001	0.50
Support for Learning	$r^*_{wg(J)}$	0.66	0.14	0.63	0.16	5.42	<.001	-0.22
	Mean	3.28	0.30	3.21	0.32	9.65	<.001	-0.40
Support for Competence	$r^*_{wg(J)}$	0.57	0.15	0.54	0.17	5.55	<.001	-0.23
	Mean	3.09	0.35	3.04	0.36	6.45	<.001	-0.26
Student-Teacher Relationship	$r^*_{wg(J)}$	0.62	0.13	0.58	0.16	8.13	<.001	-0.33
	Mean	3.14	0.30	3.12	0.33	3.86	<.001	-0.16

Note. M = mean. SD = standard deviation. T, p = test statistics of one-sided dependent T-Test of pre-post changes. ES g = effect size Hedge's g of pre-post change.  $r^*_{wg(J)}$  = agreement index according to Lindell et al. (1999).

manuscript, but we also provide separate results for the individual countries in Tables Supplement S9-S12 for interested readers. In the following, we describe the findings separately for student-related adaptive teaching and class-related adaptive teaching. Fig. 3 provides an overview of the effects.

### 7.3. Effects of student-related adaptive teaching

A higher level of student-related adaptive teaching predicted higher agreement (i.e., less within-class heterogeneity) in students' perceptions of *clarity* ( $\beta_{rwg} = .08, p = .044$ ) and *autonomy support* ( $\beta_{rwg} = .09, p = .014$ ), and higher mean levels of these dimensions (i.e., higher levels of *clarity*,  $\beta_M = .18, p < .001$ , and *autonomy support*,  $\beta_M = .39, p < .001$ ). Additionally, higher levels of student-related adaptive teaching predicted higher mean values of *support for learning* ( $\beta_{rwg} = .06, p < .001$ ) and *support for competence* ( $\beta_{rwg} = .25, p < .001$ ) as well as higher mean values of *student-teacher relationship* ( $\beta_{rwg} = .28, p < .001$ ); however, there were no statistically significant effects on heterogeneity in these dimensions. Higher levels of student-related adaptive teaching were predictive of higher mean values for *cognitive engagement* ( $\beta_M = .12, p = .001$ ), but also predicted lower agreement (i.e., larger heterogeneity) in student ratings within classes ( $\beta_{rwg} = -.09, p = .018$ ). Further, regarding the *classroom disruptions* scale, higher levels of student-related adaptive teaching predicted a lower mean (i.e., less disruptions,  $\beta_M = -.19, p < .001$ ) and lower agreement (i.e., a higher degree of heterogeneity;  $\beta_{rwg} = -.10, p = .005$ ) of students' perceptions of classroom disruptions within classes.

### 7.4. Effects of class-related adaptive teaching

Higher levels of class-related adaptive teaching predicted higher levels of agreement (i.e., lower levels of heterogeneity) of students' perceptions of *clarity* ( $\beta_{rwg} = .24, p < .001$ ), *support for learning* ( $\beta_{rwg} = .24, p < .001$ ), *support for competence* ( $\beta_{rwg} = .19, p < .001$ ), *autonomy support* ( $\beta_{rwg} = .20, p < .001$ ), and *student-teacher relationship* ( $\beta_{rwg} = .20, p < .001$ ). No significant effects emerged for *cognitive engagement* or *disruptions*. Regarding the class averages (mean levels) in student perceptions, higher levels of student-reported class-related adaptive teaching predicted higher levels of *clarity* ( $\beta_M = .50, p < .001$ ), *support for learning* ( $\beta_M = .60, p < .001$ ), *support for competence* ( $\beta_M = .56, p < .001$ ), *autonomy support* ( $\beta_M = .54, p < .001$ ), *student-teacher relationship* ( $\beta_M = .55, p < .001$ ), and, to a lesser extent, in *cognitive engagement* ( $\beta_M = .17, p < .001$ ), but not *disruptions*.

## 8. Discussion

While heterogeneity in students' perceptions of teaching quality has often been dismissed as "noise", emerging research suggests that this heterogeneity carries substantive meaning (e.g., Bardach et al., 2021; Schenke et al., 2017, 2018; Schweig, 2016). The aim of this study was to contribute to this emerging line of research by investigating, for the first time, whether heterogeneity in students' perceptions of teaching quality within a class can serve as a marker of adaptive teaching. Leveraging a large, multi-country data set (TALIS Global Teaching InSights study, OECD, 2021), we therefore tested whether adaptive teaching predicts heterogeneity in students' perceptions of teaching quality in terms of clarity of instruction, cognitive engagement, support for learning, support for competence, autonomy support, student-teacher-relationship, and classroom disruptions. In addition to effects of adaptive teaching on heterogeneity in students' teaching quality perceptions, we investigated effects of adaptive teaching on mean levels of teaching quality. The consideration of both mean levels and heterogeneity in students' teaching quality perceptions contributes to a more comprehensive understanding, and allows comparing effects on mean levels, the typical focus in research on teaching quality, and heterogeneity in perceptions as crucial extension. Another novel contribution of our study is that we investigated two aspects of adaptive teaching separately. Specifically, we distinguished between adaptive teaching related to individual students (student-related adaptive teaching) and adaptive teaching that also considers tailoring teaching to classroom processes more generally (class-related adaptive teaching).

