UNIVERSITY *of York*

This is a repository copy of *Different levels of context-specificity of teacher self-efficacy and their relations with teaching quality*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/187564/

Version: Accepted Version

Article:

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Different levels of context-specificity of teacher self-efficacy and their relations with teaching quality

Désirée Thommen[1*], Urs Grob[1], Fani Lauermann[2], Robert M. Klassen[3], Anna-Katharina Praetorius[1]

[1]Institute of Education, Faculty of Arts, University of Zurich, Switzerland, [2]Institute for School Development Research, Faculty of Education, Psychology and Sociology, Technical University of Dortmund, Germany, [3]Department of Education, University of York, United Kingdom

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

Keywords

Abstract

Word count: 140

On the basis of Bandura's social cognitive theory, researchers often assume that a teachers' self-efficacy (TSE) will have a positive effect on teaching quality. However, the available empirical evidence is mixed. Building on previous research into TSE, we examined whether assessing class-/task-specific TSE gives a more accurate indication of the associations between TSE assessments and student-rated teaching quality. The analyses were based on the English sample of the TALIS Video Study. Mathematics teachers (N = 86) rated their self-efficacy beliefs using generalized task-specific TSE items and class-/task-specific TSE items. Their students (N = 1930) rated the quality of teaching in their math class. Multilevel regression analyses revealed stronger associations between student-rated teaching quality and class-/task-specific TSE than generalized task-specific TSE. We discuss possible reasons for these results and outline the potential benefits of using class-specific assessments for future TSE research.

Contribution to the field

Teacher self-efficacy (TSE), a teacher's belief in their ability to influence student engagement and learning, is regarded as a key predictor of teaching quality. However, the empirical evidence on the relations between TSE and teaching quality is mixed. Mixed results may result in part from the use of different conceptualizations and assessments of TSE across studies. TSE assessments typically ask teachers to rate their TSE at the generalized, task-specific level and do not account for class-specific differences. However, TSE can vary significantly between different classes (e.g., depending on the subject or group of students taught). According to social cognitive theory, using class-specific TSE assessments should result in more consistent and comparatively stronger relations with teaching quality. No study to date has examined how well TSE scales with different levels of context-specificity may predict teaching quality. The key objective of the present study was to conduct comparative analyses of class-/task-specific versus generalized task-specific TSE scales and their associations with teaching quality in the same sample. Our analyses revealed stronger associations between teaching quality and class-/task-specific TSE relative to generalized task-specific TSE. Future research should consider the appropriate level of context-specificity in the conceptualization and assessment of TSE.

Ethics statements

Studies involving animal subjects
Generated Statement: No animal studies are presented in this manuscript.

Studies involving human subjects
Generated Statement: The studies involving human participants were reviewed and approved by Department Research Ethics Committee, University of Oxford Department of Education. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Inclusion of identifiable human data
Generated Statement: No potentially identifiable human images or data is presented in this study.

Data availability statement

Generated Statement: The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: **[https://www.oecd.org/education/school/global-teaching-insights-technical-documents.htm]**.

Different levels of context-specificity of teacher self-efficacy and their relations with teaching quality

Désirée Thommen[1]*, Urs Grob[1] , Fani Lauermann[2] , Robert M. Klassen[3] , Anna-Katharina Praetorius[1]

[1]Institute of Education, Faculty of Arts and Social Sciences, University of Zurich, Zurich, Switzerland

[2]Institute for Research on Education and School Development, TU Dortmund University, Dortmund, Germany

[3]Psychology in Education Research Centre, Department of Education, University of York, York, United Kingdom

*Correspondence:
Désirée Thommen
desiree.thommen@uzh.ch

Abstract

On the basis of Bandura's social cognitive theory, researchers often assume that a teachers' self-efficacy (TSE) will have a positive effect on teaching quality. However, the available empirical evidence is mixed. Building on previous research into TSE, we examined whether assessing class-/task-specific TSE gives a more accurate indication of the associations between TSE assessments and student-rated teaching quality. The analyses were based on the English sample of the TALIS Video Study. Mathematics teachers ($N = 86$) rated their self-efficacy beliefs using generalized task-specific TSE items and class-/task-specific TSE items. Their students ($N = 1930$) rated the quality of teaching in their math class. Multilevel regression analyses revealed stronger associations between student-rated teaching quality and class-/task-specific TSE than generalized task-specific TSE. We discuss possible reasons for these results and outline the potential benefits of using class-specific assessments for future TSE research.

1   Introduction

In research on teacher motivation self-efficacy is considered a key motivational characteristic of teachers, emphasizing the belief in their own ability to influence student engagement and learning, even when they encounter difficulties (Klassen & Chiu, 2010; Tschannen-Moran & Woolfolk Hoy, 2001). Teacher self-efficacy (TSE) has attracted attention in educational research in recent decades as an important contributor to outcomes such as teacher well-being, and student achievement and motivation (Caprara et al., 2006; Klassen et al., 2011). It is also assumed that teachers with high self-efficacy beliefs perceive themselves as more competent and confident in managing difficult situations in the classroom and this in turn leads to higher-quality teaching (Bandura, 1997; Tschannen-Moran et

39   al., 1998). Empirical findings on the relations between TSE and teaching quality, however, have not
40   been consistent, with studies finding both negative and positive relations between the two constructs
41   (see review by Lauermann & ten Hagen, 2021).

42   One reason for these inconsistent results could be the varied ways in which the studies conceptualize
43   and assess TSE; the studies differ with respect to their degree of context-specificity (Lauermann & ten
44   Hagen, 2021). Even though self-efficacy beliefs were originally conceptualized as context-specific
45   characteristics, meaning that they could fluctuate depending on the task or situation (Bandura, 1986,
46   1997), the vast majority of studies have treated TSE as a trait-like characteristic that can be generalized
47   across different teaching contexts (Zee et al., 2016). The different students and classes teachers teach
48   throughout the day are pivotal contextual factors that can contribute to different TSE ratings (Dellinger
49   et al., 2008). To date, only two empirical studies have investigated class-specific TSE evaluations.
50   These studies show that TSE varies considerably across different classes and that this intra-teacher
51   variance is correlated with class-specific characteristics (e.g., class size, achievement levels, and
52   student engagement; Raudenbush et al., 1992; Ross et al., 1996). Therefore, investigating TSE on a
53   general level fails to account for the context-specificity, particularly the class-specificity, of TSE.

54   It is likely that class-specificity would be particularly relevant for teachers who teach multiple classes
55   (e.g., secondary-level teachers). Recognizing the intra-teacher variance of self-efficacy beliefs across
56   different classes, several researchers have highlighted the need to assess TSE with reference to a
57   specific class (Lauermann & Berger, 2021; Lazarides et al., 2021; Raudenbush et al., 1992). Using
58   class-specific TSE scales should result in comparatively stronger associations with student-rated
59   indicators of teaching quality because individuals' self-efficacy beliefs are most accurate in predicting
60   corresponding behaviors when measured with a similar level of context specificity as their presumed
61   outcomes (Bandura, 2006; Chesnut & Burley, 2015; Pajares, 1996). Given that teaching quality, which
62   is also a context-specific characteristic, is usually assessed with reference to a specific class (Göllner
63   et al., 2018), class-specific evaluations of TSE could result in more consistent relations with teaching
64   quality than those of generalized TSE scales. A misalignment between generalized TSE and class-
65   specific teaching quality measures could be the reason for the inconsistent results.

66   A small number of studies have used class-specific adaptations of established TSE scales (Holzberger
67   et al., 2013; Perera & John, 2020), and these show significant positive associations with teaching
68   quality. However, no study to date has directly compared the predictive effect of TSE scales that use
69   different levels of context-specificity on student-rated teaching quality. A key objective of the present
70   study is to conduct comparative analyses of class-/task-specific versus generalized task-specific TSE
71   scales and their associations with teaching quality in the same sample.

72   In the subsequent sections, we present the conceptualization and presumed classroom implications of
73   TSE from the perspective of social cognitive theory. We then outline the context-specificity of existing
74   TSE measures and demonstrate why it is important to assess TSE with reference to a specific class.
75   Finally, we present the aims of the present study.

