Journal of School Psychology 55 (2016) 39–56

Contents lists available at ScienceDirect

Journal of School Psychology

journal homepage: www.elsevier.com/locate/jschpsyc

Inter- and intra-individual differences in teachers' self-efficacy: A multilevel factor exploration

Marjolein Zee [a,*], Helma M.Y. Koomen [a], Francine C. Jellesma [a], Jolien Geerlings [b], Peter F. de Jong [a]

[a] Research Institute of Child Development and Education, University of Amsterdam, The Netherlands
[b] Utrecht University, The Netherlands

ARTICLE INFO

Article history:
Received 1 March 2015
Received in revised form 24 August 2015
Accepted 10 December 2015
Available online 12 January 2016

Keywords:
Teacher self-efficacy
Upper elementary school
Intra-individual differences
Multilevel confirmatory factor analysis
Cluster bias

ABSTRACT

This study explored inter- and intra-individual differences in teachers' self-efficacy (TSE) by adapting Tschannen-Moran and Woolfolk Hoy's (2001) Teachers' Sense of Efficacy Scale (TSES) to the domain- and student-specific level. Multilevel structural equation modeling was used to evaluate the factor structure underlying this adapted instrument, and to test for violations of measurement invariance over clusters. Results from 841 third- to sixth-grade students and their 107 teachers supported the existence of one higher-order factor (Overall TSE) and four lower-order factors (Instructional Strategies, Behavior Management, Student Engagement, and Emotional Support) at both the between- and within-teacher level. In this factor model, intra-individual differences in TSE were generally larger than inter-individual differences. Additionally, the presence of cluster bias in 18 of 24 items suggested that the unique domains of student-specific TSE at the between-teacher level cannot merely be perceived as the within-teacher level factors' aggregates. These findings underscore the importance of further investigating TSE in relation to teacher, student, and classroom characteristics.

The last decades have witnessed the growth of teacher self-efficacy (TSE) studies from a small side-branch of school effectiveness research to a major area of educational psychology (Klassen, Tze, Betts, & Gordon, 2011). One of the triggers for this progress is the belief that generalized TSE, or the self-confidence with which teachers approach and bring about their daily teaching tasks, is a central determinant of teachers' behaviors and actions in the classroom (Bandura, 1997; Tschannen-Moran & Woolfolk Hoy, 2001). Both theoretical and empirical sources have surfaced the tacit notion that teachers high in self-efficacy are more likely than poorly efficacious educators to set high goals for themselves, to activate adequate effort to perform specific teaching tasks, and to persist when the goings get tough (e.g., Bandura, 1997, 2000; Gibson & Dembo, 1984; Tschannen-Moran & Woolfolk Hoy, 2001). Moreover, there is evidence to suggest that teachers with a resilient sense of self-efficacy are generally effective in providing the instructional and affective supports that match their students' needs and lead to positive learning outcomes (e.g., Guo, McDonald Connor, Yang, Roehring, & Morrison, 2012; Justice, Mashburn, Hamre, & Pianta, 2008; Leroy, Bressoux, Sarrazin, & Trouilloud, 2007).

To date, empirical research has predominantly concentrated on measuring between-teacher differences in TSE and its outcomes (cf., Ross, 1994). As such, most studies have implicitly assumed TSE to be a relatively stable, almost trait-like teacher characteristic which, at best, may fluctuate across various teaching tasks and domains (Raudenbusch, Rowan, & Cheong, 1992; Tschannen-Moran & Woolfolk Hoy, 2001). Apart from its static aspects, however, TSE has also been perceived as an inherently mutable state within teachers, which largely depends on challenges presented by different types of students in class (Raudenbusch et al., 1992; Ross,

* Corresponding author at: Research Institute of Child Development and Education, University of Amsterdam, P.O. Box 15776, NL-1001 NG Amsterdam, The Netherlands. Tel.: +31 20 525 1250.
E-mail address: M.Zee@uva.nl (M. Zee).
Action Editor: Kathy Moritz Rudasill

http://dx.doi.org/10.1016/j.jsp.2015.12.003
0022-4405/

Cousins, & Gadalla, 1996; Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998). Unfortunately, though, the examination of intra-individual variability in teachers' self-efficacy has largely gone unheeded by educational research, as its measurement and analysis have generally been presumed to be relatively complex. In the present study, therefore, we aimed to advance understanding of the multifaceted nature of teachers' sense of self-efficacy by exploring this construct across various domains of teaching and learning and particular students. Distinguishing inter- and intra-individual differences in TSE may be important for determining how these capability beliefs are shaped and what their effects are on individual students' academic adjustment in the classroom.

1. A social cognitive perspective on teacher self-efficacy

Empirical research on TSE has predominantly been grounded in Bandura's (1977, 1986, 1997) social cognitive framework. Central to this framework is the idea that people are not merely nudged by the whims of their environment or biological make-up, but rather operate within a system of triadic reciprocal causation. This complex system indicates that environmental constraints or resources are likely to operate through such important personal cognitions as self-efficacy, which organize and produce actions for given purposes (Bandura, 1997, 2006; Pajares, 1997). According to Bandura (1997), these capability beliefs provide the power to act differently from what specific contextual forces dictate, by activating and sustaining the skills, motivation, and effort required for desired achievements to be realized. Educational researchers have, for instance, highlighted the importance of TSE for teachers' ability to manage and motivate difficult students, and their level of effort and persistence in getting these students to study (e.g., Almog & Shechtman, 2007; Bandura, 1997; Lambert, McCarthy, O'Donnell, & Wang, 2009; Tschannen-Moran & Woolfolk Hoy, 2001). Accordingly, teachers' self-efficacy has generally been considered a vital predictor of behavior and action in the domain of teaching and learning.

The basic tenets of the social cognitive paradigm have offered some useful insights into how self-efficacy could be best approached. Among those guiding principles is the recognition of the “person-in-context” in capturing the construct of self-efficacy. For the domain of teaching and learning, this emphasis on environment implies that the degree of specificity of teaching tasks and domains has to be adequately identified (Bandura, 1997, 2006; Tschannen-Moran & Woolfolk Hoy, 2001). Moreover, it underscores the importance of considering environmental obstacles that embody gradations of challenge to which teachers can adjudge their sense of self-efficacy.

2. Degree of domain specificity of TSE

Teachers' sense of self-efficacy has been generally conceptualized at various levels of specificity. As such, this construct can be perceived to reside along a continuum from domain generality at one end to increasingly advanced specificity levels at the other (Lent & Brown, 2006). At the most universal level, TSE has been regarded as a single-level, trait-like construct, reflecting generalized capability beliefs that fluctuate between teachers (e.g., Schwarzer, 1992; Schwarzer & Jerusalem, 1995). Investigators taking such a theoretical stance habitually decontextualize TSE from a wider scope of tasks and domains in the classroom, resulting in one-dimensional, all-purpose measures that are widely applicable to a range of outcomes (e.g., Bandura, 1997; Pajares, 1996). Moreover, they commonly treat within-teacher variations in TSE as error variance, as these variations only represent deviations from teachers' baseline level of self-efficacy.

Generalized measures that capture between-teacher differences in TSE appear, by far, to be the most frequently used in studies conducted from 1998 to 2009 (Klassen et al., 2011). Indicative of such between-teacher tests are the oft-cited Teacher Efficacy Scale (TES; Gibson & Dembo, 1984) and Schwarzer and Jerusalem's (1995) General Efficacy Scale (GES). Despite their popularity, however, these measures have been criticized for being invalid and lacking predictive relevance (Bandura, 1997, 2006; Kagan, 1990; Pajares, 1996). For instance, domain-general scales have been suggested to be problematically ambiguous in the sense that teachers are forced to guess what the unspecified contextual details of individual items might be (Bandura, 1997; Wheatley, 2005). Items such as “I know that I can motivate my students to participate in innovative projects” (Schwarzer, Schmitz, & Daytner, 1999) may place a burden on teachers to comprehend what is being asked of them, as it leaves unspecified what “innovative projects” are. Moreover, it is likely that global measures fail to adequately match with the particular outcomes in the classroom that are of interest to the researcher (Bandura, 1997, 2006). Those potential misalliances between predictor and outcome may come at the expense of the explanatory and predictive merit of general TSE measures (Pajares, 1996).

Recognizing that further specification of TSE is required to elucidate the self-efficacy regulation of teachers' behaviors in the classroom, more recent scholars have shifted focus to subject-, task-, or domain-specific conceptualizations of TSE (Brouwers & Tomic, 2000; Dellinger, Bobbett, Olivier, & Ellett, 2008; Friedman & Kass, 2002; Riggs & Enochs, 1990; Siwatu, 2007, 2011; Tschannen-Moran & Woolfolk Hoy, 2001; Tschannen-Moran & Johnson, 2011; Tsouloupas, Carson, Matthews, Grawitch, & Barber, 2010). One of the most celebrated attempts at this domain-level of specificity comes from Tschannen-Moran and Woolfolk Hoy (2001). In a seminar on efficacy in teaching and learning, these researchers pooled and discussed both new and existing items to construct a TSE scale that assumedly considers the full range of teaching tasks and responsibilities. This measure, which is generally known as the Teachers' Sense of Self-Efficacy Scale (TSES), holds promise as a flexible research tool that can be used across grades (Tschannen-Moran & Woolfolk Hoy, 2001), subjects (Tschannen-Moran & Woolfolk Hoy, 2007), and teaching contexts (Klassen et al., 2009; Woolfolk Hoy & Burke-Spero, 2005). Moreover, its factorial, convergent, and concurrent validity has been demonstrated in several empirical studies (e.g., Heneman, Kimball, & Milanowski, 2006; Klassen et al., 2009; Tschannen-Moran & Woolfolk Hoy, 2001; Wolters & Daugherty, 2007).

The TSES has taken account of three unique teaching domains that appear to be the most germane to teachers' daily activities and students' academic adjustment. Tschannen-Moran and Woolfolk Hoy (2001) labeled these domains as TSE for instructional strategies, student engagement, and classroom management, with the first two domains usually being the most highly correlated (e.g., Tsigilis, Koustelios, & Grammatikopoulos, 2010). Recently, attention has also been drawn to another domain that may be relevant to teachers' self-efficacy for teaching and learning. This domain of emotional support involves tasks and responsibilities related to how well teachers can establish caring relationships with students, acknowledge students' opinions and feelings, and create settings in which students feel secure to explore and learn (e.g., Pianta, La Paro, & Hamre, 2008). A rich body of empirical research has indicated that emotionally supportive teacher behaviors, next to instructional, motivational, and organizational aspects of teaching and learning, are one of the strongest correlates of students' achievement, engagement, and enjoyment during learning tasks (e.g., Crosnoe, Johnson, & Elder, 2004; Hamre, Hatfield, Pianta, & Jamil, 2014; Reyes, Brackett, Rivers, White, & Salovey, 2012; Rimm-Kaufman & Chui, 2007; Rimm-Kaufman, La Paro, Downer, & Pianta, 2005; Roorda, Koomen, Spilt, & Oort, 2011). Therefore, at the domain-level of specificity, adding the emotional support domain may provide, above and beyond the domains of instructional strategies, classroom management, and student engagement, relevant insights into the multifaceted nature of TSE and its outcomes in the classroom.