Our findings relating to adaptive teaching (both student-related and class-related adaptive teaching) supported our hypothesis that heterogeneity in students' perceptions of teaching quality serves as a marker of adaptive teaching for several teaching quality aspects. For example, the effects on heterogeneity of student perceptions of clarity of instruction within classes indicate that adaptive teaching contributes to lower heterogeneity in student ratings, possibly because the teaching is comprehensible at "the right level" for all students: If teachers are successful in adapting their teaching to different students' needs and learning prerequisites, all students should receive relatively unambiguous and comprehensible messages (e.g., Bardach et al., 2021), which likely gives rise to a more shared perception of clarity of instruction within classes (i.e., lower heterogeneity). Moreover, both student-reported student-related and class-related adaptive teaching predicted higher mean levels of clarity of instruction. The effects of adaptive teaching on mean levels of clarity of instruction suggest that adaptive teaching helps to make

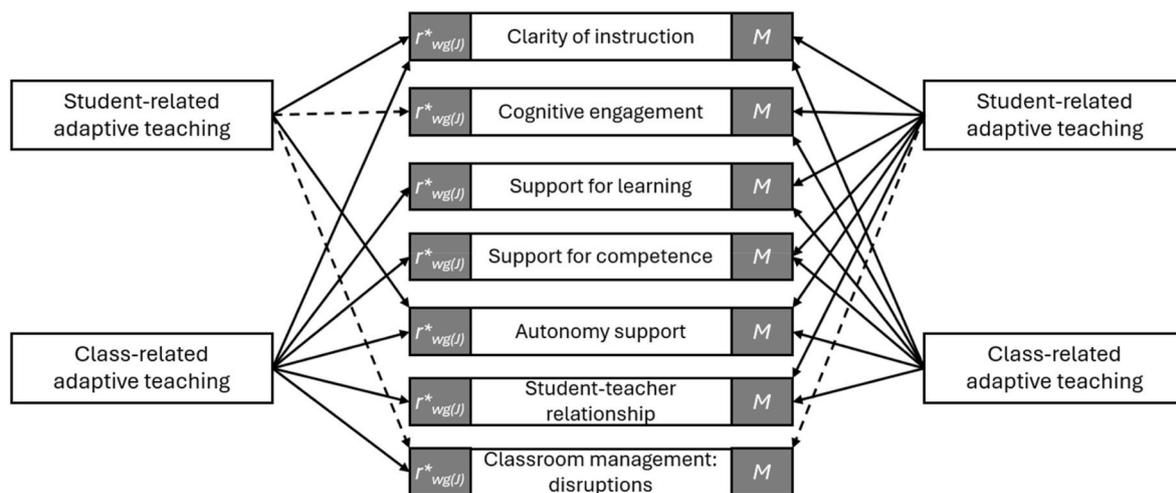


Fig. 3. Overview of the effects of student-related and class-related adaptive teaching on heterogeneity ( $r^*_{wg(I)}$ ) and mean levels ( $M$ ) of teaching quality dimensions within Classes.

Note. Solid lines: Significant positive effects. Dashed lines: Significant negative effects. No lines: No statistically significant effects.

instructions more comprehensible on average, which adds to existing findings on the positive effects of adaptive teaching documented in the literature (e.g., Brühwiler & Blatchford, 2011; Brühwiler & Vogt, 2020).

For autonomy support, our assumption that adaptive teaching is predictive of higher mean levels of and less heterogeneity in teaching quality perceptions in classes was also confirmed. With respect to heterogeneity, we propose that adaptive teaching makes it easier for *all* students to find tasks fitting their learning needs and preferences (see also Corno, 2008), which could explain the reduced heterogeneity in autonomy support perceptions. In addition, if teachers take the differing autonomy needs and learning difficulties of students into account, they should be perceived as more autonomy supportive by students, on average (i.e., higher mean levels).