76   **2   Theoretical background**

77   **2.1   Teacher self-efficacy and how it relates to teaching quality**

78   Self-efficacy is a key motivational characteristic of teachers and describes the teacher's judgment of
79   their perceived ability to influence student engagement and learning, even in difficult situations
80   (Klassen & Chiu, 2010; Tschannen-Moran & Hoy, 2001). Research on TSE builds on social cognitive
81   theory (Bandura, 1997), which posits that an individual's behavior is influenced by the interplay of

82　personal, behavioral, and environmental factors. Specifically, teacher self-efficacy beliefs are shaped
83　by efficacy-building experiences such as mastery experiences (e.g., successful student achievement),
84　vicarious experiences (e.g., observation of a successful behavior of a colleague), verbal persuasion
85　(e.g., positive feedback from a colleague), and physiological activity (e.g., heart rate) (Fackler &
86　Malmberg, 2016). Self-efficacy beliefs not only influence performance, but also goal setting, effort,
87　and perseverance in attaining goals, which then represent new sources of information for an adapted
88　estimation of one's self-efficacy (Bandura, 1997). This interplay illustrates that self-efficacy beliefs do
89　not refer to actual competences but rather to the self-evaluated levels of competence.

90　TSE has garnered increased attention in research on teacher motivation in the last thirty years and
91　appears to be a an important factor in teacher development, teaching practice, and student outcomes
92　(see reviews by Klassen et al., 2011; Lauermann & ten Hagen, 2021). There is an assumption that
93　teachers with high levels of TSE are less likely to experience burnout and more likely to be satisfied
94　with their job (e.g., Skaalvik & Skaalvik, 2007). Studies have also found positive relations between
95　TSE and student achievement and motivation (e.g., Caprara et al., 2006). Research further suggests a
96　positive association between TSE and teachers' classroom behavior. Teachers with a high level of self-
97　efficacy tend to be harder working, more persistent in the face of obstacles, and capable of
98　implementing more challenging and innovative teaching methods (Klassen & Tze, 2014; Tschannen-
99　Moran et al., 1998).

100　In research on teaching, teaching quality is considered a key determinant of student learning
101　achievement (Hattie, 2009). Over the last decades, various frameworks have been developed to
102　describe pivotal characteristics of teaching quality ~~Across a number of different models~~ (for an
103　overview see ~~e.g., Klieme et al., 2001;~~ Praetorius & Charalambous,~~et al.,~~ 2018). Among others, the
104　Three Basic Dimensions of Teaching Quality (TBD) referring to three pivotal characteristics has
105　emerged as being especially useful for describing teaching quality: 1. *Classroom management* –
106　Maximizing students' time on task by coping effectively with disruptions and implementing clear rules
107　and routines. Through effective classroom management students are provided with disruption-free
108　learning opportunities that can be used for engaged learning processes and activities. Well-organized
109　classroom management environments therefore foster student learning (Seidel & Shavelson, 2007). 2.
110　*Cognitive activation* – encompasses discursive teaching and intensive higher-order thinking, by, for
111　example, providing complex tasks and encouraging problem solving. Cognitive activating teaching
112　aims for a deeper understanding of the learning content and a depth of processing and therefore
113　promotes students' learning and achievement (Lipowsky et al., 2009). 3. *Student support* – Fostering
114　positive and supportive relations between themselves and students, for example, by providing
115　constructive feedback and adopting a positive attitude towards student errors. A supportive classroom
116　climate fosters positive engagement and a feeling of social relatedness, competence, and autonomy,
117　which enhances student motivation (Rakoczy, 2008). For a detailed overview of the three basic
118　dimensions of teaching quality and their assumed effects, see Klieme et al. (2006; 2009). Teaching
119　quality dimensions have often been assessed using student ratings, as they are based on students' day-
120　to-day classroom experience. These ratings represent a valid, reliable, and cost-effective assessment
121　perspective (Clausen, 2002; Göllner et al., 2021; Praetorius et al., 2018).

122　Contrary to theoretical expectations, empirical ~~relations~~ findings on the relations between TSE and the
123　three basic dimensions of teaching quality (classroom management, cognitive activation, and student
124　support) are rather inconsistent across the existing studies ~~(~~see reviews by Lauermann & ten Hagen,
125　2021 and by~~;~~ Zee & Koomen, 2016). Hence, ~~s~~Studies, which find positive cross-sectional links
126　between TSE and student-rated dimensions of teaching quality, seem to be as common as studies that
127　show no significant relation. ~~In~~ For example, in studies by Burić and Kim (2020), Fauth et al. (2019),

128 and Ryan et al. (2015) significant positive <u>cross-sectional</u> links have been found between TSE and the
129 three basic teaching quality dimensions<u>.</u> <s>(classroom management, cognitive activation, and student</s>
130 <s>support).</s> However, others have not been able to find significant cross-sectional links between TSE and
131 student-rated teaching quality dimensions (e.g., Guo et al., 2012; Jamil et al., 2012). Also, the few
132 longitudinal studies found inconsistent relations between TSE and student-rated teaching quality<u>:</u>
133 <u>Whereas the study by Holzberger et al. (2013) found significant positive relations between TSE and</u>
134 <u>teaching quality dimensions, the two other existing longitudinal studies by Lazarides et al. (2013) and</u>
135 <u>Praetorius et al. (2017) found no significant longitudinal relations.</u> <s>(e.g., Holzberger et al., 2013;</s>
136 <s>Lazarides et al., 2021; Praetorius et al., 2017).</s> <u>The positive longitudinal effect of cognitive activation</u>
137 <u>and classroom management on student-rated TSE in the study of Holzberger et al. (2013) indicate that</u>
138 <u>TSE may not only be a predictor but also an outcome of high quality teaching.</u> Considering the
139 importance of teaching quality in research on educational effectiveness, it is important to establish a
140 better understanding of the empirical links between TSE and teaching quality. One reason for the
141 inconsistent findings <u>across the studies</u> could be the various conceptualizations and measurements of
142 TSE used by researchers, which differ with respect to their levels of context-specificity (Lauermann &
143 ten Hagen, 2021; Lazarides et al., 2021).

144 2.2 Context-specificity of TSE measures

145 The question of what constitute appropriate conceptualizations and measurements of TSE has been a
146 topic of debate for decades (Klassen et al., 2011). Over the years various conceptualizations and
147 measures have been developed, from general to more specific levels of TSE. Early empirical research
148 mostly treated TSE as a relatively stable, almost trait-like characteristic of teachers that indicated a
149 teacher's belief in their capabilities (Gibson & Dembo, 1984; Schwarzer et al., 1999). Researchers
150 following this theoretical stance thus treated within-teacher variance in TSE as error-variance (Zee et
151 al., 2016). Generalized measures are not tailored to the teaching process itself but relate to various
152 rather broad areas of teachers' work (e.g., social interactions with parents). Even though they have
153 commonly been used for studying TSE across different school grades and subjects from 1998 to 2009
154 (see Klassen et al., 2011), these unidimensional measures have been criticized for their lack of
155 predictive validity (Bandura, 1997). This is because the items are often formulated in <u>such a</u> way that
156 <s>does</s> <u>is</u> not <s>make</s> clear what precisely is being measured. For example, items such as "I can enforce
157 changes within the model project over skeptical colleagues" are ambiguous and fail to specify
158 contextual details. Such a general, undifferentiated, perspective seems particularly problematic as it
159 does not reflect the many facets of the complex nature of teaching that teachers face in their daily life
160 (Tschannen-Moran et al., 1998). General measures neglect the basic tenets of the social cognitive
161 theory on which TSE is based, which suggests that self-efficacy does not reflect a uniform stable-trait
162 characteristic of a person. Instead, TSE is context-specific since "some situations require greater skill
163 and more arduous performances, or carry greater risk of negative consequences, than others" (Bandura,
164 1986, p. 411). Bandura <u>(2006)</u> was therefore critical of "all-purpose" self-efficacy measures, as they
165 do not refer to particular tasks and situations (<u>p. 307</u><s>2006</s>).