Investigators have increasingly supported the need to use measures of TSE in specific domains of teaching and learning (e.g., Bandura, 1997; Brouwers & Tomic, 2000; Dellinger et al., 2008; Tschannen-Moran et al., 1998; Tsouloupas et al., 2010). Tschannen-Moran et al. (1998, p. 227–228), for instance, are adamant of the idea that "teachers feel efficacious for teaching particular subjects to certain students in specific settings, and [that] they can be expected to feel more or less efficacious under different circumstances". Consistent with this Bandurian notion, a modest body of empirical research on within-teacher variations in TSE (Raudenbusch et al., 1992; Ross et al., 1996) has furthermore indicated that teachers' sense of self-efficacy may be significantly affected by contextual factors, such as subject matter, student behavior, and the type of students they teach. Apart from between-teacher differences, this within-person variation in TSE is important to recognize, as it may advance understanding of the changing states of teachers' self-efficacy beliefs across domains and particular students. Unfortunately, though, research on TSE towards particular children under different domains of functioning seems to be more the exception than the rule. For this reason, we will not only consider the degree of domain specificity of teachers' self-efficacy, but the level of student specificity as well.

3. Degree of student specificity of TSE

Taking teachers' self-efficacy to both the domain- and student-specific level without becoming too specific is no easy matter. Similar to global capability beliefs, overly particularized self-efficacy judgments have been criticized by prior research (e.g., Bandura, 1997; Pajares, 1996; Tschannen-Moran et al., 1998) for their potential lack of external validity and practical relevance to the field of education. Tschannen-Moran and Woolfolk Hoy (2001, p. 795), for instance, strikingly illustrate how such microscopically operationalized self-efficacy items as "I am confident I can teach simple subtraction to middle-income second graders in a rural setting who do not have specific learning disabilities, as long as my class is smaller than 22 students and good manipulatives are available" may loose both predictive power for other teaching contexts and students, as well as practical utility. Potentially, such issues may be circumvented by allowing the level of domain specificity to depend on obstacles against which teachers can adjudge their self-efficacy (cf. Pajares, 1996). Assumedly, the behaviors and characteristics that students bring to the classroom may function as such obstacles, determining the strength of teachers' self-efficacy across various domains of teaching and learning.

Past research on self-efficacy has since long acknowledged the importance of viewing teachers' self-efficacy in light of various environmental obstacles (Bandura, 1997, 2006; Coladarci & Breton, 1997; Pajares, 1996; Wheatley, 2005; Wyatt, 2014). Without such obstacles, the interpretation of TSE may be ambiguous, as teachers are likely to base their responses on imagined students or situations. For instance, teachers may respond confidently to such TSES-items as "How much can you do to get children to follow classroom rules?" (Tschannen-Moran & Woolfolk Hoy, 2001), but may reply far less self-confident when the question is "How much can you do to get disruptive children in your classroom to follow classroom rules?". Hence, obstacles, such as disruptive children in this case, may avoid teachers to become naïvely optimistic about their self-efficacy beliefs, and may ameliorate the predictive validity of TSE (Bandura, 1997, 2006; Wheatley, 2005; Wyatt, 2014).

Defining obstacles may present, as Tschannen-Moran and Woolfolk Hoy (2001, p. 794) aptly point out, "thorny issues," as they may substantially increase the complexity of each individual item. In existing measures of TSE, most of the items usually lack such clear obstacles. Yet, some attempts have been made to include them in a handful of TSES-items. All these embedded challenges extend to student characteristics or behaviors, including 'very capable students', 'problem students', 'students who show low interest in schoolwork', and 'students who are failing' (Tschannen-Moran & Woolfolk Hoy, 2001). Thus, the gradations of challenge to teachers' performance are likely to be mainly determined by individual students' behaviors and actions.

The idea that obstacles are predominantly reflected in student characteristics fits fairly well with the assumption that TSE may vary across different students. Moreover, with this assertion comes a way to resolve the persisting issue of how situational impediments should be defined. By letting teachers report on their self-efficacy for individual students, it becomes possible to specify the forms the impediments take, without unnecessarily complicating individual self-efficacy items, or limiting the generalizability of the TSES or other domain-specific self-efficacy instruments. In addition, through this particular manner of specifying obstacles, teachers may be less likely to respond in a socially desirable direction, as they may rather ascribe their low self-efficacy to characteristics of particular students, than to their incompetent self.

4. Present study

From the social cognitive paradigm, it follows that TSE is best approached by capturing the teaching domains and students that generate inter- and intra-individual differences in teachers' capability beliefs. Such domain-linked and student-specific self-efficacy beliefs may generally be more predictive of specific teacher behaviors and actions, due to the variations in self-efficacy percepts that occur across different task domains and specific students. Unfortunately, however, conceptual and methodological issues have largely prevented researchers to take such conditional self-efficacy states into consideration. This study, therefore, set out to explore teachers' sense of self-efficacy across various domains of teaching and learning and particular students. To this end, we took Tschannen-Moran and Woolfolk Hoy's (2001) original TSES to the domain- and student-specific level by making its individual items student-specific and including the domain of emotional support.

Initially, we examined the multilevel factor structure of the adapted instrument to explore inter-individual (trait) differences in TSE at the between-teacher level and intra-individual (state) differences at the within-teacher level. Largely consistent with the original TSES, we expected to find empirical support for a four-factor multilevel model, representing the TSE domains of Instructional Strategies, Behavior Management, Student Engagement, and Emotional Support. To meaningfully compare domain- and student-specific TSE across teachers, we subsequently tested for violations of measurement invariance over clusters, or cluster bias (Jak, Oort, & Dolan, 2013, 2014). In the present study, the absence of cluster bias would indicate that teachers' self-efficacy reports are likely to measure the same constructs across educators, and that its hypothesized domains at the between-teacher level can be perceived as the aggregate of the within-teacher level dimensions (Jak et al., 2014). As such, it can also be expected that our adapted instrument is likely to show moderate to strong correspondence with the original TSES, providing evidence for the concurrent validity of this measure.

5. Method

5.1. Participants

The participants in the present study included regular elementary school teachers and their students drawn from third- to sixth-grade classrooms in the Netherlands. After ethical approval from the Ethics Review Board of the Faculty of Social and Behavioral Sciences, University of Amsterdam, was granted (project no. 2013-CDE-3188), approximately 700 schools across the Netherlands were drawn from the total pool of 6800 regular Dutch elementary schools. To promote the sample's representativeness with respect to the variables measured in our study, we aimed at selecting a wide range of schools that were demographically diverse in terms of geographical spread, denomination, school size, urbanicity, and characteristics of the student population.

Of the schools that were initially invited, 42 ultimately agreed to take part in the study. This sample of schools appeared to represent a relatively balanced cross-section of the larger population of schools in the Netherlands (see Table 1). Non-participation was mainly due to schools' already full agendas, or their involvement in other research studies. After schools agreed to participate, information letters about the nature and purposes of the study were sent to all teachers who taught in the upper elementary grades, soliciting their voluntary participation in the study. On average, three teachers per participating school (range = 1–8 teachers;

Table 1
Demographic characteristics of participating schools

	Total sample		Total population
	N	Percentage	Percentage
Geographical region			
North	6	14.3%	10.1%
East	12	28.6%	22.5%
South	10	23.8%	19.8%
West	14	33.3%	47.6%
Denomination			
Public school	19	45.3%	33%
Protestant Christian school	10	23.8%	30%
Catholic school	10	23.8%	29%
Other	3	7.1%	8%
School size			
< 101 students	5	11.9%	18.9%
101-201 students	16	38.1%	31.7%
201-501 students	17	40.5%	44.9%
> 501 students	4	9.5%	4.5%
Urbanicity			
Urban	16	38.1%	-
Peri-Urban	15	35.7%	-
Rural	11	26.2%	-

Note. Demographic data for the total population of Dutch elementary schools are retrieved from CBS Statline (2015b).

participation rate = 70.8%) expressed their interest in participation, resulting in an original sample of 113 teachers. Teachers who refrained from participation generally were substitute teachers and educators with additional tasks and responsibilities, including mentoring, coordinating, or remedial teaching tasks. Of the original sample of 113 teachers, six (5.3%) additionally failed to complete all questionnaires due to long-term absence, sickness, or strenuous workloads. Given that these data were not missing completely at random, we decided to exclude those cases from analyses.

5.1.1. Teacher sample

Complete data were available for 107 teachers (73.5% females), ranging in age from 20 to 63 years ($M = 42.02$, $SD = 12.36$). On average, teachers had 16.58 years ($SD = 11.58$) of professional teaching experience, with the least experienced teacher working only half a year in primary education, and the most experienced teacher having a 44-year teaching career. These demographic characteristics are comparable to those of the larger population of Dutch teachers, who generally have a mean age of 43.25 years (range = 19–67 years), and are typically female (84%; DUO, 2014).

Some past empirical research has suggested that teachers' years of professional experience may positively add to their sense of self-efficacy (e.g., Klassen & Chiu, 2010; Morris-Rothschild & Brassard, 2006). Other studies, however, have indicated that TSE is likely to decrease over time (Cantrell, Young, & Moore, 2003) or may not at all be associated with teaching experience (e.g., Ghaith & Yaghi, 1997; Soodak & Podell, 1996). In the present sample, analyses of variance showed that teachers with little experience (<5 years), average experience (5–10 years), or high experience (>10 years) did not differ in their domain- and student-specific self-efficacy beliefs, $p > .05$.

5.1.2. Student sample

For the student sample, both the first and fourth authors randomly selected four boys and four girls from each teacher's classroom. This sample contained children from grades 3 ($N = 54$), 4 ($N = 262$), 5 ($N = 270$) and 6 ($N = 255$), respectively. The students ranged from 7 to 13 years of age ($M = 10.83$, $SD = 1.04$) and the gender composition was evenly distributed with 420 boys (49.9%) and 421 girls (50.1%). Most students had a Dutch origin (73%), with the remaining 27% of students representing other ethnic backgrounds. Based on teacher reports of parents' working status and educational level, most students were considered to have an average to high socioeconomic status. Both parents were employed in 65.9% of the families, 27.5% had at least one employed parent, and only 4.9% of the families included two unemployed parents. In addition, teachers indicated the majority of the parents to have finished senior vocational education (48.8%) or higher education (39.3%), leaving 9% of the parents to only have finished primary education. For less than 3% of the students, teachers failed to provide information on parents' working status and educational background.

The student sample appeared to be relatively similar to the larger population of third- to sixth-graders in the Netherlands in terms of gender (50.5% male students) and ethnicity (15% non-Dutch origin; CBS Statline, 2015a, 2015b). Moreover, previous studies using nationally representative elementary school samples (e.g., Hornstra, van der Veen, Peetsma, & Volman, 2013; Zee, Koomen, & van der Veen, 2013) reported demographic characteristics for third- and sixth-graders that resemble those of the students included in the present study. Hence, although the participating schools, teachers, and students cannot be considered to be fully representative in this study, they seem to reasonably approximate the larger population.