Students' reports of student-related and class-related adaptive teaching were predictive of higher mean levels of *cognitive engagement*, but not to a reduction in within-class heterogeneity. The fact that we found consistent support for positive effects of both types of adaptive teaching on mean levels, in line with our hypothesis, indicates that adaptive teaching helps to increase the extent to which students in a class feel cognitively engaged. Since cognitive engagement is a key precursor to successful domain-specific learning processes (e.g., Kunter et al., 2013), identifying its predictors—such as the mean levels of different forms of adaptive teaching—is relevant for both research and educational practice. On the other hand, the non-significant effect of class-related adaptive teaching on heterogeneity, and the finding that higher levels of student-related adaptive teaching even predicted higher heterogeneity contradict our hypothesis and raise questions. Cognitive engagement is a strongly “cognitively loaded” teaching aspect, and it may be particularly challenging for teachers to adapt their teaching so that everyone is cognitively engaged to the same or a highly similar degree. For example, high-quality adaptive teaching not only requires a large pool of suitable tasks at different difficulty levels, it is necessary also for either the teacher to assign the different tasks to students in a way that is befitting their ability or competency, or for students to be adequately trained to choose the appropriate level of difficulty for themselves (Bondie et al., 2019). Especially the adaptation of tasks to the learning prerequisites of individual students within a class (i.e., student-related adaptive teaching) may be challenging for teachers in that it may even increase gaps in the perceptions of different groups of students within classes, as manifested in higher levels of heterogeneity in cognitive engagement.

Furthermore, it should be noted that the cognitive engagement dimension differs from the other dimensions in two key respects. First, it is the only scale that exclusively uses “I” as the item referent (e.g., “I think intensively about the mathematical content”), whereas most other scales refer to the “teacher” (e.g., “My mathematics teacher is interested in my well-being”) or to the classroom environment (e.g., “There is much disruptive noise in this classroom”). This distinction has been discussed in several studies on student perceptions of teaching quality, particularly regarding how the use of “I” versus “we” in item wording may affect interpretation (Jaekel et al., 2022; see also den Brok et al., 2006). While minor differences did emerge in some cases, they warrant further investigation. In our study, we did not systematically examine item formulation effects within the same scale (see, e.g., Jaekel et al., 2022). Future research using the same scale with two versions of item referents—differing by addressee—and investigating the effects of adaptive teaching on heterogeneity in both versions could provide further clarity. Second, the cognitive engagement dimension differs from the others in that it is the only one used here that focuses on the individual student's experience and learning behavior during mathematics lessons, rather than on the teaching provided by the teacher—such as the use of cognitively activating tasks and instructional methods. Unfortunately, we were unable to include the cognitive activation scale in our analyses for comparison purposes due to its low reliability.

For support for learning, support for competence, and student-teacher

relationships, the same pattern of findings emerged: higher levels of class-related adaptive teaching predicted less heterogeneity in students' perceptions of these three dimensions, whereas student-related adaptive teaching showed no significant effects. For both types of adaptive teaching, however, effects on the mean levels of these dimensions were observed. It is also possible that class-related adaptive teaching, having been assessed more comprehensively than student-related adaptive teaching, may better capture aspects that contribute to the more shared student perceptions with respect to these dimensions. With respect to student-teacher relationships, it is possible that providing tasks at different levels—as examined here through student-related adaptive teaching—may do little to change heterogeneity in perceptions (i.e., no significant effect): Students receiving easier tasks due to lower achievement levels might feel less liked or valued, while higher-achieving students may perceive the teacher more positively. Similar mechanisms may unfold for support for learning and support for competence. Class-related adaptive teaching, by contrast, may capture aspects such as situational adaptations (Corno, 2008) that are more relevant for predicting heterogeneity in these dimensions. Such situational *microadaptations* (e.g., adjusting the structure of a lesson that most students find difficult) focus less on making differences between students more salient, as the student-related adaptive teaching scale might, and more on connecting with and supporting everyone.

Lastly, the findings revealed that students' ratings of greater student-related adaptive teaching predicted lower mean levels of perceived *classroom disruptions* and greater heterogeneity in students' perceptions of the presence of disruptions. The observed effect of student-related adaptive teaching in reducing mean levels of classroom disruptions suggests that it can facilitate smoother teaching and learning processes, thus contributing to the growing body of literature highlighting the positive effects of adaptive teaching (e.g., Brühwiler & Blatchford, 2011; Brühwiler & Vogt, 2020). Interestingly, student-related adaptive teaching increased heterogeneity in perceptions of classroom disruptions. We propose that student-related adaptive teaching likely requires a more individualized form of support from the teacher (McCaslin et al., 2015), such as one-on-one discussions about individual questions or problems with the respective student. As such, student-related adaptive teaching perhaps presents a particular challenge when it comes to observing all students, their behavior, and to reacting promptly and effectively to emergent disruptions (McCaslin et al., 2015), which could manifest in increasingly heterogeneous perceptions. It is also conceivable that the larger heterogeneity in perceptions of disruptions within classes could be attributed to the fact that students' expectations, sensitivities and needs regarding noise levels during periods of the lesson in which student-related adaptive teaching prevails may differ, as reflected in greater heterogeneity in perceptions of classroom disruptions. Moreover, no statistically significant effects occurred for students' ratings of class-related adaptive teaching on classroom disruptions. These findings underscore that the two adaptive teaching forms are differentially related to students' perceptions of classroom disruptions. Importantly, student-related adaptive teaching seems to capture content more relevant to both mean levels and heterogeneity in students' perceptions of classroom disruptions, and we propose that this may be due to the stronger focus of student-related adaptive teaching on individualized adaptations.