166 In early research the context-specificity of TSE was largely ignored; general TSE ratings with little or
167 no connection to the relevant teaching task or situation were favored (Lazarides & Warner, 2020).
168 Recognizing the drawbacks of general measurements, later researchers started putting a stronger
169 emphasis on the context-specific nature of TSE and developing new measurements (Zee et al., 2016).
170 This resulted in a shift from general to task-specific conceptualizations of TSE. One of the most
171 prominent scales is the Teachers' Sense of Efficacy Scale (TSES) developed by Tschannen-Moran and
172 Woolfolk Hoy (2001). This scale <s>is</s> comprise<s>d of</s> three fundamental teaching-related tasks in a
173 teachers' daily life: TSE for classroom management, instructional strategies, and student engagement.

174   The assumption is that a teacher may feel efficacious about, for example, dealing with classroom
175   disruptions, while perceiving him/herself as less effective in building supportive relationships with
176   students. The TSES is applicable across different grades and school subjects (Klassen et al., 2009;
177   Tschannen-Moran & Woolfolk Hoy, 2007).

178   Even though the development of task-specific TSE measurements moved the field towards a more
179   valid approach for assessing the self-efficacy beliefs of teachers by tailoring their items toward specific
180   teaching-related tasks, the vast majority of studies on TSE still implicitly assume that TSE is
181   generalizable across different teaching situations (Dellinger et al., 2008). Researchers following
182   Bandura's notion that TSE is task- and situation-specific argue that TSE fluctuates not only across
183   teaching-related tasks but also across different teaching situations (e.g., Tschannen-Moran & Woolfolk
184   Hoy, 2001; Zee et al., 2016). Dellinger et al. (2008), for example, adhered to the idea that TSE
185   represents a "teacher's individual beliefs in their capabilities to perform specific teaching tasks at a
186   specified level of quality in a specified situation" (p. 752). Thus, the authors argued that a teacher might
187   experience different levels of self-efficacy across various teaching-related tasks and teaching situations
188   (specific schools, classrooms, students). A pivotal situational context that varies in teachers' daily work
189   is the different classes that they teach, as teachers deal with different kinds of environments and
190   challenges in each class (Raudenbush et al., 1992).

191   There are several reasons why assessing TSE not only via task-specific but also class-specific items,
192   such as the tailoring of TSE items to specific classes, could be productive. First, individual studies have
193   shown that between 21 % and 44 % of the total variation of teachers' self-efficacy beliefs reflect within-
194   teacher variation across classrooms (e.g., Raudenbush et al., 1992; Ross et al., 1996). Despite the
195   limitation of TSE being assessed with a single item in both studies, the findings confirm that teachers'
196   self-efficacy beliefs are not stable and generalizable across different teaching situations but vary across
197   different classrooms. Second, considering within-teacher variation of TSE across classes is particularly
198   important for research in secondary schools or high schools, where teachers usually have multiple
199   classes. Previous studies that examined TSE at this level used generalized measures for assessing TSE
200   (e.g., Burić & Kim, 2020; Künsting et al., 2016; Praetorius et al., 2017) and therefore failed to consider
201   the class-specificity of TSE. Thus, the evaluation of TSE might be ambiguous and open for
202   interpretation, since it is unclear which class is being referred to (Zee et al., 2016). A teacher might
203   answer the same item differently depending on whether they are thinking of a comparatively easy or
204   difficult class. With reference to the four main sources of TSE (see section 2.1), external norm criteria
205   such as past or present experiences with a particular class, contextual cues (e.g., classroom
206   characteristics), or references (e.g., class comparisons) might influence teachers when they are
207   reporting their level of self-efficacy towards a specific class (Zee et al., 2018). For example, a teacher
208   might interpret high student achievement in their class as a kind of mastery experience, indicating their
209   teaching success, which then might positively affect the nature of their self-efficacy beliefs (Fackler &
210   Malmberg, 2016). By contrast, the same teacher might assess TSE differently if the items are related
211   to a different class with which they experience frequent stress and frustration in class. It is therefore
212   important that TSE items refer to a specific class. Third, assessing TSE with class-specific instead of
213   generalized measures also seems beneficial in terms of its predictive validity, as self-efficacy scales
214   are deemed most predictive when measured ~~in~~ as context-specific ~~contextually specific a manner~~ as
215   possible (Bandura, 1997, 2006). A recent meta-analysis confirmed that generalized TSE measures
216   suffer from low predictive validity and fail to uncover relations with context-specific outcomes
217   (Chesnut & Burley, 2015). This study concurs with a recent review by Lauermann and ten Hagen
218   (2021) and indicates that context-specific TSE measures have a higher magnitude of relations with
219   contextualized outcomes than generalized measures. A misalignment of context level between
220   predictor and outcome might therefore have contributed to the inconsistency of the findings of studies

221 investigating the relation between TSE and teaching quality to date. While TSE is usually assessed in
222 general terms, items for teaching quality dimensions are mostly context-specific and tailored to a
223 specific class because teaching quality is considered to be a classroom-level phenomenon (Aditomo &
224 Köhler, 2020; Göllner et al., 2018). Therefore, assessing TSE on a class-specific level might increase
225 predictive validity and strengthen associations with class-specific teaching quality and several
226 researchers have recently called for a more context-specific assessment of TSE (e.g., Bandura, 2006;
227 Lazarides & Warner, 2020; Zee et al., 2018).

228 To the best of our knowledge, only two studies have included class-specific adaptations of established
229 and ad-hoc TSE scales (see review of Lauermann and ten Hagen, 2021) to study the relations between
230 teaching quality dimensions and TSE (Holzberger et al., 2013; Perera & John, 2020). The introductory
231 sentence of both of those self-efficacy questionnaires referred to a target class, aligning them to class-
232 specific teaching quality. The increased validity of such class-specific TSE measurements could have
233 contributed to the significant relations found in both studies between class-specific TSE and teaching
234 quality dimensions. By contrast, a study by Praetorius et al. (2017) that used the same TSE measures
235 as Holzberger et al. (2013), but without tailoring the instrument to a specific class, found that the
236 relations were not significant. These preliminary findings support the assumption that context-specific
237 judgments of TSE have higher predictive power for relations with contextualized outcomes (Chesnut
238 & Burley, 2015). Despite the growing literature and the call for more context-specific TSE measures,
239 no study has yet conducted a direct comparison of how different levels of context-specificity in TSE
240 relate to teaching quality. A direct comparison would enable, for the first time, an analysis of whether
241 class-specific TSE measures have advantages for assessing teaching quality. This might go some way
242 towards clarifying the findings of inconsistent relations between TSE and teaching quality.

2.3 The present study

244 Encouraged by the previous findings of context-specific TSE (Bandura, 2006; Chesnut & Burley,
245 2015), the present study aimed to investigate TSE not only in relation to a specific teaching-related
246 task, but also to a specific class. We extend the study by Holzberger et al. (2013) which referred to a
247 specific class but neglected the task-specificity of TSE as they used the general measure of Schwarzer
248 et al. (1999). We have incorporated the generalized task-specific TSE measure of Tschannen-Moran
249 and Hoy (2001), but also tailored the introductory sentence and all items to a specific class. This should
250 align the TSE measurement more closely to teaching quality. By directly comparing two TSE scales
251 with different levels of context-specificity (the generalized task-specific TSE scale vs. the adapted
252 class-/task-specific TSE scale), we also aim to explore their predictive validity. Specifically, the study
253 explores the following research questions:

254 1. How is class-/task-specific TSE related to the three basic dimensions of teaching quality?

255 2. How do the relations to teaching quality dimensions differ between class-/task-specific TSE and
256 generalized task-specific TSE?

257 Based on previous results, we expect that class-/task-specific TSE will be positively related to
258 classroom management [H1a], cognitive activation [H1b], and student support [H1c]. We also expect
259 that the relations of class-/task-specific TSE and classroom management [H2a], cognitive activation
260 [H2b], and student support [H2c] are significantly stronger than the ones with generalized task-specific
261 TSE.