5.2. Instruments

5.2.1. Overall teacher self-efficacy

Teachers' perceptions of their overall level of self-efficacy were measured using a short, 12-item version of Teachers' Sense of Efficacy Scale (TSES; Tschannen-Moran & Woolfolk Hoy, 2001). The TSES is specifically designed to evaluate teachers' perceptions of their competence across a variety of important teaching tasks. Analogous to the original 24-item instrument, the short TSES has been evidenced to comprise three interrelated dimensions of teacher self-efficacy, which are labeled Instructional Strategies (IS), Classroom Management (CM), and Student Engagement (SE). The domain of IS (4 items) measures the extent to which teachers feel able to use various instructional methods that enable and enhance student learning. The CM domain (4 items) taps teachers' perceptions of their ability to organize and guide students' behavior. TSE for SE (4 items) captures teachers' perceived ability to activate students' interest in their schoolwork. Example items for each of these domains of TSE include "To what extent can you provide an alternative explanation or example when students are confused?", "How much can you do to get children to follow classroom rules?", and "How much can you do to help your students value learning?", respectively. Although the TSES is usually measured on a 9-point rating scale, teachers in the present study responded on a 7-point rating scale, ranging from 1 (*nothing*) to 7 (*a great deal*). Reason to deviate was that prior research (e.g., Diefenbach, Weinstein, & O'Reilly, 1993) has indicated that 7-point scales generally outperform 2-, 5-, 9-, 11-, 12-, and 100-point scales on accuracy, perceived ease of use, and agreement of scale-derived ranks with direct rankings.

The psychometric properties of the short form of the TSES have been shown to be adequate and largely comparable to those of the long form (e.g., Tschannen-Moran & Woolfolk Hoy, 2001). In prior research, alpha coefficients ranged between .71 and .87 for IS, .83 and .94 for CM, and .74 and .88 for SE, respectively (e.g., Klassen et al., 2009; Tschannen-Moran & Woolfolk Hoy, 2001). In addition to these adequate alpha coefficients, Klassen et al. (2009) found evidence of strong structural and measurement invariance in groups of teachers who differed on language, cultural practices and beliefs, teaching environment, and school level. Correlations between the TSES dimensions and adjacent constructs, including personal efficacy, teaching efficacy and job satisfaction, also lend credence to the convergent and concurrent validity of the short TSES (Klassen et al., 2009; Tschannen-Moran & Woolfolk

Hoy, 2001). Together, these reliability and validity assessments seem to support the appropriateness of the short TSES for use in different contexts.

To evaluate the reliability and factorial validity of the short TSES in the present study, we performed a confirmatory factor analysis for complex survey data, using robust maximum likelihood estimation (MLR) in M*plus* 7.11 (Muthén & Muthén, 1998-2012). This method takes the non-independence of data due to clustering into account, and provides a mean-adjusted chi-square test and standard errors that are robust for non-normality (Muthén & Muthén, 1998-2012; Yuan & Bentler, 2000). Both a three-factor solution ($\chi^2(50) = 93.20$, $p < .001$, RMSEA = .033 (90% CI [.022, .043]), CFI = .89, SRMR = .076) and a one-factor solution ($\chi^2(43) = 77.12$, $p < .001$, RMSEA = .031 (90% CI [.020, .043]), CFI = .89, SRMR = .066) yielded a reasonable fit, after adding a theoretically plausible correlation residual to both models. Although the CFI was below the conventional threshold of .90 for satisfactory fit, the model showed quite sound goodness of fit according to established cutoff values of .08 for the RMSEA and SRMR (Bentler, 1992; Browne & Cudeck, 1993; Hu & Bentler, 1999; Kline, 2011). Factor loadings ranged between .47 and .85 in the three-factor model, and between .37 and .79 in the one-factor solution. Alpha coefficients were .84 for Overall TSE, .71 for IS, .76 for CM and .77 for SE, respectively.

5.2.2. Domain- and student-specific teacher self-efficacy

To measure teachers' self-efficacy towards particular children in different domains of functioning, we developed a new instrument, based on the original, 24-item TSES of Tschannen-Moran and Woolfolk Hoy (2001). The adaptation process began with the adjustment of the original TSES items to the student-specific level (see Appendix A). For instance, the item "How much can you do to get children to follow classroom rules?" was changed into "How much can you do to get *this student* to follow classroom rules?". Classroom Management items 12 ("How well can you establish a classroom management system with each group of students?") and 16 ("How well can you establish routines to keep activities running smoothly?") of the original scale were omitted, as they could not be accurately made specific to the level of individual students. Notably, whereas all other adapted items of this scale concentrated on teachers' perceived ability to manage the behavior of individual students, items 12 and 16 mainly focused on aspects of classroom management. As such, these two items also reflected a slightly different construct. In addition, several original TSES items (items 8, 13, 19, and 21) included embedded obstacles that embody gradations of challenge to teachers' tasks in a given teaching domain. Examples of such obstacles are "very capable students" (item 8), "problem students" (item 13), "students who show low interest in schoolwork" (item 19), and "students who are failing" (item 21). By evaluating teachers' self-efficacy beliefs in relation to individual students, however, it becomes possible to specify the forms the impediments take in all TSES items, without unnecessarily complicating these items. In the process of making the original TSES items student-specific, we therefore consistently removed all embedded obstacles from items 8, 13, 19, and 21.

After adjusting the original TSES items to be student-specific, we further shortened the original TSES by removing four less relevant items. The first item ("To what extent can you use a variety of assessment strategies?") was discarded because this item was not representative of the regular teaching tasks of Dutch elementary school teachers. Furthermore, this item appeared to have one of the poorest factor loadings in samples of elementary school teachers (e.g., Heneman et al., 2006; Klassen et al., 2009), suggesting that this item may be more relevant for secondary school teachers. The main reason to remove the fifth item ("How well can you respond to difficult questions from your students?") involved the ambiguous nature of this item. Specifically, this item might either relate to students' difficulties regarding instruction or learning tasks, or refer to issues of a more personal nature, such as family problems. Probably, this ambiguity is also reflected in the relatively low factor loading of this item in previous research (e.g., Wolters & Daugherty, 2007; Tschannen-Moran & Woolfolk Hoy, 2001). This may explain why this item is not part of the short form of the original TSES.

Additionally, after adjusting the level of specificity of item 14 ("How well can you respond to defiant students?"), a substantial overlap between this question and other Classroom Management items was recognized. Therefore, this item was removed as well. Given that the reported factor loadings of Classroom Management items are usually quite substantial (> .70), and of roughly equal magnitude (e.g., Heneman et al., 2006; Klassen et al., 2009; Wolters & Daugherty, 2007), the removal of this item did probably not affect the consistency of this scale. Lastly, item 20 ("How much can you assist families in helping their children do well in school?") was considered to have too little in common with the domain of Student Engagement. Moreover, prior research reporting on the factor structure of the TSES consistently showed that this item is least stable and has the poorest factor loading in general (Heneman et al., 2006; Klassen et al., 2009; Wolters & Daugherty, 2007). The removal of six items in total (in stage one and two) resulted in a total of 18 adapted items (6 IS, 5 CM, and 7 SE items) that were retained in the new instrument.

Subsequent to adapting the three original TSES domains, we used the CLASS framework (for an overview, see Hamre et al., 2013 and Pianta et al., 2008) to construct seven new items that aimed to cover the domain of Emotional Support. These items were based on the common metric used to describe positive dimensions of the CLASS-domain of Emotional Support, including Positive Climate, Teacher Sensitivity, and Regard for Student Perspectives (Pianta et al., 2008). Three items concerned teachers' perception of their ability to establish a warm connection with individual students (Positive Climate). Two other pairs of items measured teachers' perceived ability to be aware of, and responsive to individual students' academic and emotional needs (Teacher Sensitivity) and to emphasize students' viewpoints and interest (Regard for Student Perspectives). The addition of these items resulted in a 25-item instrument, reflecting the domains of Instructional Strategies (IS; 6 items), Behavior Management (BM; 5 items), Student Engagement (SE; 7 items), and Emotional Support (ES; 7 items), respectively. Largely similar to the original TSES, responses to each of these items were given on a seven-point rating scale, ranging from 1 (*nothing*) to 7 (*a great deal*).

The translation of the TSES, lastly, was performed using a standard forward-backward procedure, involving two forward translators and one backward translator. In the first step of the translation process, the first and second author, both native Dutch speakers, independently translated the original English version of the TSES into the Dutch language. After the translations were completed, they compared all items, and critically evaluated them on parameters like difficulties in translation, doublets of items, and relevance for the Dutch school context. Any discrepancies between the two translations were solved by consensus with the other authors. This process resulted in a single conditional forward translation of the student-specific TSES, which offered some alternative wordings for (parts of) items that appeared to be difficult to translate, and included the seven new items on Emotional Support. This provisional version was back-translated by a native English speaker from Dutch into English, and checked by the first author.

In the second step of the translation process, the student-specific TSES was pilot tested with six elementary school teachers, who reviewed the items for content validity, clarity of wording, and relevance of the response scale. Based on their analysis, the first two authors slightly reworded one adapted TSES-item (item 1) that was deemed too complex, without altering its meaning.

5.3. Procedure

Data for this study were collected between January and March 2014. Prior to data collection, participating schools were asked to distribute a letter to students' parents, explaining the nature and purposes of the study and providing a form to refuse permission, which could be returned to school. All parents voluntarily gave their consent to their child's participation in this study. Participating teachers signed a written informed consent form at the start of data collection.

To avoid common method variance, teacher survey data were collected in two parts. The first, written part of the survey was administered during a planned school visit, and consisted of demographic items and the short TSES, respectively. Teachers who were not present at the time of data collection could return the survey by regular mail. The second part of the survey was distributed directly after the school visit, by sending an e-mail invitation that contained an anonymous survey link. This digital survey, which was completed for eight randomly selected students from teachers' classrooms, had a forced response format and involved the newly developed student-specific TSES, and some general questions regarding parents' socioeconomic status. Teachers were asked to return the digital survey within two weeks after the invitation was sent. To improve the participation rate, reminders were sent to non-responding teachers. Ultimately, six teachers failed to fill out the survey and another four teachers completed the survey for less than eight students, due to time constraints. This resulted in a total response rate of 94.6%.

5.4. Data analysis

We used multilevel confirmatory factor analysis (MCFA) to test the factor structure of the student-specific TSES. With this analytic technique, model fit and parameter estimate biases can be avoided by decomposing the total sample covariance matrix into a pooled within-group (Σ_{WITHIN}) and a between-group ($\Sigma_{BETWEEN}$) covariance matrix (Muthén, 1994). In addition, MCFA is well suited to detect violations of measurement invariance across clusters, or cluster bias, in multilevel data (Jak et al., 2013, 2014). This relatively new technique is particularly useful when collecting the same measure from qualitatively different groups or individuals operating in distinct contexts, as it aims to take differences in response processes into account (Muthén & Asparouhov, 2013; Ryu, 2014). Generally, cluster bias indicates that teachers might answer differently on the self-efficacy items, despite having similar beliefs in their capability. These systematic differences in observed self-efficacy scores seem to occur when contextual factors or personal teacher characteristics implicitly affect teachers' interpretation of self-efficacy items. Thus, in this study, the presence of cluster bias would indicate that the dimensions of the student-specific TSES do not measure the same constructs over teachers, and that part of the variance in teachers' student-specific self-efficacy beliefs may be attributed to teacher and/or classroom characteristics.

5.4.1. Modeling procedure

In line with the strategies of Jak et al. (2014) for the investigation of cluster bias, we followed four analytical steps. First, to determine whether multilevel modeling was required, we calculated the intraclass correlation coefficients (ICC) for each of the model's indicators and tested whether the between-teacher level variance and covariance deviated significantly from zero. To this end, we fitted a Null Model ($\Sigma_{BETWEEN} = 0$, $\Sigma_{WITHIN} =$ free) and an Independence Model ($\Sigma_{BETWEEN} =$ diagonal, $\Sigma_{WITHIN} =$ free) to the data (Jak et al., 2013, 2014; Muthén, 1994). Generally, poor fit of these models are indicative of meaningful between-teacher level variance and covariance (Hox, 2002).