### 8.1. Implications for research and practice

By testing whether heterogeneity in students' perceptions of adaptive teaching serves as a marker of adaptive teaching, the present study contributes to the emerging line of research focusing on the substantive meaning of heterogeneity in students' ratings of teaching quality, and to respective theory building. Our work showed that higher level of heterogeneity in students' perceptions of autonomy support, clarity of instruction, and – with regard to class-related adaptive teaching – support for learning, for competence and student-teacher relationship can

indicate a lack of adaptive teaching; at the same time, our study highlights important differences across teaching quality aspects and types of adaptive teaching, given that adaptive teaching did not predict heterogeneity in students' perceptions of cognitive engagement, for example, and given that we obtained more significant effects for class-related than student-related adaptive teaching. The finding that adaptive teaching even predicted an increase in heterogeneity in students' perceptions of classroom disruption is relevant for theory as well, as it underscores that adaptive teaching likely does not function in the same way across teaching quality aspects when the focus is on heterogeneity in students' ratings. Thus, there is a clear need for research on heterogeneity in students' perceptions to think thoroughly about contents of teaching quality aspects and their conceptual links to different potential sources of heterogeneity (in our study: adaptive teaching), and to let these conceptual considerations guide the theory-informed selection of constructs and the specification of hypotheses.

In addition, our study has implications for recent conceptualizations of teaching quality, which emphasize the importance of adaptive teaching for all teacher behaviors and interactions with students by considering adaptive teaching as grounding layer of teaching quality aspects (Charalambous & Praetorius, 2020). Our findings that (student-reported) adaptive teaching was consistently related to positive changes in mean levels of all teaching quality aspects support this theoretical proposition. Importantly, our work also suggests that these conceptualizations may benefit from expanding their focus to include heterogeneity in students' teaching quality perceptions as a feature that had so far received no or little attention in conceptual work on teaching quality.

Our work has implications for practice too. The obtained effects of adaptive teaching on positive changes in mean levels of all teaching quality aspects highlight the importance of adaptive teaching in creating learning environments that students perceive to be autonomy-supportive, characterized by high levels of clarity of instruction and few classroom disruptions, and beneficial for their cognitive engagement. Therefore, assisting teachers and prospective teachers in improving their adaptive teaching competencies represent important goals for teacher education, professional development programs, and interventions (e.g., Vogt & Rogalla, 2009).

Our findings may also have implications for the use of student surveys as feedback for the further development of teaching quality (e.g., student feedback; Röhl et al., 2021), even though we acknowledge that replications of our study's findings, especially regarding effects of adaptive teaching on heterogeneity in student-perceived teaching quality are needed before definite implications for practice can be derived. Nonetheless, we suggest that when using student surveys as feedback tools for teachers, it may be promising to complement reports of item or scale averages with respective measures of heterogeneity. Furthermore, teachers may need to learn to interpret heterogeneity within classes as a potential indication of the need for greater adaptive teaching with respect to certain aspects of teaching quality. As shown in a recent meta-analysis, a discussion of the student feedback between the teacher and the class with the aim of a collective development of possible improvements has proven to be very effective in improving teaching practices (Röhl et al., 2025). For this purpose, the heterogeneity measures of the individual scales could also be used, as this exchange may enable to obtain more detailed information on specific learning requirements of the students within a class and ways to better adapt teaching students' needs.

## 8.2. Limitations and directions for future research

Several limitations and promising directions for future work should be noted. Even though student ratings are considered as a reliable measure of teaching quality, they have some potential limitations. Students may tend to rate the teaching quality of teachers they like more highly (Bijlsma & Röhl, 2023). If a teacher is liked by all students in a

class, this could lead to less heterogeneity in student perceptions of teaching quality within a class. However, conversely, it can also be assumed that teachers with a higher level of adaptive teaching are more liked by the students because they take the students and their needs into account. This is also evident in the high correlations between class-related adaptive teaching and the average student-teacher relationship (Supplemental Table S7). Therefore, using further data sources would have been desirable. However, the video ratings available as part of the TALIS GTI study are only cross-sectional and could therefore not be considered in this longitudinal study. In addition, the TALIS-GTI study also provides teacher ratings, but they are generally considered less meaningful than student ratings (e.g., Wagner et al., 2015).