3 Methods

6

263 **3.1   Participants and procedure**

264   Data was drawn from the Teaching and Learning International Survey (TALIS) Video Study. The main
265   data collection of the study was conducted in 2017 and 2018 (OECD, 2020a). The present study is
266   based on country-level data from England. The English data sample was selected because it included
267   the planned sample size of $N = 85$ classes and the instrument quality of the target scales was judged to
268   be sufficient (this was not the case for some of the other countries). In the final sample of this study,
269   ratings of $N = 86$ mathematics secondary teachers from 78 schools (all state-funded, 74% located in
270   urban areas) and their $N = 1930$ students were collected. All teachers taught the focal topic of quadratic
271   equations within the target year groups (year 8 to 11) with the majority (71%) of the students being in
272   school year 10. The mean number of students per class was 23.6 students ($SD = 6.50$). A total of 58 %
273   of the teachers were female and their average age was 35.7 years old ($SD = 8.40$) with an average work
274   experience of 9.9 years ($SD = 7.00$). Students were 14.8 years old ($SD = 0.61$) on average, with 54 %
275   of them being female. Study participation was voluntary for both teachers and students.

276

277 **3.2   Measures**

278 **3.2.1 Teacher self-efficacy**

279 *Generalized task-specific TSE*

280   Teachers were asked to rate their self-efficacy beliefs during teaching with a short version of the task-
281   specific TSES devised by Tschannen-Moran and Woolfolk Hoy (2001). The TALIS Video Study had
282   used a shortened version with six items of this TSES to keep the size of the questionnaire manageable
283   (for the items used, see the Appendix). The questionnaire included questions about teachers' self-
284   efficacy beliefs about key teaching tasks such as classroom management, instructional strategies, and
285   student engagement. The introductory stem was "In your teaching in general, to what extent can you
286   do the following?" and the six items were recorded on a Likert-type scale ranging from 1 (*not at all*)
287   to 4 (*a lot*). A sample item was "Help my students value learning". Cronbach's alpha for this scale was
288   .79.

289 *Class-/task-specific TSE*

290   A modified version of the task-specific TSES questionnaire that included a class-specific component
291   was also used. The introductory sentence and all items in it referred to a specific class: The introductory
292   stem of the class-/task-specific version was "In your teaching, to what extent can you do the following
293   in the target class?" and the six items were recorded on a Likert-type scale ranging from 1 (*not at all*)
294   to 4 (*a lot*). A sample item was "Help these students value learning". Cronbach's alpha for this scale
295   was .69.

296 **3.2.2 Teaching quality**

297   The students in each class rated teaching quality in mathematics based on classroom management,
298   cognitive activation, and student support (for the items used, see the Appendix). The ratings included
299   items that were adapted from PISA (2003, 2012). All items were rated on a 4-point Likert scale ranging
300   from 1 (*strongly disagree*) to 4 (*strongly agree*). Multilevel McDonald's omega indices reflect the
301   level-specific reliability of the teaching quality scales (see Geldhof et al., 2014).

302 *Classroom management* was assessed with a 10-item scale including items about routines, monitoring,
303 and disruptions, e.g., "When the lesson begins, our mathematics teacher has to wait quite a long time
304 for us to quieten down." Within-level ω was .76 and between-level ω was .99.

305 *Cognitive activation* was assessed with a 7-item scale including items about students' cognitive
306 engagement and participation in discourse, e.g., "Our mathematics teacher presents tasks for which
307 there is no obvious solution." Within-level ω was .71 and between-level ω was .87.

308 *Student support* was assessed with an 8-item scale including items about the student-teacher
309 relationship and teacher support, e.g., "My mathematics teacher makes me feel she/he really cares
310 about me". Within-level ω was .89 and between-level ω was .99.

3.3 Statistical analyses

312 *Multilevel path analyses*

313 MPLUS 8.6 was used for all analyses (Muthén & Muthén, 1998-2017), applying maximum likelihood
314 estimation with robust standard errors (MLR). Missing data was handled with full-information
315 maximum likelihood estimation (FIML), as missing data on all variables were below 5%.

316 A multi-level path analysis was conducted to account for the hierarchical structure of the data (students
317 nested within classrooms). The three dimensions of teaching quality were included as dependent
318 variables. Measures were based on student ratings, which were combined to manifest scale values per
319 student and subsequently decomposed into within-class level (level 1) and between-class level (level
320 2) variance components (for advantages of latent aggregation see Lüdtke et al., 2008). For the first
321 research question, On level 2, manifest z-standardized scales of both TSE scales were used on level 2
322 as predictor variables. Due to sample size constraints on level 2, we refrained from using latent
323 modeling of TSE and a doubly-latent operationalization of teaching quality dimensions and instead
324 used sum scores for the variables.

325 For the second research question, we used the MODEL CONSTRAINT option to create additional
326 difference parameters to compare the structural paths between the two different TSE scales and
327 teaching quality dimensions. To test the difference parameters against zero, the variances of both
328 predictors on level 2 had to be equal. In order to express the relations in the form of standardized
329 regression coefficients, both the predictors and the criteria were standardized. As Mplus does not
330 standardize the variables separately on both levels when using the DEFINE STANDARDIZE function,
331 both predictors and all three dependent variables were standardized on level 2 by means of a linear
332 transformation within Mplus (subtraction of the level 2 mean, division by the square root of the level
333 2 variance).

334 As our hypotheses are directional, one-tailed tests were used with a significance level of $p < .05$ (Cho
335 & Abe, 2013; Ruxton & Neuhäuser, 2010). The final model was fully saturated; model fit was therefore
336 trivially perfect.

4 Results

4.1 Descriptive statistics and correlations

339 Table 1 reports the means, standard deviations, and intercorrelations for the variables on level 2 (latent
340 mean aggregation of student-rated teaching quality) along with the intraclass correlations (ICC1) and
341 the reliability of the class-aggregated scores (ICC2) for the three teaching quality dimensions.

342 Results showed that the two TSE scales were highly correlated. The three teaching quality dimensions
343 were also highly intercorrelated. ICC(1) values for the student-rated teaching quality dimensions
344 ranged from .16 to .36, indicating that between 16 % and 36 % of the total variance occurred due to
345 systematic between-class differences, supporting the decision to use multilevel analysis. ICC2 values,
346 which show the degree of consistency in students' ratings within a class, indicated a high consistency
347 across all three teaching quality dimensions (see Table 1).

348 Class-/task-specific TSE was positively associated with classroom management and student support,
349 whereas generalized task-specific TSE was unrelated to all three teaching quality dimensions.

350 [PLEASE INCLUDE TABLE 1 HERE]

4.2   Multilevel path analyses

352 The cross-sectional structural paths between class-/task-specific TSE and teaching quality dimensions
353 were tested in a multilevel path analysis. In line with Hypothesis 1a and 1c, class-/task-specific TSE
354 was significantly positively related with classroom management and student support (see Table 2 and
355 Figure 1). The relation between class-/task-specific TSE and cognitive activation, however, was not
356 significant. Thus, Hypothesis 1b was not confirmed.

357 As a next step, we compared the cross-sectional structural paths between class-/task-specific TSE and
358 teaching quality dimensions to the ones with generalized task-specific TSE. From a descriptive
359 perspective, greater positive relations were found between class-/task-specific TSE and classroom
360 management and student support compared to generalized task-specific TSE. This descriptive pattern
361 was not found for cognitive activation, as both TSE scales were unrelated to cognitive activation.
362 Despite this, none of the three pairs of structural paths between both sets of TSE measurements and
363 the teaching quality dimensions differed significantly, as indicated by their corresponding difference
364 parameters (see Table 2). Thus, Hypotheses 2a-c were not confirmed.

365 [PLEASE INCLUDE TABLE 2 HERE]

366 [PLEASE INCLUDE FIGURE 1 HERE]

5   Discussion

368 As previous findings do not provide a clear indication of whether TSE is associated with teaching
369 quality, we aimed to investigate whether a class-specific perspective on TSE, rather than a generalized
370 one, might yield a clearer picture. We followed the often-neglected assumption of social cognitive
371 theory that suggests that TSE measures are not only task- but also situation-specific (e.g., class-
372 specific) and most predictive when they are aligned with the behavioral outcome (Bandura, 1997,
373 2006).