In step two, we first conducted a confirmatory factor analysis on the sample pooled-within covariance matrix to determine the factor structure at the within-group level only (Dyer, Hanges, & Hall, 2005; Hox, 2002; Muthén, 1994). Apart from the proposed four-factor model, we also considered several alternative models, including one-factor and three-factor solutions, and Tschannen-Moran and Woolfolk Hoy's (2001) original three-factor and higher-order factor models, to determine potential sources of model misspecification.

In the third step, we used the measurement model that was established in step two to investigate cluster bias. We started with a fully constrained model, in which all factor loadings were constrained to be equal across the within- and between-teacher level, and residual variances at the between-teacher level were fixed at zero. To test whether strong factorial invariance held across clusters, we sequentially allowed the between-teacher level residual variances to be freely estimated. Generally, residual variances greater than zero are indicative of cluster bias in their corresponding indicators (Jak et al., 2013, 2014).

Subsequently, we evaluated whether factor loadings could be considered equal across clusters. Unequal factor loadings indicate that the unique domains of student-specific TSE at the between-teacher level cannot merely be assumed to be the within-teacher level factor's aggregates.

Finally, in the fourth step, we fitted a restricted factor analysis (RFA; Oort, 1992) to investigate the concurrence between Tschannen-Moran and Woolfolk Hoy's (2001) original TSES and the student-specific TSES. To this end, we extended the multilevel measurement model to include correlations between the generalized TSES and the student-specific TSES at the between-level of measurement. Depending on the presence of cluster bias, we expected a moderate to strong correspondence between the original and student-specific TSES.

5.4.2. Model goodness-of-fit

Multilevel models were fitted in M*plus* 7.11, using robust maximum likelihood estimation (MLR; Muthén & Muthén, 1998-2012). This method of estimation offers a mean-adjusted χ^2, which is asymptotically equivalent to Yuan and Bentler's (2000) T_2-test statistic and generates adjusted standard error estimates that are robust for non-normality (Muthén & Muthén, 1998-2012). Generally, the adjusted χ^2 test statistic indicates a good overall model fit when it does not reach the significance threshold. However, as even trivial discrepancies between the expected and the observed model may lead to the model's rejection (Chen, 2007), other criteria in evaluating fit were used as well. These included the root mean square of approximation (RMSEA) and standardized root mean square residual (SRMR), with values ≤.05 reflecting a close fit, and ≤.08 a satisfactory fit (Browne & Cudeck, 1993; Hu & Bentler, 1999; Kline, 2011), and the comparative fit index (CFI), with values ≥.95 indicating close fit, and values ≥.90 indicating acceptable fit (Bentler, 1992). To compare alternative models, we employed the (Satorra–Bentler scaled) chi-square difference test (TRd; Satorra, 2000; Satorra & Bentler, 2010), with non-significant chi-squares indicating equivalent fit, and the CFI-difference, with CFI changes ≥.02 being indicative of model nonequivalence (Cheung & Rensvold, 2002).

6. Results

6.1. Data screening and descriptive statistics

Inspection of the distributional properties of both the total score and the three subscales of the original, overall TSES-domains revealed no serious departures from normality and linearity. Skewness levels were −0.20 for Overall TSE, −0.54 for IS, −0.78 for CM, and −0.30 for SE, and kurtosis values −0.42 for Overall TSE, −0.01 for IS, 0.71 for CM, and 0.22 for SE, respectively. Teachers' mean responses on the original TSES, reported on a 7-point scale, were lowest for SE ($M = 5.46$, $SD = 0.69$), followed by IS ($M = 5.67$, $SD = 0.65$), and CM ($M = 5.90$, $SD = 0.67$). The mean total score of teachers' generalized self-efficacy was 5.70 ($SD = 0.52$). These relatively high means and small standard deviations are consistent with previous findings (e.g., Heneman et al., 2006; Tschannen-Moran & Woolfolk Hoy, 2001).

Teachers' responses on the Student-Specific TSES domains of IS and SE were approximately normally distributed. In these domains, most items did not reach the skewness threshold of ±1.00 (range = −0.63 to −1.07 for IS and −0.64 to −1.19 for SE). Moreover, kurtosis values ranged from −0.17 to 1.62 for items comprising the IS domain, and from 0.00 to −1.16 for SE-items. Items appeared to be highly skewed, however, in the domains of BM (range = 1.16 to −1.84) and ES (range = −0.75 to −1.41), and were characterized by high kurtosis (range = 1.16 to 3.82 for BM and 0.35 to 2.32 for ES). To deal with these high skewness levels, we used robust maximum likelihood estimation to obtain parameter estimates (Muthén & Muthén, 1998-2012), as this estimator is robust to non-normality and enables the adjustment of standard errors.

Table 2 displays the means, within-teacher standard deviations, and between-teacher standard deviations of the student-specific TSES items. The descriptive statistics indicate that all item means were relatively high and largely comparable with the averages found for the original TSES domains. Notably, the highest item means were found for items comprising the BM and ES domains of student-specific self-efficacy. Inspection of the partitioned standard deviations, which provide an indication of self-efficacy differences within and between teachers, furthermore shows that there is more variability within teachers than between teachers. This is in line with Bandura's (1997) premise that self-efficacy is more likely to reflect a dynamic state, than a relatively stable trait.

6.2. Multilevel confirmatory factor analysis of the student-specific TSES

6.2.1. Step 1: Evaluating between-teacher level variance and covariance

The intraclass correlations (ICCs) for the Student-Specific TSES items (see Table 2) ranged between .09 (item 7) and .36 (item 24), with a mean ICC of .23. Fit indices of the Null Model, χ^2 (302) = 3244.27, RMSEA = .108, CFI = .77, $SRMR_{WITHIN}$ = .10, $SRMR_{BETWEEN}$ = .65, and the Independence Model, χ^2 (278) = 1898.58, RMSEA = .083, CFI = .88, $SRMR_{WITHIN}$ = .09, $SRMR_{BETWEEN}$ = .65, suggested that there is meaningful between-teacher level variance and covariance. Hence, these clustering effects were substantial enough to warrant the use of MCFA.

6.2.2. Step 2: Evaluating the measurement model at the within-teacher level

Using the sample pooled-within covariance matrix, we examined the hypothesized four-factor model. The overall fit of the model was reasonable, with RMSEA and SRMR values below .08 and a CFI greater than .90, χ^2 (269) = 1484.15, $p < .001$, RMSEA = .073 (90% CI [.070–.077]), CFI = .91, SRMR = .05. To diagnose systematic patterns of misfit, we inspected the model's

Table 2
Item means and standard deviations of the student-specific TSES

Item	*M*	SD_{within}	$SD_{between}$	ICC
TSE for instructional strategies				
IS1	5.87	0.83	0.54	.30
IS2	5.46	1.11	0.61	.23
IS3	5.43	1.05	0.60	.25
IS4	5.71	0.87	0.55	.29
IS5	5.83	0.96	0.53	.24
IS6	5.43	1.01	0.59	.25
TSE for behavior management				
BM1	6.07	1.17	0.36	.09
BM2	6.13	1.11	0.39	.11
BM3	6.18	1.08	0.37	.11
BM4	6.15	1.08	0.41	.13
BM5	6.30	0.84	0.44	.21
TSE for student engagement				
SE1	5.87	0.93	0.50	.23
SE2	5.72	1.19	0.49	.15
SE3	5.72	1.21	0.52	.16
SE4	5.67	1.11	0.51	.18
SE5	5.46	1.11	0.63	.24
SE6	5.26	1.05	0.69	.31
SE7	5.81	1.10	0.45	.14
TSE for emotional support				
ES1	6.30	0.81	0.38	.19
ES2	6.20	0.77	0.44	.25
ES3	6.12	0.79	0.49	.28
ES4	5.81	0.92	0.63	.32
ES5	5.63	0.90	0.62	.32
ES6	5.65	0.92	0.68	.36
ES7	5.43	0.98	0.64	.30

Note. Item means are reported on a 7-point scale. TSE = teacher self-efficacy.

modification indices. These indices suggested model improvement by adding a correlation between the residuals of SE-items 13 ("To what extent can you help this student to value learning?") and 14 ("To what extent can you motivate this student for his/her schoolwork?"). These two items showed a considerable conceptual overlap, both focusing on teachers' perceived capability to motivate individual students for their schoolwork. Following Tabachnick and Fidell's (2007) cut-off criteria, we additionally removed item 22 ("To what extent can you timely recognize that this student does not feel well?"), which loaded poorly on its corresponding factor (<.40). These alterations resulted in a satisfactory fit to the data: χ^2 (245) = 1229.13, $p < .001$, RMSEA = .069 (90% CI [.065–.073]), CFI = .93, SRMR = .05.

6.2.2.1. Alternative models. Although the fit of the hypothesized model was acceptable, there might be alternative models that generate roughly similar, or even better predicted covariances (Kline, 2011). To justify the appropriateness of the hypothesized model, we therefore examined a series of theoretically plausible competing models, including one-factor, three-factor, and higher-order factor models.

The first two competing models tested were a one-factor model and a three-factor model, in which the SE and ES dimensions were combined to create a single Engaging Strategies factor. Comparison of the four-factor model with these one-factor, $\Delta\chi^2$ (30) = 2407.90, $p < .001$, ΔCFI = .17, and three-factor alternatives, $\Delta\chi^2$ (27) = 345.72, $p < .001$, ΔCFI = .02, indicated that both alternative models had a poorer fit to the data, and had slightly worse structural parameter estimates. These results lend credence to the proposed four-factor structure of the student-specific TSES.

Secondly, we evaluated whether the original factor structure proposed by Tschannen-Moran and Woolfolk Hoy (2001) held in the present sample. To this end, we fitted a three-factor model in which all Emotional Support items were omitted. This model obtained an acceptable fit, χ^2 (132) = 708.87, $p < .001$, RMSEA = .072 (90% CI [.067–.077]), CFI = .95, SRMR = .04. The results of this model suggest that the additional domain of Emotional Support can be distinguished from the original TSES-domains and may provide information about teachers' perceived capabilities that goes above and beyond their self-efficacy for Instructional Strategies, Behavior Management, and Student Engagement.

Thirdly, and largely consonant with Tschannen-Moran and Woolfolk Hoy's findings, we considered a hierarchical factor model, in which one second-order factor of teachers' general self-efficacy beliefs towards particular students was hypothesized to underlie the four proposed TSE domains of teaching and learning. Such higher-order models are particularly relevant when hypothesizing general constructs that comprise several closely related domains (Chen, West, & Sousa, 2006). Although this model fitted the data reasonably well, the χ^2 difference test statistic suggested that the hypothesized four-factor model is to be preferred over its higher-order equivalent, $\Delta\chi^2$ (2) = 107.54, $p < .001$, ΔCFI = .01. Based on these comparisons, we gleaned that the proposed four-factor model is most likely the preferred solution.