Bifurcating adaptive teaching into individual student- and class-related aspects was supported by the present findings as the two adaptive teaching aspects differed in their effects of student perceptions of some teaching quality dimensions. However, it remains a limitation that student-related adaptive teaching was only surveyed with one item. It would be useful to differentiate between the different types of adaptive teaching in further studies and to use two more comprehensive multiple-item scales. In addition, this would allow for the assessment of psychometric properties, including reliabilities and factorial structures, of all adaptive teaching scales—something that was not possible for student-related adaptive teaching in the current study.

We were able to use data from two measurement points; however, the interval between the measurement points was relatively short, so in typical longitudinal study, only a few changes in teaching practice would be expected. Nonetheless, many of the participating teachers reported that they prepared their lessons more carefully than usual between the measurements due to the video recordings (Opfer et al., 2020), making changes in student perceptions of lesson quality more likely. Without this unique situation, changes in student perceptions of a teacher's teaching practice would probably not have been expected, even over a longer period—particularly if teaching quality was assessed using questionnaires referring to a longer period (Gaertner & Brunner, 2018). In future longitudinal studies, our findings could, for example, be examined using a targeted intervention design with changes in adaptive teaching over time as treatment. Another promising approach could be to capture the actual dynamics of classroom processes relating to adaptive teaching as they unfold at shorter time scales (e.g., from day to day or even moment to moment). We envision future work following up on our study and using intensive longitudinal study designs (e.g., Davison et al., 2023) to shed light on fine-grained, dynamic, and potentially reciprocal associations between adaptive teaching and heterogeneity in students' perceptions of teaching quality. In addition, we employed one specific agreement index in our study. The use of other interrater agreement indices could possibly lead to slightly different effect sizes (for an overview of their advantages and disadvantages, see O'Neill, 2017), and we deem it worthwhile to systematically compare different indices in future work.

Another limitation of the present study lies in an implicit assumption underlying our argument: that students, regardless of differences in ability level, interest, or prior knowledge, are equally accurate in their evaluations of teaching quality. This assumption may not fully hold, as students' characteristics could systematically influence how they perceive and interpret teaching practices (e.g., Golke et al., 2022). Future research building on our work could also examine differences in student ratings across diverse learner profiles.

Lastly, we used class level aggregated data in our analyses, due to our interest in the class level and given that some of our variables (heterogeneity) can only be located at the class level. One limitation of our approach, however, is that information on the level of individual students was not used. Using individual student data and disaggregating information on the individual student and class level in future research would make it possible to reap the advantages of more complex statistical approaches, such as doubly latent multi-level modelling (e.g., Marsh et al., 2012). Applying our latent change score approach within a

doubly latent multi-level modelling framework would, for example, allow to simultaneously correct for both measurement and sampling error for constructs other than heterogeneity in perceptions (which is captured by one manifest index/class).

## 9. Conclusions

What does it *mean* when students in the same classroom differ in how they perceive teaching quality? This study contributes to the scarce body of research exploring the significance of heterogeneity in students' perceptions of teaching quality and its theoretical implications. Our findings suggest that adaptive teaching predicted lower heterogeneity in students' perceptions—particularly regarding autonomy support, clarity of instruction, and, to some extent, support for learning, competence, and student–teacher relationships. Thus, the present study provided initial evidence for our claim that heterogeneity in student-perceived teaching quality can serve as a marker of adaptive teaching.

## CRedit authorship contribution statement

**Sebastian Röhl:** Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Kyle Davison:** Writing – review & editing, Writing – original draft. **Lisa Bardach:** Writing – review & editing, Writing – original draft, Supervision, Resources, Funding acquisition, Conceptualization.

## Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 471695001, a Fellowship from the Elite Program for post docs of the Baden-Württemberg Foundation, and a Jacobs Foundation Research Fellowship (awarded to Lisa Bardach). Lisa Bardach is supported by research funding No P-EDU-23-28, which is co-funded by the European Union (the project “Breakthrough in Educational Research” No 10-044-P-0001) under the April 1, 2025 Agreement with the Research Council of Lithuania (RCOL, LMTLT) and the 14th April 2025 Joint Activity Agreement with Kaunas University of Technology.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.learninstruc.2025.102245>.

## Data availability

Original data is freely available from the OECD. The source code for the statistical analyses and the aggregated data used are available at <https://osf.io/dnqzt/>.