5.1   Relations between class-/task-specific TSE and teaching quality

375 With our first research question, we investigated the relations between class-/task-specific TSE and
376 teaching quality. Our analyses revealed significant positive cross-sectional relations between class-
377 /task-specific TSE and student-rated classroom management and student support. When teachers felt

378 confident in their teaching capabilities, students rated their teaching quality as higher, resulting in better
379 classroom management and student support. This corroborates with the two existing studies on class-
380 specific adaptions of TSE scales (Holzberger et al., 2013; Perera & John, 2020).

381 Interestingly, no significant relation was found between class-/task-specific TSE and cognitive
382 activation. This finding might be attributed to the fact that cognitive activation represents a complex
383 and high inference characteristic of teaching quality that requires a higher level of idiosyncratic
384 interpretation and is more difficult to observe than classroom management and student support (see
385 e.g., Praetorius et al., 2014). This usually results in a lower agreement between student evaluations, as
386 shown by low ICC values (see e.g., Fauth et al., 2020; Kunter et al., 2008; Thommen et al., 2021),
387 which is also true in this study, (see Table 1) and in lower teacher-student agreement (see e.g.,
388 Wisniewski et al., 2020) than for classroom management and student support. Students seem to find it
389 more difficult to evaluate cognitive activating teaching reliably. This might explain why the
390 associations between TSE and teaching quality are usually greater and more consistent when teachers
391 instead of students assess their teaching (Lauermann & ten Hagen, 2021). For example, Schiefele and
392 Schaffner (2015) found significant positive relations between TSE and teacher-rated cognitive
393 activation, but none with student-rated cognitive activation. Only a few studies have investigated the
394 relation between TSE and teaching quality from different rater perspectives. We recommend that future
395 studies investigate teaching quality from different perspectives including, for example, external
396 observer ratings as they are deemed promising (Clausen, 2020). Apart from that, the various
397 conceptualizations and operationalizations of cognitive activation used in previous studies could have
398 also contributed to the inconsistent research findings on the relations between TSE and cognitive
399 activation. In our study, cognitive activation was assessed by two core subdimensions discursive
400 teaching and support of higher-order thinking. However, there are various other approaches to
401 measuring cognitive activation (see Praetorius & Charalambous, 2018). Developing a shared
402 understanding of these constructs and their measurement in the research community would benefit the
403 aim of cumulative research on teaching (see Charalambous et al., 2021).

404 The absence of significant relations between TSE and cognitive activation might also be attributed to
405 validity issues: The original TSES is assumed to be conceptually close to teaching quality dimensions
406 as their underlying sub-dimensions refer to crucial teaching-related tasks (Tschannen-Moran &
407 Woolfolk Hoy, 2001). However, a close analysis of the items of the three sub-dimensions of TSES,
408 shows that only TSE for classroom management ("To what extent can you do the following: […]
409 control disruptive behavior in this classroom") and TSE for student engagement (" […] get students in
410 this class to believe they can do well in school work") include aspects similar to the basic teaching
411 quality dimensions of classroom management and student support. By contrast, items of the sub-
412 dimension TSE for instructional strategies such as "[…] use a variety of assessment strategies in this
413 class" relate more strongly to the adaptability and flexibility of a teacher than to cognitive activating
414 teaching. This potential threat to validity caused by a content-related misalignment might therefore
415 have contributed to the absence of a significant relation between the two constructs.

416 Lastly, it might be that teachers' self-efficacy beliefs only have an indirect predictive effect on (student-
417 rated) cognitive activation. As recently discussed in the review by Lauermann and ten Hagen (2021),
418 it might be that the effect of TSE on student-rated cognitive activation is mediated by teachers' levels
419 of effort and persistence and classroom processes (e.g., mastery-oriented instructional practices).
420 However, available evidence on direct and indirect effects is scarce and needs further investigation.

421 **5.2   Comparison of the different levels of context-specificity of TSE scales and their relations**
422 **with teaching quality**

423 With our second research question, we aimed to compare two different context-specific levels of TSE
424 scales directly to get further insight into whether a class- and task-specific TSE scale could be useful
425 for examining the relation with dimensions of teaching quality.

426 Our findings indicate stronger relations between class-/task-specific TSE and teaching quality than that
427 with generalized task-specific TSE. Significant positive relations between class-/task-specific TSE and
428 classroom management and student support were found. In contrast, no significant relations between
429 teaching quality and generalized task-specific TSE were found.

430 The difference parameters between the two TSE measures were not statistically significant (see Table
431 2). However, a non-significant *p*-value should be interpreted carefully as it does not indicate whether
432 there is an actual absence of an effect or possibly a Type II error (see e.g., Edelsbrunner & Thurn,
433 2020; Mehler et al., 2019). It may be that the *p*-values > .05 stem from the rather small sample size on
434 level 2 and high standard errors with limited power to find statistically significant effects. To verify if
435 the sample was indeed too small to find significant effects, a power analysis would be appropriate.
436 However, as post-hoc power analyses are conducted on the basis of sample-based mean differences
437 and conceptually flawed, several researchers advise against conducting such analyses in retrospect (see
438 also Zhang et al., 2019). Thus, future studies should consider *apriori* power analyses to get information
439 about sample sizes needed to detect statistically significant effects.

440 Another explanation for the non-significant difference parameters might be that the rather low
441 reliability (α = .69) of the class-/task-specific TSE might have influenced our findings to some extent.
442 The low reliability stems from the shortened version of the original TSES with only six instead of 12
443 or 24 items. Future studies should preferably use the original scale to ensure higher reliability.

444 Taken together, the non-significant difference parameters in this study do not yield conclusive
445 information on the added value of class-/task-specific TSE compared to generalized task-specific TSE
446 when examining the relation with teaching quality. Our preliminary findings should therefore be
447 interpreted carefully. Despite the non-significant difference parameters, this study indicates that it
448 makes a difference whether a teacher is asked about his/her self-efficacy beliefs in general or their TSE
449 with reference to a specific class. Both TSE scales seem to be highly correlated (see Figure 1), but
450 there seems still enough within-teacher variance that could be explained by contextual factors such as
451 classroom characteristics. This seems in line with the findings of Raudenbush et al. (1992) and Ross et
452 al. (1996) and suggests that teachers' self-efficacy beliefs should not be treated as generalizable across
453 different classrooms. Assessing TSE with reference to a specific class seems especially important for
454 research in secondary schools or high schools, where a teacher usually teaches more than one class at
455 a time and generalized TSE measures would not indicate which class is being referred to. As our study
456 is the first to specifically investigate different levels of context-specificity of TSE and their associations
457 with teaching quality, further research is needed. As all teachers in our study were only assessed with
458 respect to teaching one particular class, the possibility of a variance decomposition (ICC values) for
459 TSE is not given. It might be interesting for future studies to investigate whether differences in the self-
460 efficacy of a teacher can be identified between different classes.~~It~~ Moreover, it might be interesting to
461 examine which classroom characteristics (e.g., class size, number of students with special educational
462 needs, achievement level, achievement related heterogeneity) best explain the within-teacher variance
463 of TSE (see e.g., Raudenbush et al., 1992).

5.3   Limitations

465 The following limitations should be considered when interpreting our findings.

466 First, our analyses are based on cross-sectional data, which cannot be used to infer causality.[1] This
467 study was based on the theoretical assumption, drawn from prior studies, that higher self-efficacy
468 beliefs lead to higher teaching quality (Perera & John, 2020). However, from the point of view of social
469 cognitive theory, the relations between the two constructs are reciprocal. As shown by Holzberger et
470 al. (2013), a well-functioning classroom can be interpreted by a teacher as an indicator of achievement
471 and serve as a source of mastery experience, influencing future self-efficacy beliefs. Future studies
472 should therefore use longitudinal data with multiple measurement points to provide clearer information
473 on causal effects between TSE and teaching quality.

474 Second, the English sample of the TALIS Video Study is not considered representative of the national
475 population of schools, teachers, or students since voluntary participation led to selective sampling and
476 the number of schools was rather small (OECD, 2020a). The relatively small teacher sample might
477 have led to an underestimation of the variance of TSE and teaching quality and, therefore, of the
478 relations between them. Future studies should replicate our findings with larger samples, to be able to
479 make general conclusions on the added value of a class-specific TSE assessment.