6.2.3. Step 3: Detecting violations of measurement invariance across clusters

In the third step, we established a measurement model at both the within-teacher (state) and between-teacher (trait) level, resulting in a poor overall fit, χ^2 (540) = 2817.20, $p < .001$, RMSEA = .071, CFI = .83, $SRMR_{WITHIN}$ = .075, $SRMR_{BETWEEN}$ = .290. Similar to the within-teacher level model, this baseline model appeared to poorly explain the observed correlation between items 13 and 14, indicating that a correlation between the residuals of these items may be required. In tests of cluster bias, however, residual variances on the between-teacher level have to be fixed at zero, while constraining the factor loadings at the within- and between-teacher level to be equal (Jak et al., 2014). To obtain an estimate of this residual covariance, we therefore re-parameterized the measurement model by allowing items 13 and 14 to load on an additional factor, which is uncorrelated to the four student-specific self-efficacy domains. Moreover, we fixed the factor loading of these two items at one, such that the obtained factor variance equals the estimate of the residual covariance (Jak, 2014).

Although the re-parameterized, fully constrained four-factor model significantly improved on the baseline model, TRd (2) = 307.09, ΔCFI = .02, it did not converge to an admissible solution and yielded an unacceptable fit, χ^2 (538) = 2559.23, $p < .001$, RMSEA = .069, CFI = .85, $SRMR_{WITHIN}$ = .075, $SRMR_{BETWEEN}$ = .283. Generally, the pattern of discrepancies between the model and the data indicated that strong factorial invariance across teachers does not hold. Moreover, the substantial factor correlations (see Table 3) suggested that models with fewer latent factors might provide a more plausible alternative. Based on the model's parameters and theory, we therefore successively fitted a one-factor model, a three-factor model in which the IS and SE domains were combined, and a three-factor model in which the ES and SE domains were combined. Neither the one-factor solution, TRd (12) = 2255.24, ΔCFI = .20, nor both three-factor alternatives, TRd (6) = 74.66, ΔCFI = .01; TRd (6) = 94.72, ΔCFI = .01, significantly improved the model's fit.

Given that TSE likely resides along a continuum from domain generality to domain- and student specificity, we explored whether the four specified domains of teachers' self-efficacy towards particular students may be accounted for by one common underlying higher-order construct of General Self-Efficacy. This model with four first-order factors and one second-order factor showed no convergence problems and had a slightly better fit than the one-factor, TRd (5) = 2740.44, ΔCFI = .00, and three-factor alternatives, TRd (1) = 12.44, ΔCFI = .00; TRd (1) = 41.80, ΔCFI = .00.

Taking the model with four first-order factors and one second-order factor as a baseline, we subsequently tested the significance of the between-teacher level residual variances. Based on the modification indices, we successively freed 18 of 24 residual variances, resulting in a statistically significant improvement of model fit, TRd (18) = 2818.49, ΔCFI = .05. Further improvement of fit was established by allowing the factor loadings of nine items (4, 7, 11, 13, 14, 15, 17, 18, 19) to be freely estimated across teachers. These factor loadings were all more indicative of higher student-specific TSE at the between-teacher level, suggesting that the domains of TSE, and especially Student Engagement, do not have the same interpretation across teachers. Hence, these violations of measurement invariance across clusters suggest that the domains of student-specific TSE at the between-teacher level cannot merely be assumed to be the within-teacher level factor's aggregates. The final, partially constrained model, had an acceptable fit to the data, χ^2 (518) = 1864.92, $p < .001$, RMSEA = .056, CFI = .90, $SRMR_{WITHIN}$ = .068, $SRMR_{BETWEEN}$ = .152. The standardized factor loadings of the final model are depicted in Fig. 1.

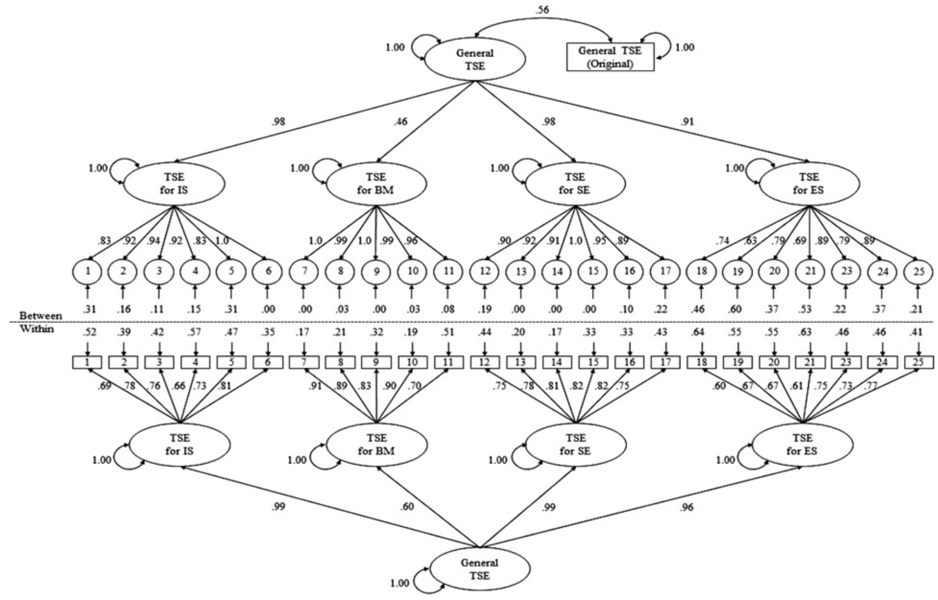
6.2.4. Step 4: Evaluating the concurrence between the generalized and student-specific TSES

To investigate the concurrence between the generalized and student-specific TSES, we allowed the total score of the generalized TSES to correlate with the second-order common self-efficacy factor at the between-teacher level of the final model from step 3 (see Fig. 1). Addition of this correlation resulted in a satisfactory model fit, χ^2 (541) = 1941.95, $p < .001$, RMSEA = .055, CFI = .90; $SRMR_{WITHIN}$ = .068, $SRMR_{BETWEEN}$ = .151. Although the chi-square value of this model indicated a statistically significant lack of fit, the CFI of .90 was reasonable, and the RMSEA of .055 and $SRMR_{WITHIN}$ of .068 were smaller than Hu and Bentler's (1999) cutoff value of .08, suggesting acceptable fit. The $SRMR_{BETWEEN}$ value of .151 indicated that the component fit of the between part was slightly worse than the within part of the model. This poorer fit at the between-level has been noted by previous research as well (cf., Dyer et al., 2005). Assessment of the correlation coefficient pointed to a statistically association between generalized TSE and teachers' student-specific TSE, $r = .59$, $p < .001$). This association suggests a moderate correspondence between the original TSES and the adapted, student-specific TSES.

Table 3
Estimated correlations for the latent factors.

	1	2	3	4	5
1. TSE for instructional strategies	1.00				
2. TSE for behavior management	.59	1.00			
3. TSE for student engagement	.98	.60	1.00		
4. TSE for emotional support	.95	.57	.95	1.00	
5. General TSE	.99	.60	.99	.96	1.00

Note. All correlations are statistically significant ($p < .001$). TSE = teacher self-efficacy.



Note. Parameter estimates are standardized and statistically significant ($p < .001$). For reasons of parsimony, the residual correlation between items 13 and 14 are not displayed in the model.

Fig. 1. Final model of teachers' sense of domain- and student-specific self-efficacy.

7. Discussion

Since long, empirical studies have mainly equated teachers' sense of self-efficacy with a relatively stable omnibus trait that generates inter-individual differences between teachers. Following the basic tenets of social cognitive theory, however, TSE could also be considered to embody domain-linked cognitive states that depend on challenges presented by particular students (e.g., Bandura, 1997; Tschannen-Moran et al., 1998). As such, a premium has been placed on the effort to disentangle within-teacher fluctuations in TSE across various teaching tasks, domains, and students (Raudenbusch et al., 1992; Ross et al., 1996; Tschannen-Moran et al., 1998). The present study is one of the first to come to grips with trait and state-variability in TSE, by evaluating these capability beliefs in relation to particular students, and across various domains of teaching and learning. Recognizing the existence of both inter- and intra-individual differences in TSE has important theoretical and practical implications for the investigation of TSE.

7.1. Domain specification of TSE

In line with prior theory and research (e.g., Bandura, 1997; Lent & Brown, 2006; Tschannen-Moran et al., 1998; Tschannen-Moran & Woolfolk Hoy, 2001), we hypothesized teachers' self-efficacy beliefs to reside along a continuum from domain generality to domain specificity. The present study's findings generally afforded credence to this idea. Initially, evidence was found for the presence of a single, higher-order construct that potentially reflects teachers' generalized sense of self-efficacy. This common factor of general TSE took the commonality among the lower-order domains of self-efficacy into consideration, thereby providing a strong rationale for the unidimensional total score of these capability beliefs. In their seminal study, Tschannen-Moran and Woolfolk Hoy (2001) have also found evidence for such a second-order construct of teacher self-efficacy, which accounted for 75% of the variance and showed a high internal consistency ($\alpha = .94$).

In our study, the substantial factor loadings of the generalized teacher self-efficacy factor indicated that between 21% and 98% of the variance is shared between the TSE domains at the lower level of the structural hierarchy. Still, the markedly poorer fit of a first-order single factor solution, as well as several other plausible alternatives, suggested that specific dimensions of TSE can be distinguished. Markedly, the strongest support was found for the unique domain of behavior management, which evaluates the extent to which teachers feel able to promote positive behavior in a particular child. The interrelationships between this factor and other domains of self-efficacy were moderate, suggesting that tasks and capabilities related to behavior management may be relatively distinct from other core responsibilities, such as providing the instructional, motivational, and emotional supports that generate gains in learning. As such, these results substantiate previous findings from related studies, in which the classroom management domain was also found to be the most distinctive (e.g., Fives, Hamman, & Olivarez, 2007; Tschannen-Moran & Woolfolk Hoy, 2001). The potential uniqueness of the behavior management factor may explain, in part, why this domain of TSE has increasingly gained popularity among educational researchers as a separate field of study (cf., Emmer & Hickman, 1991; O'Neill & Stephenson, 2011).

The final second-order model's factor structure also provided evidence for the existence of the unique TSE domains of instructional strategies and student engagement. These domains tap into teachers' perceived capability to use various instructional methods that enable and enhance individual students' learning, and activate their interest in their schoolwork. Notably, the inter-factor correlations between TSE for instructional strategies and student engagement appeared to be the highest, which is consistent with previous empirical findings (e.g., Tschannen-Moran & Woolfolk Hoy, 2001; Tsigilis et al., 2010). Following classroom-based research (e.g., Hamre & Pianta, 2005), these strong links may be explained in terms of the important role teachers' instructional strategies play in making content relevant, meaningful, and enjoyable to their students. Thereby, such skills and capabilities may set the stage for students' motivation and engagement in schoolwork, and may play a key role in enhancing students' knowledge and skills (Hamre & Pianta, 2005; Hamre et al., 2013; Hardré & Sullivan, 2009).

From a methodological viewpoint, the high correspondence among the instructional strategies and student engagement domains can also be explicated by the less stable structure of the SE factor in prior studies. The factor analytic results from Wolters and Daugherty (2007), for instance, revealed a pattern of cross-loadings of items related to the student engagement domain that was indicative of poor discriminant validity between the instructional strategies and student engagement subscales of the original TSES. Moreover, Henson (2002) noted that caution should be exercised when using scores from the student engagement subscale, as the evidence for the existence of the third domain of the original TSES is far from conclusive. Further large-scale research using the student-specific TSES is therefore needed to verify the uniqueness of the self-efficacy domains of instructional strategies and student engagement.