## References

- Bardach, L., Boda, Z., Neuendorf, C., Wagner, W., Davison, K., & Trautwein, U. (2024). Seeing the teacher through my peers' eyes? A social network study on adolescents' teaching quality perceptions. *Trends in Neuroscience and Education*, Article 100224. <https://doi.org/10.1016/j.tine.2024.100224>
- Bardach, L., Lüftenegger, M., Yanagida, T., Schober, B., & Spiel, C. (2018). The role of within-class consensus on mastery goal structures in predicting socio-emotional outcomes. *British Journal of Educational Psychology*, 89(2), 239–258. <http://doi.org/10.1111/bjep.12237>
- Bardach, L., Lüftenegger, M., Yanagida, T., Spiel, C., & Schober, B. (2019). Achievement or agreement – Which comes first? *Learning and Instruction*, 61, 72–83. <https://doi.org/10.1016/j.learninstruc.2019.01.003>
- Bardach, L., Yanagida, T., Morin, A. J. S., & Lüftenegger, M. (2021). *Is everyone in class in agreement and why (not)? Learning & instruction*. Advance Online Publication. <https://doi.org/10.1016/j.learninstruc.2020.101400>
- Bernacki, M. L., Greene, M. J., & Lobczowski, N. G. (2021). A systematic review of research on personalized learning: Personalized by whom, to what, how, and for what purpose(s)? *Educational Psychology Review*, 33(4), 1675–1715. <https://doi.org/10.1007/s10648-021-09615-8>
- Bijlsma, H. J. E., & Röhl, S. (2023). Fostering effective teaching at schools through measurements of student perceptions: Processes, risks and chances. In R. Maulana, M. Helms-Lorenz, & R. M. Klassen (Eds.), *Effective teaching around the world: Theoretical, empirical, methodological and practical insights* (pp. 119–135). Springer. [https://doi.org/10.1007/978-3-031-31678-4\\_6](https://doi.org/10.1007/978-3-031-31678-4_6)
- Bliese, P. (2022). *Multilevel: Multilevel functions* [Computer software]. <https://CRAN.R-project.org/package=multilevel>
- Bondie, R. S., Dahnke, C., & Zusho, A. (2019). How does changing “one-size-fits-all” to differentiated instruction affect teaching? *Review of Research in Education*, 43(1), 336–362. <https://doi.org/10.3102/0091732X18821130>
- Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction*, 21(1), 95–108. <https://doi.org/10.1016/j.learninstruc.2009.11.004>
- Brühwiler, C., & Vogt, F. (2020). Adaptive teaching competency: Effects on quality of instruction and learning outcomes. *Journal for Educational Research Online*, 12(1), 119–142. <https://doi.org/10.18747/phsg-coll3/id/1213>
- Brunner, M., Keller, L., Stallasch, S. E., Kretschmann, J., Hasl, A., Preckel, F., Lüdtke, O., & Hedges, L. V. (2022). Meta-analyzing individual participant data from studies with Complex Survey designs: A tutorial on using the two-stage approach for data from educational large-scale assessments. *Research synthesis methods*. Advance online publication. <https://doi.org/10.1002/jrsm.1584>
- Burke, M. J., Finkelstein, L. M., & Dusig, M. S. (1999). On average deviation indices for estimating interrater agreement. *Organizational Research Methods*, 2, 49–68. <https://doi.org/10.1177/109442819921004>
- Charalambous, C. Y., & Praetorius, A. K. (2020). Creating a forum for researching teaching and its quality more synergistically. *Studies In Educational Evaluation*, 67, Article 100894.
- Charalambous, C. Y., Praetorius, A.-K., Sammons, P., Walkowiak, T., Jentsch, A., & Kyriakides, L. (2021). Working more collaboratively to better understand teaching and its quality: Challenges faced and possible solutions. *Studies In Educational Evaluation*, 71, Article 101092. <https://doi.org/10.1016/j.stueduc.2021.101092>
- Connor, C. M., Mazocco, M. M., Kurz, T., Crowe, E. C., Tighe, E. L., Wood, T. S., & Morrison, F. J. (2018). Using assessment to individualize early mathematics instruction. *Journal of School Psychology*, 66, 97–113. <https://doi.org/10.1016/j.jsp.2017.04.005>
- Corno, L. (2008). On teaching adaptively. *Educational Psychologist*, 43(3), 161–173. <https://doi.org/10.1080/00461520802178466>
- Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, 14(2), 81–100. <https://doi.org/10.1037/a0015914>
- Darling-Hammond, L. (2021). Defining teaching quality around the world. *European Journal of Teacher Education*, 44(3), 295–308. <https://doi.org/10.1080/02619768.2021.1919080>
- Davison, K., Malmberg, L.-E., & Sylva, K. (2023). Academic help-seeking interactions in the classroom: A microlongitudinal study. *British Journal of Educational Psychology*, 93, 33–55. <https://doi.org/10.1111/bjep.12538>
- Ferguson, R. F., & Danielson, C. (2014). How framework for teaching and tripod 7Cs evidence distinguish key components of effective teaching. In T. J. Kane, K. A. Kerr, & R. C. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (pp. 98–143). Jossey-Bass. <https://doi.org/10.1002/9781119210856.ch4>
- Gaertner, H., & Brunner, M. (2018). Once good teaching, always good teaching? The differential stability of student perceptions of teaching quality. *Educational Assessment, Evaluation and Accountability*, 30(2), 159–182. <https://doi.org/10.1007/s11092-018-9277-5>
- Gasser, L., Grüter, J., Buholzer, A., & Wettstein, A. (2017). Emotionally supportive classroom interactions and students' perceptions of their teachers as caring and just. *Learning and Instruction*, 54, 82–92. <https://doi.org/10.1016/j.learninstruc.2017.08.003>
- Golke, S., Steininger, T., & Wittwer, J. (2022). What Makes Learners Overestimate Their Text Comprehension? The Impact of Learner Characteristics on Judgment Bias. *Educational Psychology Review*, 34(4), 2405–2450. <https://doi.org/10.1007/s10648-022-09687-0>
- Hardy, I., Decristan, J., & Klieme, E. (2019). Adaptive teaching in research on learning and instruction. *Journal for Educational Research Online*, 11(2), 169–191. <https://doi.org/10.25656/01:18004>
- Herbert, B., Fischer, J., & Klieme, E. (2022). How valid are student perceptions of teaching quality across education systems? *Learning and Instruction*, 82, Article 101652. <https://doi.org/10.1016/j.learninstruc.2022.101652>
- Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 44(3), 486–507. <https://doi.org/10.1177/0049124114543236>
- Klieme, E., & Schweig, J. (2020). Opportunities to learn. In *OECD (Ed.), global teaching InSights: A video Study of teaching* (pp. 246–265). OECD.
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.
- Kuhfeld, M. R. (2017). When students grade their teachers: A validity analysis of the tripod student survey. *Educational Assessment*, 22(4), 253–274. <https://doi.org/10.1080/10627197.2017.1381555>
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Eds.). (2013). *Cognitive activation in the mathematics classroom and professional competence of teachers*. Springer US. <https://doi.org/10.1007/978-1-4614-5149-5>
- LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, 11, 815–852. <https://doi.org/10.1177/1094428106296642>