480 Third, we have examined TSE based on self-assessments because they are best placed to report on their
481 belief in their abilities. However, when interpreting the rather high mean TSE values in our study,
482 methodological biases such as self-desirability or faking should be considered when using self-reports
483 of teacher motivation (Bardach et al., 2021). ~~These~~ Following these authors, ~~therefore suggest~~ using
484 complementary measures such as situational judgment tests for TSE evaluations should be considered
485 in future studies.

486 Lastly, because the shortened version of the TSES had only two items per sub-dimension, in our
487 analyses we used the total TSE scores to examine the relations between TSE and the dimensions of
488 teaching quality. However, as discussed in Section 5.1, stronger relations are expected when predictor
489 and outcome refer to the same entity. There is some evidence to suggest that assessing the relation
490 between matched sub-dimensions of TSE and teaching quality, for example, between TSE for
491 classroom management and student-perceived classroom management is promising (Lazarides et al.,
492 2020). Future research needs to validate our findings with the original version of the TSES and could
493 examine the relations of matched sub-dimensions of TSE and teaching quality separately.

494 **6   Conclusion**

495 By adopting a class-specific perspective on TSE, our study aimed to clarify why research findings on
496 the relations between TSE and teaching quality have been inconsistent. Our results suggest significant
497 positive associations between class-/task-specific TSE and student-rated teaching quality. This study
498 is also the first to directly compare different context-specific levels of TSE and their relations with
499 teaching quality. Our results do not provide conclusive information about the added value of the class-
500 /task-specific TSE compared to the generalized task-specific TSE scale. However, based on the
501 descriptive results, it seems promising to continue assessing TSE from a class-specific perspective and
502 replicate our findings with a larger sample. We believe that more consistent use of context-specific

---

12

[1] It should be noted that the TALIS Video Study is based on a longitudinal data structure. However, we decided to use teaching quality ratings from the pre-questionnaire for our analyses, as in England the time interval between pre- and post data collection was rather short (only around two weeks; for further information see Ingram et al., 2020). This resulted in very high stabilities for the teaching quality dimensions, which indicate that there was very little time for changes in teaching quality to happen.

503 TSE scales, as suggested by Bandura's social cognitive theory, would also help synthesize future
504 research findings.

505 **7   Declaration of conflicts of interest:**

506 The authors declare that the research was conducted in the absence of any commercial or financial
507 relationships that could be construed as a potential conflict of interest.

508 **8   Author contributions:**

509 DT and AKP conceived and designed the study. UG and FL provided advice on the study design and
510 analyses. DT carried out the main data analysis and wrote the first draft of the manuscript. UG verified
511 the analyses. All authors discussed the results and were involved in writing the final manuscript. AKP
512 supervised the project.

513 **Acknowledgement:**

518 **Data availability statement:**

519 The original data reported in the study are publicly available. They can be found here:
520 [https://www.oecd.org/education/school/global-teaching-insights-technical-documents.htm]

521 **References**

522 Aditomo, A., & Köhler, C. (2020). Do student ratings provide reliable and valid information about teaching
523   quality at the school level? Evaluating measures of science teaching in PISA 2015. *Educational*
524   *Assessment, Evaluation and Accountability*, *32*(3), 275–310. https://doi.org/10.1007/s11092-020-
525   09328-6
526 Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. *Prentice-Hall series*
527   *in social learning theory*. Prentice-Hall.
528 Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman and Company.
529 Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-Efficacy Beliefs of Adolescents*, *5*(1),
530   307–337.
531 Bardach, L., Klassen, R., & Perry, N. E. (2021). Teachers' psychological characteristics: Do they matter for
532   teacher effectiveness, Teachers' well- being, retention, and interpersonal relations? An integrative
533   review. *Educational Psychology Review*.
534   https://www.researchgate.net/publication/349761009_Teachers%27_Psychological_Characteristics_D
535   o_They_Matter_for_Teacher_Effectiveness_Teachers%27_Well-
536   _being_Retention_and_Interpersonal_Relations_An_Integrative_Review
537 Burić, I., & Kim, L. E. (2020). Teacher self-efficacy, instructional quality, and student motivational beliefs:
538   An analysis using multilevel structural equation modeling. *Learning and Instruction*, *66*, 1–12.
539   https://doi.org/10.1016/j.learninstruc.2019.101302
540 Charalambous, C. Y., Praetorius, A.-K., Sammons, P., Walkowiak, T., Jentsch, A., & Kyriakidēs, L. (2021).
541   Working more collaboratively to better understand teaching and its quality: Challenges faced and

possible solutions. *Studies in Educational Evaluation, 71.*
https://doi.org/10.1016/j.stueduc.2021.101092

Caprara, G. V., Barbaranelli, C., Steca, P., & Malone, P. S. (2006). Teachers' self-efficacy beliefs as
determinants of job satisfaction and students' academic achievement: A study at the school level.
Journal of School Psychology, *44*(6), 473–490. https://doi.org/10.1016/j.jsp.2006.09.001

Chesnut, S. R., & Burley, H. (2015). Self-efficacy as a predictor of commitment to the teaching profession:
A meta-analysis. *Educational Research Review*, *15*, 1–16.
https://doi.org/10.1016/j.edurev.2015.02.001

Cho, H.-C., & Abe, S. (2013). Is two-tailed testing for directional research hypotheses tests legitimate?
Journal of Business Research, *66*(9), 1261–1266. https://doi.org/10.1016/j.jbusres.2012.02.023

Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive? Empirische Analysen zur
Übereinstimmung, Konstrukt- und Kriteriumsvalidität* [*Teaching quality: A question of perspective?
Empirical analyses of agreement, construct and criterion validity]. Pädagogische Psychologie und
Entwicklungspsychologie. Vol. 29*. Waxmann.

Clausen, M. (2020). Commentary regarding the section "The role of different perspectives on the
measurement of teaching quality". *Zeitschrift Für Pädagogik*, *66*(66), 173–178.

Dellinger, A. B., Bobbett, J. J., Olivier, D. F., & Ellett, C. D. (2008). Measuring teachers' self-efficacy
beliefs: Development and use of the TEBS-Self. *Teaching and Teacher Education*, *24*(3), 751–766.
https://doi.org/10.1016/j.tate.2007.02.010

Edelsbrunner, P. A., & Thurn, C. M. (2020). *Improving the utility of non-significant results for educational
research*. PsyArXiv [Preprint]. Available at: https://doi.org/10.31234/osf.io/j93a2.

Fackler, S., & Malmberg, L.-E. (2016). Teachers' self-efficacy in 14 OECD countries: Teacher, student
group, school and leadership effects. *Teaching and Teacher Education*, *56*, 185–195.
https://doi.org/10.1016/j.tate.2016.03.002

Fauth, B., Decristan, J., Decker, A.-T., Büttner, G., Hardy, I., Klieme, E., & Kunter, M. (2019). The effects
of teacher competence on student outcomes in elementary science education: The mediating role of
teaching quality. *Teaching and Teacher Education*, *86.* https://doi.org/10.1016/j.tate.2019.102882

Fauth, B., Wagner, W., Bertram, C., Göllner, R., Roloff, J., Lüdtke, O., Polikoff, M. S., Klusmann, U., &
Trautwein, U. (2020). Don't blame the teacher? The need to account for classroom characteristics in
evaluations of teaching quality. *Journal of Educational Psychology*, *112*(6), 1284–1302.
https://doi.org/10.1037/edu0000416

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory
factor analysis framework. *Psychological Methods*, *19*(1), 72–91. https://doi.org/10.1037/a0032138

Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational
Psychology*, *76*(4), 569–582. https://doi.org/10.1037/0022-0663.76.4.569

Göllner, R., Fauth, B., & Wagner, W. (2021). Student Ratings of Teaching Quality Dimensions: Empirical
Findings and Future Directions. In W. Rollett, H. Bijlsma, & S. Röhl (Eds.), *Student Feedback on
Teaching in Schools: Using Student Perceptions for the Development of Teaching and Teachers*
(pp. 111–122). Springer International Publishing. https://doi.org/10.1007/978-3-030-75150-0_7

Göllner, R., Wagner, W., Eccles, J. S., & Trautwein, U. (2018). Students' idiosyncratic perceptions of
teaching quality in mathematics: A result of rater tendency alone or an expression of dyadic effects
between students and teachers? *Journal of Educational Psychology*, *110*(5), 709–725.
https://doi.org/10.1037/edu0000236

Guo, Y., Connor, C. M., Yang, Y., Roehrig, A. D., & Morrison, F. J. (2012). The effects of teacher
qualification, teacher self-efficacy, and classroom practices on fifth graders' literacy outcomes. *The
Elementary School Journal*, *113*(1), 3–24. https://doi.org/10.1086/665816

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge
Taylor & Francis Group.