Apart from the three domains proposed by Tschannen-Moran and Woolfolk Hoy (2001), the student-specific TSES also appeared to be targeted to teachers' emotional support. Comparison of the four-factor solution with the original three-factor model suggested that teachers' self-efficacy for emotional support can be distinguished separately from other domains of TSE. Yet, this dimension of self-efficacy also corresponded highly with domains of instructional strategies and student engagement. It might well be that teachers with a strong sense of self-efficacy for emotional support are generally better attuned and responsive to individual students' needs, ideas, and thoughts. Theoretical and empirical work from Hamre et al. (2013, 2014) substantiates this notion, suggesting that the strategies teachers use to foster students' learning and engagement in the classroom are likely to be based on individual students' basic, affective needs for relatedness, autonomy, and competence. This sensitivity to students' perspectives might explain why these considerable associations were found in the present study.

Adding the emotional support dimension to the extant domains of TSE may be of particular importance for studies investigating outcomes related to teaching and learning. A sizeable literature has provided evidence, both theoretically and empirically, that sensitive and emotionally supportive teachers may provide students with experiences that foster their motivation and learning outcomes in the classroom (Crosnoe et al., 2004; Hamre et al., 2014; Pianta, La Paro, Payne, Cox, & Bradley, 2002; Roeser, Eccles, & Sameroff, 2000). Moreover, teachers' emotional support has frequently been shown to reduce the risk of low-quality student–teacher relationships, especially for students who display uncontrollable or disruptive behavior (Ahnert, Pinquart, & Lamb, 2006; Buyse, Verschueren, Doumen, Van Damme, & Maes, 2008; Hamre & Pianta, 2005; La Paro, Pianta, & Stuhlman, 2004). Building self-efficacy around the domain of emotional support may therefore advance further understanding of the multifaceted ways in which teachers' self-percepts of efficacy function.

Taken together, the overall, higher-order factor of TSE seems to account for substantial amounts of variance shared by the four hypothesized domains of self-efficacy beliefs. As such, it may be compelling to expand the original structure of the TSES by adding one higher-order dimension, without losing sight of the relevance and potential independence of the four TSE domains of teaching and learning. These domains remain essential, both theoretically and practically, yet their commonality is not negligible. Thus, adapting the hierarchical structure of the student-specific TSES, which suggests a continuum from domain generality to domain and student specificity, may potentially advance our understanding of the nature of teachers' sense of efficacy.

7.2. Inter- and intra-individual differences in teachers' self-efficacy

Results of this study indicated that the adapted, student-specific TSES may be suitable for capturing both inter- and intra-individual differences in TSE. Generally, there was significant state and trait variability for each of the model's items. Intraclass correlations showed that the variability at the state (within-teacher) level was larger than at the trait (between-teacher) level. These larger within-teacher differences mirror the social cognitive view that teachers' self-efficacy beliefs, despite reflecting some degree of trait variability, may vary across realms of activity, situational demands, and characteristics of the students toward whom their behaviors and actions are directed (Bandura, 1997; Tschannen-Moran et al., 1998).

Notably, the within-teacher variability seemed to be the largest in teachers' student-specific self-efficacy for behavior management. There might be several reasons for the smaller amount of variation in TSE for behavior management at the between-teacher level. First, this lack of variability might in part be attributable to the process of revising the original TSES. Three out of eight items related to classroom management were removed from the adapted instrument, as these could not be accurately made specific to the level of individual students, or overlapped too substantially with other items. As a consequence, the domain of behavior management seemed to reflect a greater focus on student behavior issues, thereby concentrating less on classroom routines and organization of time and resources (e.g., Emmer & Stough, 2001; O'Neill & Stephenson, 2011). Second, teachers' beliefs about their capability to deal with individual students' classroom behaviors may depend more heavily than other TSE beliefs on interpersonal aspects of teaching. Prior research suggests that teachers tend to appraise individual students' behavior on the basis of relationship beliefs, feelings, and expectations, which usually stem from teachers' previous affective experiences and day-to-day interactions with the child (Bandura, 1997; Spilt & Koomen, 2009; Stuhlman & Pianta, 2002). Whereas positive appraisals may lead teachers to believe in their capabilities to positively affect the child's behavior, negative appraisals may thwart teachers' self-efficacy for behavior management and subsequent behavior towards this child (ibid.). Arguably, teachers who doubt their ability to effectively deal with individual students' behaviors may unintentionally convey poor expectations and ideals, thereby potentially further stimulating undesirable behavior and attributes in the child and confirming their already poor efficacy beliefs. Hence, further research on the reciprocal relationships between teachers' appraisals of individual students' behavior and relationship representations is needed to explain fluctuations in teachers' self-efficacy across individual students.

To meaningfully compare variations in domain- and student-specific TSE between teachers, we tested for violations of measurement invariance over clusters. Recent research (e.g., Jak et al., 2013, 2014; Muthén & Asparouhov, 2013; Ryu, 2014) underscored the necessity of using this relatively new, but complex technique, as it attempts to take account of differences in response processes that may result from personal and contextual characteristics, while still allowing for comparisons of groups on similar latent variables.

In the present study, cluster bias was detected in 18 of 24 items, with the lowest amount of bias found in the behavior management items and, after that, the emotional support items. The partial absence of cluster bias in those items might be due to the smaller amount of variance in these items across teachers. Furthermore, nine additional factor loadings could not be considered equal across educators. These factor loadings were all indicative of higher TSE at the between-teacher level, especially with respect to the domain of student engagement. Hence, (the domains of) TSE at the between-teacher level cannot be merely perceived as the aggregate of within-teacher level self-efficacy beliefs (Jak et al., 2013, 2014), which is also reflected in the moderate correlation between the total score of the original TSES and the second-order common self-efficacy factor at the between-teacher level.

Importantly, the presence of cluster bias in teachers' self-efficacy underscores the complexity of purely estimating these elusive capability beliefs. Both teachers' and students' idiosyncratic characteristics and behaviors are likely to shape a unique classroom environment that ultimately affects how teachers judge and interpret their own sense of efficacy. There is some literature to suggest, for instance, that teachers' knowledge and provision of instructional strategies may be dependent on prior education, years of teaching experience, and satisfaction with past performance (Tschannen-Moran & Woolfolk Hoy, 2001, 2007). Contextual factors, including school and classroom climate, principal leadership, student behavior, available teaching materials, and collective efficacy have also been proposed as sources of teachers' self-efficacy (Goddard & Goddard, 2001; Hipp & Bredeson, 1995; Moore & Esselman, 1992; Tschannen-Moran & Woolfolk Hoy, 2001). Variability in such distinct features may potentially lead to inconsistencies in self-efficacy reports across teachers, which are not accounted for by the common factor structure. This indicates that two teachers with equal values on the latent domain(s) of self-efficacy but from different classrooms are likely to vary in their expected observed test score (Jak et al., 2013, 2014).

7.3. Limitations

The present study's results should be interpreted in light of several limitations. First, the generalizability of our findings remains to be established across teachers and classrooms. Although our sample appeared to be largely comparable to the larger population of Dutch schools, teachers, and students, it primarily consisted of female teachers with relatively high levels of teaching experience. Moreover, a small amount of participating teachers (5.3%) dropped out before data collection as a result of long-term sickness, strenuous workloads, or burnout. This dropout might have given rise to both non-response bias and bias across clusters, suggesting that teachers may report different self-efficacy scores, despite having similar beliefs in their capability. Following Bandura's (1997) notions of triadic reciprocal causality, it may be reasonable to assume that teachers' responses to individual items do not only rely on their self-efficacy beliefs for a specific domain and/or student, but also on personal characteristics and the context in which teachers operate. Yet, such biases across teachers might have implications for the psychometric quality of self-efficacy measures, as well as the interpretation of inter- and intra-individual differences in TSE. An important next step for future research, therefore, is to explore the explaining factors underlying the cluster bias by including features of both teachers and the classroom as potentially biasing attributes. Moreover, additional tests for measurement invariance across (subgroups of) teachers may warrant considerations in future research, to establish whether observed differences in teachers' reports reflect systematic response biases across teachers, or substantive differences in TSE per se.

Second, and in a related vein, participating students in this study were predominantly Dutch, and had relatively high socioeconomic backgrounds. Probably, the nature of the student sample might have affected teachers' responses on the student-specific TSES. Indeed, previous research has suggested that teachers may hold different self-efficacy beliefs in relation to different students, depending on students' demographic backgrounds and behaviors (e.g., Raudenbusch et al., 1992; Ross et al., 1996; Spilt & Koomen, 2009; Spilt, Koomen, & Thijs, 2011). In any attempt to replicate the results, it is therefore recommended that future researchers consider individual student characteristics as covariates of teachers' sense of student-specific self-efficacy.

Third, it should be noted that the response rate among schools invited to participate was very low. This low response rate may have biased the present study's results, since schools with self-efficacious teachers and an open mind to research were probably more likely to take part than schools with already full agendas or strenuous workloads. Nonetheless, a sincere attempt was made to increase the response rate among teachers within the participating schools, by rewarding participation with school reports containing a conceptual overview of the study's results and gift vouchers. As a result, more than 70% of the teachers was willing to participate, which may to some extent compensate for the low participation rate among schools.

Fourth, analytic techniques such as multilevel structural equation modeling are subject to assumptions of multivariate normality of continuous data. In the present study, several student-specific TSES items were found to be skewed, indicating potential non-normality of the data. Essentially, violations of assumptions of multivariate normality may result in bias in the model's parameter estimates and fit indices. However, the size of our sample was substantial, and robust maximum likelihood was used to deal with the non-normality of some student-specific TSES items.

Fifth, the student-specific TSES was filled out by participating teachers for a limited number of randomly selected students, thereby possibly raising questions of selective bias. It should be noted, however, that Snijders and Bosker (1999) have demonstrated that inclusion of all students from each classroom is insensible and needless when the cluster size of the sample is sufficient, as is the case in the present study. Moreover, including the full amount of students per class would have made the data gathering process excessively time-consuming burdensome for teachers.

7.4. Implications for research and practice

Despite these shortcomings, the present study may provide some promising avenues for further research and practice in the field of teaching and learning. First, the present study generally maintains the view that unique TSE domains of teachers' functioning can be distinguished, but may also converge to an overarching construct of general teacher self-efficacy. Although more research is evidently needed to refine and further confirm the instrument's distinct dimensions, the adapted TSES may be already relevant for educational researchers and practitioners alike. Specifically, our new instrument might provide meaningful and relevant profiles of teachers' self-efficacy judgments across various domains of functioning, each of which require specific knowledge, skills, and competencies (cf., Bandura, 1997). Uncovering such distinctive patterns of TSE across teaching tasks and domains may help school psychologists in their quest to develop intervention strategies for a myriad of sources that may influence teachers' sense of efficacy and associated performance in instructional, affective, and behavioral teaching domains. For instance, helping teachers to selectively focus on their performance attainments and to monitor their physiological reactions to inefficacious control of difficult students or challenging teaching tasks may raise teachers' self-efficacy for teaching domains in which they feel less confident (Bandura, 1997).

Second, the student-specific TSES is one of the first measures to empirically support the social cognitive view that TSE is a multifaceted phenomenon that fluctuates over teaching domains and particular students (Bandura, 1997; Tschannen-Moran et al., 1998). Compared to the original TSES, which exclusively focuses on inter-individual differences in TSE across domains, our instrument seems well suited to grasp teachers' unique sense of self-efficacy in relation to different students as well. This interpersonal view on TSE may be relevant for understanding teachers' differential treatment of particular students in class, and fluctuations in the affective quality of dyadic student–teacher relationships. Existing research has increasingly encouraged such an interpersonal focus of analysis to comprehend the mechanisms behind teachers' and students' behaviors and actions in the classroom (Pianta, Hamre, & Stuhlman, 2003; Spilt et al., 2011). In light of this recommendation, the unique domain of emotional support seems a meaningful addition to the construct of teacher self-efficacy. Insights into teachers' sense of self-efficacy towards individual children in this particular domain may help school psychologists to coach teachers in emotionally connecting with, and getting through to particular students in class.