- Lindell, M. K., Brandt, C. J., & Whitney, D. J. (1999). A revised index of interrater agreement for multi-item ratings of a single target. *Applied Psychological Measurement*, 23(2), 127–135. <https://doi.org/10.1177/01466219922031257>
- Lüdtke, O., Trautwein, U., Kunter, M., & Baumert, J. (2006). Reliability and agreement of student ratings of the classroom environment – A reanalysis of TIMSS data. *Learning Environments Research*, 9, 215–230. <https://doi.org/10.1007/s10984-006-9014-8>
- Maulana, R., & Helms-Lorenz, M. (2016). Observations and student perceptions of the quality of preservice teachers' teaching behaviour: Construct representation and predictive quality. *Learning Environments Research*, 19(3), 335–357. <https://doi.org/10.1007/s10984-016-9215-8>
- McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, 60, 577–605. <https://doi.org/10.1146/annurev.psych.60.110707.163612>
- McCaslin, M., Sotardi, V. A., & Vega, R. I. (2015). Teacher support and students' self-regulated learning: Co-regulation and classroom management. In E. T. Emmer, & E. J. Sabornie (Eds.), *Handbook of classroom management* (2nd ed., pp. 322–343). Routledge.
- Mihaly, K., Klieme, E., Fischer, J., & Doan, S. (2021). Questionnaire scale characteristics. In *Global teaching InSights: Technical report*. Paris: OECD. <https://www.oecd.org/edu/school/GTI-TechReport-Chapter18.pdf>.
- Nelson, P. M., Hall, G., & Christ, T. J. (2016). The stability of student ratings of the class environment. *Journal of Applied School Psychology*, 32(3), 254–267. <https://doi.org/10.1080/15377903.2016.1183543>
- O'Neill, T. A. (2017). An overview of interrater agreement on likert scales for researchers and practitioners. *Frontiers in Psychology*, 8, 777. <https://doi.org/10.3389/fpsyg.2017.00777>
- OECD. (2020). *Global teaching InSights: A video study of teaching*. OECD. <https://doi.org/10.1787/20d6f36b-en>
- OECD. (2021). *Global teaching InSights: Technical report*. <https://www.oecd.org/education/school/global-teaching-insights-technical-documents.htm>
- Opfer, V. D., Bell, C. A., Klieme, E., McCaffrey, D. F., Schweig, J., & Stecher, B. M. (2020). Understanding and measuring mathematics teaching practice. In *OECD (Ed.), global teaching InSights: A video study of teaching* (pp. 33–47). OECD.
- Parsons, S. A., Vaughn, M., Scales, R. Q., Gallagher, M. A., Parsons, A. W., Davis, S. G., Pierczynski, M., & Allen, M. (2018). Teachers' instructional adaptations: A research synthesis. *Review of Educational Research*, 88(2), 205–242. <https://doi.org/10.3102/0034654317743198>
- Pozas, M., & Schneider, C. (2019). Shedding light into the convoluted terrain of differentiated instruction (DI): Proposal of a taxonomy of differentiated instruction in the heterogeneous classroom. *Open Education Studies*, 1, 73–90. <https://doi.org/10.1515/edu-2019-0005>
- Praetorius, A. K., & Charalambous, C. Y. (2018). Classroom observation frameworks for studying instructional quality: Looking back and looking forward. *ZDM*, 50(3), 535–553. <https://doi.org/10.1007/s11858-018-0946-0>
- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of three basic dimensions. *ZDM*, 50(3), 407–426. <https://doi.org/10.1007/s11858-018-0918-4>
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Robinson, K. A. (2023). Motivational climate theory: Disentangling definitions and roles of classroom motivational support, climate, and microclimates. *Educational Psychologist*, 58(2), 92–110. <https://doi.org/10.1080/00461520.2023.2198011>