590 Holzberger, D., Philipp, A., & Kunter, M. (2013). How teachers' self-efficacy is related to instructional
591   quality: A longitudinal analysis. *Journal of Educational Psychology*, *105*(3), 774–786.
592   https://doi.org/10.1037/a0032198

593 Ingram, J., Lindorff, A., McCann, E., Riggall, A., & Sani, N. (2020). TALIS Video Study national report.
594   London, United Kingdom: DfE.

595 Jamil, F. M., Downer, J. T., & Pianta, R. C. (2012). Association of pre-service teachers' performance,
596   personality, and beliefs with teacher self-efficacy at program completion. *Teacher Education*
597   *Quarterly*, 119–138.

598 Klassen, R. M., Bong, M., Usher, E. L., Chong, W. H., Huan, V. S., Wong, I. Y., & Georgiou, T. (2009).
599   Exploring the validity of a teachers' self-efficacy scale in five countries. *Contemporary Educational*
600   *Psychology*, *34*(1), 67–76. https://doi.org/10.1016/j.cedpsych.2008.08.001

601 Klassen, R. M., & Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction: Teacher
602   gender, years of experience, and job stress. *Journal of Educational Psychology*, *102*(3), 741–756.
603   https://doi.org/10.1037/a0019237

604 Klassen, R. M., & Tze, V. M. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A
605   meta-analysis. *Educational Research Review*, *12*, 59–76. https://doi.org/10.1016/j.edurev.2014.06.001

606 Klassen, R. M., Tze, V. M. C., Betts, S. M., & Gordon, K. A. (2011). Teacher efficacy research 1998–2009:
607   Signs of progress or unfulfilled promise? *Educational Psychology Review*, *23*(1), 21–43.
608   https://doi.org/10.1007/s10648-010-9141-8

609 Klieme, E., Lipowsky, F., Rakoczy, K., & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von
610   Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts
611   "Pythagoras" [Quality dimensions and effectiveness of mathematics education. Theoretical
612   foundations and selected results of the "Pythagoras" project]. In M. Prenzel & L. Allolio-Näcke (Eds.),
613   *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms*
614   (pp. 127–146).

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

618 Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I:
619   "Aufgabenkultur" und Unterrichtsgestaltung [Mathematics teaching in lower secondary school: "task
620   culture" and lesson design]. In E. Klieme & J. Baumert (Eds.), *TIMSS - Impulse für Schule und*
621   *Unterricht: Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente* (pp. 43–57).
622   Bundesministerium für Bildung und Forschung.

623 Künsting, J., Neuber, V., & Lipowsky, F. (2016). Teacher self-efficacy as a long-term predictor of
624   instructional quality in the classroom. *European Journal of Psychology of Education*, *31*(3), 299–322.
625   https://doi.org/10.1007/s10212-015-0272-7

626 Kunter, M., Tsai, Y.-M., Klusmann, U., Brunner, M., Krauss, S., & Baumert, J. (2008). Students' and
627   mathematics teachers' perceptions of teacher enthusiasm and instruction. *Learning and Instruction*,
628   *18*(5), 468–482. https://doi.org/10.1016/j.learninstruc.2008.06.008

629 Lauermann, F., & Berger, J.-L. (2021). Linking teacher self-efficacy and responsibility with teachers' self-
630   reported and student-reported motivating styles and student engagement. *Learning and Instruction*,
631   101441. https://doi.org/10.1016/j.learninstruc.2020.101441

632 Lauermann, F., & ten Hagen, I. (2021). Do teachers' perceived teaching competence and self-efficacy affect
633   students' academic outcomes? A closer look at student-reported classroom processes and outcomes.
634   *Educational Psychologist*, 1–18. https://doi.org/10.1080/00461520.2021.1991355

635 Lazarides, R., Fauth, B., Gaspard, H., & Göllner, R. (2021). Teacher self-efficacy and enthusiasm: Relations
636   to changes in student-perceived teaching quality at the beginning of secondary education. *Learning*
637   *and Instruction*, *73*, 1–10. https://doi.org/10.1016/j.learninstruc.2020.101435

638   Lazarides, R., & Warner, L. M. (2020). Teacher Self-Efficacy. In *Oxford Research Encyclopedia of*
639   *Education.* https://doi.org/10.1093/acrefore/9780190264093.013.890
640   Lazarides, R., Watt, H. M., & Richardson, P. W. (2020). Teachers' classroom management self-efficacy,
641   perceived classroom management and teaching contexts from beginning until mid-career. *Learning*
642   *and Instruction*, *69*, 101346. https://doi.org/10.1016/j.learninstruc.2020.101346
643   Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of
644   geometry instruction and its short-term impact on students' understanding of the Pythagorean
645   Theorem. *Learning and Instruction*, *19*(6), 527–537.
646   https://doi.org/10.1016/j.learninstruc.2008.11.001
647   Mehler, D. M. A., Edelsbrunner, P. A., & Matić, K. (2019). Appreciating the significance of non-significant
648   findings in psychology. *Journal of European Psychology Students*, *10*(4), 1–7.
649   https://doi.org/10.5334/e2019a
650   Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén &
651   Muthén.
652   OECD (2020a). Global Teaching InSights: A Video Study of Teaching. Paris: OECD
653   Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*(4), 543–
654   578.
655   Perera, H. N., & John, J. E. (2020). Teachers' self-efficacy beliefs for teaching math: Relations with teacher
656   and student outcomes. *Contemporary Educational Psychology*, *61*.
657   https://doi.org/10.1016/j.cedpsych.2020.101842
658   Praetorius, A.-K., & Charalambous, C. Y. (2018). Classroom observation frameworks for studying
659   instructional quality: looking back and looking forward. *ZDM*, *50*(3), 535–553.
660   https://doi.org/10.1007/s11858-018-0946-0
661   Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The
662   German framework of three basic dimensions. *ZDM*, *50*(3), 407–426. https://doi.org/10.1007/s11858-
663   018-0918-4
664   Praetorius, A.-K., Lauermann, F., Klassen, R. M., Dickhäuser, O., Janke, S., & Dresel, M. (2017).
665   Longitudinal relations between teaching-related motivations and student-reported teaching quality.
666   *Teaching and Teacher Education*, *65*, 241–254. https://doi.org/10.1016/j.tate.2017.03.023
667   Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need?
668   Stability of instructional quality across lessons. *Learning and Instruction*, *31*, 2–12.
669   https://doi.org/10.1016/j.learninstruc.2013.12.002
670   Rakoczy, K. (2008). *Motivationsunterstützung im Mathematikunterricht: Unterricht aus der Perspektive von*
671   *Lernenden und Beobachtern* [Motivational support in mathematics education: teaching from the
672   perspective of learners and observers]. *Pädagogische Psychologie und Entwicklungspsychologie: Vol.*
673   *65*. Waxmann. http://deposit.d-nb.de/cgi-
674   bin/dokserv?id=3045550&prov=M&dok_var=1&dok_ext=htm
675   Raudenbush, S. W., Rowan, B., & Cheong, Y. F. (1992). Contextual effects on the self-perceived efficacy of
676   high school teachers. *Sociology of Education*, *65*(2), 150–167. https://doi.org/10.2307/2112680
677   Ross, J. A., Cousins, J. B., & Gadalla, T. (1996). Within-teacher predictors of teacher efficacy. *Teaching and*
678   *Teacher Education*, *12*(4), 385–400. https://doi.org/10.1016/0742-051X(95)00046-M
679   Ruxton, G. D., & Neuhäuser, M. (2010). When should we use one-tailed hypothesis testing? *Methods in*
680   *Ecology and Evolution*, *1*(2), 114–117. https://doi.org/10.1111/j.2041-210X.2010.00014.x
681   Ryan, A. M., Kuusinen, C. M., & Bedoya-Skoog, A. (2015). Managing peer relations: A dimension of
682   teacher self-efficacy that varies between elementary and middle school teachers and is associated with
683   observed classroom quality. *Contemporary Educational Psychology*, *41*, 147–156.
684   https://doi.org/10.1016/j.cedpsych.2015.01.002