Third, and in a related vein, the present study's results seem to underscore the importance of investigating teachers' self-efficacy in relation to the particular context in which they perform their daily job. Although there are some relatively stable, trait aspects of TSE, these capability beliefs cannot be merely perceived as context-free attributes of a teacher. Rather, features of the classroom context, individual students, and teachers themselves may play an important role in producing fluctuations in teachers' capability beliefs. For future researchers, it may be a challenge to uncover important variables that may further explain variations in self-efficacy between and within teachers.

Lastly, self-efficacy measures that are tailored to various teaching domains and specific students may increase the predictive power of the self-efficacy construct, and potentially afford better explanation of teachers' supportive behaviors and students' school adjustment in the classroom (Bandura, 1986, 1997). To date, evidence regarding the consequences of TSE for students' and teachers' classroom performances seems far from conclusive (e.g., Klassen et al., 2011). Probably, the lack of particularized instruments has, in large part, prevented improvement in interpretations of these complex relationships. Measuring teachers' self-efficacy in relation to individual students may provide a rich context for understanding and interpreting teachers' differential treatment of, and day-to-day interactions with, particular students in the classroom. Educational researchers and practitioners such as school psychologists may use these insights to develop training programs and interventions targeting teacher' student-specific efficacy beliefs as a means of improving students' outcomes, and especially those at risk of academic failure.

Acknowledgments

This research was supported by grant 411-12-036 from the Netherlands Organization for Scientific Research.

Appendix A. Appendix

Table 1
Original and student-specific TSES items.

Domain	Item	Original TSES	Domain	Item	Student-specific TSES
IS	1	To what extent can you use a variety of assessment strategies?			-
IS	2	To what extent can you provide an alternative explanation or example when students are confused?	IS	1	To what extent can you provide an alternative explanation or example when this student is confused?
IS	3	To what extent can you craft good questions for your students?	IS	2	To what extent can you craft stimulating questions for this student?
IS	4	How well can you implement alternative strategies in your classroom?	IS	3	How well can you let this student apply alternative problem solving strategies?
IS	5	How well can you respond to difficult questions from your students?		–	
IS	6	How much can you do to adjust your lessons to the proper level for individual students?	IS	4	How well can you adjust your lessons to the proper level for this student?
IS	7	To what extent can you gauge student comprehension of what you have taught?	IS	5	To what extent can you gauge this student's comprehension of what you have taught?
IS	8	How well can you provide appropriate challenges for very capable students?	IS	6	How well can you provide appropriate challenges for this student?
CM	9	How much can you do to control disruptive behavior in the classroom?	BM	7	How well can you control disruptive behavior in this student?
CM	10	How much can you do to get children to follow classroom rules?	BM	8	How well can you get this student to follow classroom rules?
CM	11	How much can you do to calm a student who is disruptive or noisy?	BM	9	How well can you calm this student when he/she is disruptive or noisy?
CM	12	How well can you establish a classroom management system with each group of students?			-
CM	13	How well can you keep a few problem students from ruining an entire lesson?	BM	10	How well can you prevent this student from negatively affecting the classroom atmosphere?
CM	14	How well can you respond to defiant students?			-
CM	15	To what extent can you make your expectation clear about student behavior?	BM	11	To what extent can you make your behavioral expectations clear to this student?
CM	16	How well can you establish routines to keep activities running smoothly?			-
SE	17	How much can you do to get students to believe they can do well in schoolwork?	SE	12	How well can you get this student to believe he/she can do well in schoolwork?
SE	18	How much can you do to help your students value learning?	SE	13	To what extent can you help this student to value learning?
SE	19	How much can you do to motivate students who show low interest in schoolwork?	SE	14	To what extent can you motivate this student for his/her schoolwork?
SE	20	How much can you assist families in helping their children do well in school?			-
SE	21	How much can you do to improve the understanding of a student who is failing?	SE	15	How well can you help this student to understand the learning content?
SE	22	How much can you do to help your students think critically?	SE	16	How well can you help this student to think critically?
SE	23	How much can you do to foster student creativity?	SE	17	To what extent can you help this student to explore new things?
SE	24	How much can you do to get through to the most difficult students?	SE	18	How well can you get through to this student?
			ES	19	How well can you respond positively and sincerely to this student in the classroom?
			ES	20	To what extent can you provide positive feedback to this student?
			ES	21	How well can you provide a safe and secure environment for this student?
			ES	22	To what extent can you timely recognize that this student does not feel well?
			ES	23	How well can you timely provide support to this student?
			ES	24	To what extent can you provide this student with the space to make his/her own choices?
			ES	25	To what extent can you adjust learning tasks to this student's needs and interests?

References

Ahnert, L., Pinquart, M., & Lamb, M. E. (2006). Security of children's relationships with nonparental care providers: A meta-analysis. *Child Development*, *77*, 664–679. http://dx.doi.org/10.1111/j.1467-8624.2006.00896.x.

Almog, O., & Shechtman, Z. (2007). Teachers' democratic and efficacy beliefs and styles of coping with behavioural problems of pupils with special needs. *European Journal of Special Needs Education*, *22*, 115–129. http://dx.doi.org/10.1080/08856250701267774.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191–215. http://dx.doi.org/10.1037/0033-295X.84.2.191.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman.

Bandura, A. (2000). Exercise of human agency through collective efficacy. *Current Directions in Psychological Science*, *9*, 75–78. http://dx.doi.org/10.1111/1467-8721.00064.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares, & T. Urdan (Eds.), *Adolescence and education. Self-efficacy and adolescence*, *Vol. 5*. (pp. 307–337). Greenwich, CT: Information Age.

Bentler, P. M. (1992). On the fit of models to covariances and methodology to the bulletin. *Psychological Bulletin*, *112*, 400–404. http://dx.doi.org/10.1037/0033-2909.112.3.400.

Brouwers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education*, *16*, 239–253. http://dx.doi.org/10.1016/S0742-051X(99)00057-8.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Buyse, E., Verschueren, K., Doumen, S., Van Damme, J., & Maes, F. (2008). Classroom problem behavior and teacher–child relationships in kindergarten: The moderating role of the classroom climate. *Journal of School Psychology*, *46*, 367–391. http://dx.doi.org/10.1016/j.jsp.2007.06.009.

Cantrell, P., Young, S., & Moore, A. (2003). Factors affection science teaching efficacy of preservice elementary teachers. *Journal of Science Teacher Education*, *14*, 177–192. http://dx.doi.org/10.1023/A:1025974417256.

CBS Statline (2015a). (Speciaal) basisonderwijs; culturele minderheden, (achterstands)leerlingen. [(Special) primary education; cultural minorities, (at-risk) students]. Retrieved from http://statline.cbs.nl/Statweb/

CBS Statline (2015b). (Speciaal) basisonderwijs en speciale scholen; leerlingen, schoolregio. [(Special) primary education and special schools; students, school region]. Retrieved from http://statline.cbs.nl/Statweb/

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504. http://dx.doi.org/10.1080/10705510701301834.

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, *41*, 189–225. http://dx.doi.org/10.1207/s15327906mbr4102_5.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255. http://dx.doi.org/10.1207/S15328007SEM0902_5.

Coladarci, T., & Breton, W. A. (1997). Teacher efficacy, supervision, and the special education resource-room teacher. *The Journal of Educational Research*, *90*, 230–239. http://dx.doi.org/10.1080/00220671.1997.10544577.

Crosnoe, R., Johnson, M. K., & Elder, G. H. (2004). Intergenerational bonding in school: The behavioral and contextual correlates of student–teacher relationships. *Sociology of Education*, *77*, 60–81. http://dx.doi.org/10.1177/003804070407700103.

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*, 751–766. http://dx.doi.org/10.1016/j.tate.2007.02.010.

Diefenbach, M. A., Weinstein, N. D., & O'Reilly, J. (1993). Scales for assessing perceptions of health hazard susceptibility. *Health Education Research*, *8*, 181–192.

DUO (2014). Leeftijd van personeel in het primair onderwijs. [Age of employees in primary education]. Retrieved from http://www.onderwijsincijfers.nl/kengetallen/primair-onderwijs/personeelpo/leeftijd-personeel

Dyer, N. G., Hanges, P. J., & Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The Leadership Quarterly*, *16*, 149–167. http://dx.doi.org/10.1016/j.leaqua.2004.09.009.

Emmer, E., & Hickman, J. (1991). Teacher efficacy in classroom management and discipline. *Educational and Psychological Measurement*, *51*, 755–765. http://dx.doi.org/10.1177/0013164491513027.

Emmer, E. T., & Stough, L. M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist*, *36*, 103–112. http://dx.doi.org/10.1207/S15326985EP3602_5.

Fives, H., Hamman, D., & Olivarez, A. (2007). Does burnout begin with student-teaching? Analyzing efficacy, burnout, and support during the student-teaching semester. *Teaching and Teacher Education*, *23*, 916–934. http://dx.doi.org/10.1016/j.tate.2006.03.013.

Friedman, I. A., & Kass, E. (2002). Teacher self-efficacy: A classroom-organization conceptualization. *Teaching and Teacher Education*, *18*, 675–686. http://dx.doi.org/10.1016/S0742-051X(02)00027-6.

Ghaith, G., & Yaghi, M. (1997). Relationships among experience, teacher efficacy and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education*, *13*, 451–458. http://dx.doi.org/10.1016/S0742-051X(96)00045-5.

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

Goddard, R. D., & Goddard, Y. L. (2001). A multilevel analysis of teacher and collective efficacy. *Teaching and Teacher Education*, *17*, 807–818. http://dx.doi.org/10.1016/S0742-051X(01)00032-4.

Guo, Y., McDonald Connor, C., Yang, Y., Roehring, 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*, 3–24. http://dx.doi.org/10.1086/665816.

Hamre, B. K., & Pianta, R. L. (2005). Can instructional and emotional support in the first grade classroom make a difference for children at risk of school failure? *Child Development*, *76*, 949–967. http://dx.doi.org/10.1111/j.1467-8624.2005.00889.x.

Hamre, B. K., Pianta, R. C., Downer, J. T., Decoster, J., Jones, S., Brown, J., Cappella, E., ... Kaefer, T. (2013). Teaching through interactions: testing a developmental framework of effective teaching in over 4,000 classrooms. *The Elementary School Journal*, *113*, 461–487. http://dx.doi.org/10.1086/669616.

Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher–child interactions: Associations with preschool children's development. *Child Development*, *85*, 1257–1284. http://dx.doi.org/10.1111/cdev.12184.

Hardré, P. L., & Sullivan, D. W. (2009). Motivating adolescents: high school teachers' perceptions and classroom practices. *Teacher Development*, *13*, 1–16. http://dx.doi.org/10.1080/13664530902858469.

Heneman, H. G., III, Kimball, S., & Milanowski, A. (2006). The teacher sense of efficacy scale: validation evidence and behavioral prediction. *WCER Working Paper No. 2006-7*. Wisconsin Center for Education Research (NJ1).