- Röhl, S., Bijlsma, H. J. E., & Rollett, W. (2021). The process model of student Feedback on Teaching (SFT): A theoretical framework and introductory remarks. In W. Rollett, H. J. E. Bijlsma, & S. Röhl (Eds.), *Student feedback on teaching in schools: Using student perceptions for the development of teaching and teachers* (pp. 1–14). Springer Nature. [https://doi.org/10.1007/978-3-030-75150-0\\_1](https://doi.org/10.1007/978-3-030-75150-0_1)
- Röhl, S., Bijlsma, H., & Schwichow, M. (2025). Can feedback from students to teachers improve different dimensions of teaching quality in primary and secondary schools? A hierarchical meta-analysis. *Educational Assessment, Evaluation and Accountability*. <https://doi.org/10.1007/s11092-024-09450-9>. Advance online publication.
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://www.jstatsoft.org/v48/i02/>.
- Schenke, K., Ruzek, E., Lam, A. C., Karabenick, S. A., & Eccles, J. S. (2017). Heterogeneity of student perceptions of the classroom climate: A latent profile approach. *Learning Environments Research*, 20(3), 289–306. <https://doi.org/10.1007/s10984-017-9235-z>
- Schenke, K., Ruzek, E., Lam, A. C., Karabenick, S. A., & Eccles, J. S. (2018). To the means and beyond: Understanding variation in students' perceptions of teacher emotional support. *Learning and Instruction*, 55, 13–21.
- Schweig, J. D. (2016). Moving beyond means: Revealing features of the learning environment by investigating the consensus among student ratings. *Learning Environments Research*, 19(3), 441–462. <https://doi.org/10.1007/s10984-016-9216-7>
- Seidel, T. (2006). The role of student characteristics in studying micro teaching-learning environments. *Learning Environments Research*, 9(3), 253–271.
- Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454–499.
- Spitzer, M. W., Bardach, L., Strittmatter, Y., Meyer, J., & Moeller, K. (2024). Evaluating the content structure of intelligent tutor systems—a psychological network analysis. *Computers and Education Open*, Article 100198.
- Tetzlaff, L., Hartmann, U., Dumont, H., & Brod, G. (2022). Assessing individualized instruction in the classroom: Comparing teacher, student, and observer perspectives. *Learning and Instruction*, 82, Article 101655. <https://doi.org/10.1016/j.learninstruc.2022.101655>
- van de Grift, W. (2007). Quality of teaching in four European countries: a review of the literature and application of an assessment instrument. *Educational Research*, 49(2), 127–152. <https://doi.org/10.1080/00131880701369651>
- Vaughn, M., Parsons, S. A., & Gallagher, M. A. (2021). Challenging scripted curricula with adaptive teaching. *Educational Researcher*, 1–11. <https://doi.org/10.3102/0013189X211065752>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48.
- Vogt, F., & Rogalla, M. (2009). Developing adaptive Teaching Competency through coaching. *Teaching and Teacher Education*, 25(8), 1051–1060. <https://doi.org/10.1016/j.tate.2009.04.002>
- Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B., & Trautwein, U. (2015). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology*, 108(5), 705–721. <https://doi.org/10.1037/edu0000075>
- Wittwer, J. (2008). What influences the agreement among student ratings of science instruction? *Zeitschrift für Erziehungswissenschaft*, 10, 205–220. [https://doi.org/10.1007/978-3-531-91815-0\\_11](https://doi.org/10.1007/978-3-531-91815-0_11)

Dr. Sebastian Röhl is interim professor at Freiburg University of Education. His research topics include teaching quality, feedback, professional development of teachers and school leaders.

Dr. Kyle Davison is Post-Doctoral Researcher at Giessen University. His research focuses interpersonal processes and the co-regulation of learning.

Dr. Lisa Bardach's research focuses on the role of individual differences in learning contexts and cultural diversity in schools. She holds a position as Full Professor for Psychology, Giessen University.