685 Schiefele, U., & Schaffner, E. (2015). Teacher interests, mastery goals, and self-efficacy as predictors of
686   instructional practices and student motivation. *Contemporary Educational Psychology*, *42*, 159–171.
687   https://doi.org/10.1016/j.cedpsych.2015.06.005
688 Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory
689   and research design in disentangling meta-analysis results. *Review of Educational Research*, *77*(4),
690   454–499. https://doi.org/10.3102/0034654307310317
691 Schwarzer, R., Schmitz, G. S., & Daytner, G. T. (1999). *The teacher self-efficacy scale*. Available at:
692   www.ralfschwarzer.de
693 Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors,
694   perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology*, *99*(3),
695   611–625. https://doi.org/10.1037/0022-0663.99.3.611
696 Thommen, D., Sieber, V., Grob, U., & Praetorius, A.-K. (2021). Teachers' motivational profiles and their
697   longitudinal associations with teaching quality. *Learning and Instruction*, 101514.
698   https://doi.org/10.1016/j.learninstruc.2021.101514
699 Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct.
700   *Teaching and Teacher Education*(17), 783–805.
701 Tschannen-Moran, M., & Woolfolk Hoy, A. (2007). The differential antecedents of self-efficacy beliefs of
702   novice and experienced teachers. *Teaching and Teacher Education*, *23*(6), 944–956.
703   https://doi.org/10.1016/j.tate.2006.05.003
704 Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and
705   measure. *Review of Educational Research*, *68*(2), 202–248.
706   https://doi.org/10.3102/00346543068002202
707 Wisniewski, B., Zierer, K., Dresel, M., & Daumiller, M. (2020). Obtaining secondary students' perceptions
708   of instructional quality: Two-level structure and measurement invariance. *Learning and Instruction*,
709   *66*, 101303. https://doi.org/10.1016/j.learninstruc.2020.101303
710 Zhang, Y., Hedo, R., Rivera, A., Rull, R., Richardson, S., & Tu, X. M. (2019). Post hoc power analysis: Is it
711   an informative and meaningful analysis? *General Psychiatry*, 32(4), e100069.
712   https://doi.org/10.1136/gpsych-2019-100069
713 Zee, M., & Koomen, H. M. Y. (2016). Teacher self-efficacy and its effects on classroom processes, student
714   academic adjustment, and teacher well-being. *Review of Educational Research*, *86*(4), 981–1015.
715   https://doi.org/10.3102/0034654315626801
716 Zee, M., Koomen, H. M. Y., Jellesma, F. C., Geerlings, J., & Jong, P. F. de (2016). Inter- and intra-
717   individual differences in teachers' self-efficacy: A multilevel factor exploration. *Journal of School*
718   *Psychology*, *55*, 39–56. https://doi.org/10.1016/j.jsp.2015.12.003
719 Zee, M., Koomen, H. M. Y., & Jong, P. F. de (2018). How different levels of conceptualization and
720   measurement affect the relationship between teacher self-efficacy and students' academic
721   achievement. *Contemporary Educational Psychology*, *55*, 189–200.
722

723   **Tables**

724   Table 1. Descriptive statistics and intercorrelations of the study variables on level 2

	(1)	(2)	(3)	(4)	(5)
(1) Generalized task-specific TSE	3.44 (0.42)				
(2) Class-/task-specific TSE	.50**	3.41 (0.39)			
(3) Classroom management ICC(1) = 0.36, ICC(2) = 0.93	.12	.26*	2.95 (0.28)		
(4) Cognitive activation ICC(1) = 0.16, ICC(2) = 0.82	.12	.10	.62**	2.81 (0.22)	
(5) Student support ICC(1) = 0.24, ICC(2) = 0.88	.08	.21*	.58**	.63**	3.15 (0.26)

725   *Note.* Mean values and standard deviations of the variables are presented on the diagonal. $* p < .05; ** p < .01$ (two-tailed).

726

727 Table 2. Multilevel path analysis to estimate the associations of the two TSE scales and teaching quality

728 dimensions

	Classroom management		Cognitive activation		Student support	
	β (*SE*)	*p*-value	β (*SE*)	*p*-value	β (*SE*)	*p*-value
Class-/task-specific TSE	.26* (0.14)	. 04	.06 (0.15)	.35	.23* (0.12)	.03
Generalized task-specific TSE	-.004 (0.14)	. 49	.09 (0.13)	.25	-.03 (0.10)	.37
Difference parameters	.26 (0.25)	.14	-.03 (0.25)	.46	.26 (0.19)	.08

729 *Note*. Standardized coefficients for the reported relations were estimated. *p < .05 (one-tailed).

730

731 **Figure legends**

732 *Figure 1*. Multilevel path model predicting teaching quality by generalized task-specific and class-/task-specific TSE. Standardized
733 regression coefficients. Saturated model. *p < .05 (one-tailed).

734

735 **Appendix: Scale documentation**

736 **Generalized task-specific TSE**

737 In your teaching in general, to what extent can you do the following?

738 1. Get students to believe they can do well in school work.
739 2. Help my student's value learning.
740 3. Craft good questions for my students.
741 4. Control disruptive behavior in the classroom.
742 5. Get students to follow classroom rules.
743 6. Provide an alternative explanation for examples when students are confused.

744 **Class-/task-specific TSE**

745 In your teaching, to what extent can you do the following in the <target class>?

746 1. Get students in this <class> to believe they can do well in school work.
747 2. Help these students' value learning.
748 3. Craft good questions for these students.
749 4. Control disruptive behavior in this classroom.
750 5. Get students in this <class> to follow classroom rules.
751 6. Provide an alternative explanation for examples in this <class> when students are confused.

752 **Classroom management**

753 To what extent do you disagree or agree with the following statements?

1. When the lesson begins, our mathematics teacher has to wait quite a long time for us to quieten down.
2. We lose quite a lot of time because of students interrupting the lesson.
3. There is much disruptive noise in this classroom.
4. In our teacher's <class>, we are aware of what is allowed and what is not allowed.
5. In our teacher's <class>, we know why certain rules are important.
6. Our teacher manages to stop disruptions quickly.
7. Our teacher reacts to disruptions in such a way that the students stop disturbing learning.
8. In our teacher's <class>, transitions from one phase of the lesson to the other (e.g., from <class> discussions to individual work) take a lot of time.
9. Our teacher is immediately aware of students doing something else.
10. Our teacher is aware of what is happening in the classroom, even if he or she is busy with an individual student.

Cognitive activation

And how often does your mathematics teacher do the following things?

1. Our mathematics teacher presents tasks for which there is no obvious solution.
2. Our mathematics teacher presents tasks that require us to apply what we have learned to new contexts.
3. Our mathematics teacher gives tasks that require us to think critically.
4. Our mathematics teacher asks us to decide on our own procedures for solving complex tasks.
5. Our mathematics teacher gives us opportunities to explain our ideas.
6. Our mathematics teacher encourages us to question and critique arguments made by other students.
7. Our mathematics teacher requires us to engage in discussions among ourselves.

Student support

To what extent do you disagree or agree with the following statements?

1. Our mathematics teacher gives extra help when we need it.
2. Our mathematics teacher continues teaching until we understand.
3. Our mathematics teacher helps us with our learning.
4. I get along well with my mathematics teacher.
5. My mathematics teacher is interested in my well-being.
6. My mathematics teacher really listens to what I have to say.
7. My mathematics teacher treats me fairly.
8. My mathematics teacher makes me feel she/he really cares about me.

Figure 1.TIF