Henson, R. K. (2002). From adolescent angst to adulthood: Substantive implications and measurement dilemmas in the development of teacher efficacy research. *Educational Psychologist*, *37*, 137–150. http://dx.doi.org/10.1207/S15326985EP3703_1.

Hipp, K. A., & Bredeson, P. V. (1995). Exploring connections between teacher efficacy and principals' leadership behavior. *Journal of School Leadership*, *5*, 136–150.

Hornstra, L., van der Veen, I., Peetsma, T., & Volman, M. (2013). Developments in motivation and achievement during primary school: A longitudinal study on group-specific differences. *Learning and Individual Differences*, *23*, 195–204. http://dx.doi.org/10.1016/j.lindif.2012.09.004.

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: conventional versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. http://dx.doi.org/10.1080/10705519909540118.

Jak, S. (2014). Testing strong factorial invariance using three-level structural equation modeling. *Frontiers in Psychology*, *5*, 1–7. http://dx.doi.org/10.3389/fpsyg.2014.00745.

Jak, S., Oort, F. J., & Dolan, C. V. (2013). A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling*, *20*, 265–282. http://dx.doi.org/10.1080/10705511.2013.769392.

Jak, S., Oort, F. J., & Dolan, C. V. (2014). Measurement bias in multilevel data. *Structural Equation Modeling*, *21*, 31–39. http://dx.doi.org/10.1080/10705511.2014.856694.

Justice, L. M., Mashburn, A. J., Hamre, B. K., & Pianta, R. C. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly*, *23*, 51–68. http://dx.doi.org/10.1016/j.ecresq.2007.09.004.

Kagan, D. M. (1990). Ways of evaluating teacher cognition: Inferences concerning the goldilocks principle. *Review of Educational Research*, *60*, 419–469. http://dx.doi.org/10.3102/00346543060003419.

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

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

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

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford.

La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The Classroom Assessment Scoring System: Findings from the prekindergarten year. *The Elementary School Journal*, *104*, 409–426.

Lambert, R. G., McCarthy, C., O'Donnell, M., & Wang, C. (2009). Measuring elementary teacher stress and coping in the classroom: Validity evidence for the Classroom Appraisal of Resources and Demands. *Psychology in the Schools*, *10*, 973–988. http://dx.doi.org/10.1002/pits.20438.

Lent, R. W., & Brown, S. D. (2006). On conceptualizing and assessing social cognitive constructs in career research: A measurement guide. *Journal of Career Assessment*, *14*, 12–35. http://dx.doi.org/10.1177/1069072705281364.

Leroy, N., Bressoux, P., Sarrazin, P., & Trouilloud, D. (2007). Impact of teachers' implicit theories and perceived pressures on the establishment of an autonomy supportive climate. *European Journal of Psychology of Education*, *22*, 529–545. http://dx.doi.org/10.1007/BF03173470.

Moore, V., & Esselman, M. (1992, April). Teacher efficacy, empowerment, and focused instructional climate: Does student achievement benefit? *Paper presented at the annual meeting of the American Educational Research Association, San Francisco*.

Morris-Rothschild, B. K., & Brassard, M. R. (2006). Teachers' conflict management styles: The role of attachment styles and classroom management efficacy. *Journal of School Psychology*, *44*, 105–121. http://dx.doi.org/10.1016/j.jsp.2006.01.004.

Muthén, B., & Asparouhov, T. (2013). New methods for the study of measurement invariance with many groups. *Technical report* (Retrieved from http://www.statmodel.com).

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, *22*, 376–398. http://dx.doi.org/10.1177/0049124194022003006.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

O'Neill, S. C., & Stephenson, J. (2011). The measurement of classroom management self efficacy: A review of measurement instrument development and influences. *Educational Psychology*, *31*, 261–299. http://dx.doi.org/10.1080/01443410.2010.545344.

Oort, F. J. (1992). Using restricted factor analysis to detect item bias. *Methodika*, *6*, 150–166.

Pajares, F. (1996, April). Assessing self-efficacy beliefs and academic outcomes: The case for specificity and correspondence. *Paper presented at the annual meeting of the American Educational Research Association, New York*.

Pajares, F. (1997). Current directions in self-efficacy research. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *International advances in self research* (pp. 1–49). Greenwich, CT: Information Age Publishing.

Pianta, R. C., Hamre, B., & Stuhlman, M. (2003). Relationships between teachers and children. In W. M. Reynolds, & G. E. Miller (Eds.), *Handbook of psychology: Educational psychology*, *Vol. 7*. (pp. 199–234). Hoboken, NJ: Wiley.

Pianta, R. C., La Paro, K., & Hamre, B. K. (2008). *Classroom assessment scoring system*. Baltimore, MD: Brookes.

Pianta, R. C., La Paro, K. M., Payne, C., Cox, M. J., & Bradley, R. (2002). The relation of kindergarten classroom environment to teacher, family, and school characteristics and child outcomes. *The Elementary School Journal*, *102*, 225–238.

Raudenbusch, S. W., Rowan, B., & Cheong, Y. F. (1992). Contextual effects on the self-perceived efficacy of high school teachers. *Sociology of Education*, *65*, 150–167.

Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology*, *104*, 700–712. http://dx.doi.org/10.1037/a0027268.

Riggs, I. M., & Enochs, L. G. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education*, *74*, 625–637. http://dx.doi.org/10.1002/sce.3730740605.

Rimm-Kaufman, S. E., & Chui, Y. J. I. (2007). Promoting social and academic competence in the classroom: An intervention study examining the contribution of the Responsive Classroom approach. *Psychology in the Schools*, *44*, 397–413. http://dx.doi.org/10.1002/pits.20231.

Rimm-Kaufman, S. E., La Paro, K. M., Downer, J. T., & Pianta, R. C. (2005). The contribution of classroom setting and quality of instruction to children's behavior in kindergarten classrooms. *The Elementary School Journal*, *105*, 377–394. http://dx.doi.org/10.1086/429948.

Roeser, R. W., Eccles, J. S., & Sameroff, A. J. (2000). School as a context of early adolescents' academic and social-emotional development: A summary of research findings. *The Elementary School Journal*, *100*, 443–447.

Roorda, D. L., Koomen, H. M., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher–student relationships on students' school engagement and achievement a meta-analytic approach. *Review of Educational Research*, *81*, 493–529. http://dx.doi.org/10.3102/0034654311421793.

Ross, J. A. (1994). The impact of an in-service to promote cooperative learning on the stability of teacher efficacy. *Teaching and Teacher Education*, *10*, 381–394.

Ross, J. A., Cousins, J. B., & Gadalla, T. (1996). Within-teacher predictors of teacher efficacy. *Teaching and Teacher Education*, *12*, 385–400.

Ryu, E. (2014). Factorial invariance in multilevel confirmatory factor analysis. *British Journal of Mathematical and Statistical Psychology*, *67*, 172–194. http://dx.doi.org/10.1111/bmsp.12014.

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.), *Innovations in multivariate statistical analysis. A Festschrift for Heinz Neudecker* (pp. 233–247). London, UK: Kluwer Academic Publishers.

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*, 243–248. http://dx.doi.org/10.1007/s11336-009-9135-y.

Schwarzer, R. (Ed.). (1992). *Self-efficacy: Thought control of action*. Washington, DC: Hemisphere.

Schwarzer, R., & Jerusalem, M. (1995). Generalized self-efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio* (pp. 35–38). Windsor: NFER-Nelson.

Schwarzer, R., Schmitz, G. S., & Daytner, G. T. (1999). The Teacher Self-Efficacy scale. Retrieved from http://www.fu-berlin.de/gesund/skalen/t_se.htm

Siwatu, K. O. (2011). Preservice teachers' culturally responsive teaching self-efficacy forming experiences: A mixed methods study. *The Journal of Educational Research*, *104*, 360–369. http://dx.doi.org/10.1080/00220671.2010.487081.

Siwatu, K. O. (2007). Pre-service teachers' culturally responsive teaching self-efficacy and outcome expectancy beliefs. *Teaching and Teacher Education*, *23*, 1086–1101. http://dx.doi.org/10.1016/j.tate.2006.07.011.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London, UK: Sage Publishers.

Soodak, L. C., & Podell, D. M. (1996). Teacher efficacy: toward the understanding of a multi-faceted construct. *Teaching and Teacher Education*, *12*, 401–411.

Spilt, J. L., & Koomen, H. M. Y. (2009). Widening the view on teacher–child relationships: Teachers' narratives concerning disruptive versus nondisruptive children. *School Psychology Review*, *38*, 86–101.

Spilt, J. L., Koomen, H. M. Y., & Thijs, J. T. (2011). Teacher wellbeing: The importance of teacher–student relationships. *Educational Psychology Review*, *23*, 457–477. http://dx.doi.org/10.1007/s10648-011-9170-y.

Stuhlman, M. W., & Pianta, R. C. (2002). Teachers' narratives about their relationships with children: Associations with behavior in classrooms. *School Psychology Review*, *31*, 148–163.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn and Bacon.
Tschannen-Moran, M., & Johnson, D. (2011). Exploring literacy teachers' self-efficacy beliefs: Potential sources at play. *Teaching and Teacher Education*, *27*, 751–761. http://dx.doi.org/10.1016/j.tate.2010.12.005.
Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*, 783–805. http://dx.doi.org/10.1016/S0742-051X(01)00036-1.
Tschannen-Moran, M., & Woolfolk Hoy, A. (2007). The differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and Teacher Education*, *23*, 944–956. http://dx.doi.org/10.1016/j.tate.2006.05.003.
Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, *68*, 202–248. http://dx.doi.org/10.3102/00346543068002202.
Tsigilis, N., Koustelios, A., & Grammatikopoulos, V. (2010). Psychometric properties of the Teachers' Sense of Efficacy Scale within the Greek educational context. *Journal of Psychoeducational Assessment*, *28*, 153–162. http://dx.doi.org/10.1177/0734282909342532.
Tsouloupas, C. N., Carson, R. L., Matthews, R., Grawitch, M. J., & Barber, L. K. (2010). Exploring the association between teachers' perceived student misbehaviour and emotional exhaustion: The importance of teacher efficacy beliefs and emotion regulation. *Educational Psychology*, *30*, 173–189. http://dx.doi.org/10.1080/01443410903494460.
Wheatley, K. F. (2005). The case for reconceptualizing teacher efficacy research. *Teaching and Teacher Education*, *21*, 747–766. http://dx.doi.org/10.1016/j.tate.2005.05.009.
Wolters, C. A., & Daugherty, S. G. (2007). Goal structures and teachers' sense of efficacy: Their relation and association to teaching experience and academic level. *Journal of Educational Psychology*, *99*, 181–193.
Woolfolk Hoy, A., & Burke-Spero, R. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education*, *21*, 343–356. http://dx.doi.org/10.1016/j.tate.2005.01.007.
Wyatt, M. (2014). Towards a re-conceptualization of teachers' self-efficacy beliefs: Tackling enduring problems with the quantitative research and moving on. *International Journal of Research & Method in Education*, *37*, 166–189. http://dx.doi.org/10.1080/1743727X.2012.742050.
Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, *30*, 165–200. http://dx.doi.org/10.1111/0081-1750.00078.
Zee, M., Koomen, H. M. Y., & van der Veen, I. (2013). Student–teacher relationship quality and academic adjustment in upper elementary school: The role of student personality. *Journal of School Psychology*, *51*, 517–533. http://dx.doi.org/10.1016/j.jsp.2013.05.003.