Teachers' and principals' perceptions of school emphasis on academic success: measurement invariance, agreement, and relations to student achievement

Wangqiong Ye[1,2]* , Ronny Scherer[1,2] and Sigrid Blömeke[1,2]

*Correspondence:
Wangqiong Ye
wangqiong.ye@cemo.uio.no
[1]Centre for Research on Equality in Education (CREATE), Faculty of Educational Sciences, University of Oslo, Postbox 1161, Oslo, Norway
[2]Centre for Educational Measurement (CEMO), Faculty of Educational Sciences, University of Oslo, Postbox 1161, Oslo 0318, Norway

Abstract

A school's emphasis on academic success (SEAS) is a crucial aspect of school climate that can be rated by different informants, such as teachers or principals, and with respect to their perception of different groups' emphasis on academic success, such as parents or students. We know little about whether there is agreement between these ratings. Utilizing Israeli Trends in International Mathematics and Science Study (TIMSS) 2019 data and multilevel structural equation modeling, we examined how teachers and principals rate teachers', students', and parents' emphasis on academic success and how these ratings are related to student achievement. We found that (a) ratings are invariant across informants; (b) teachers and principals agree on parents' and students' yet not teachers' emphasis; (c) teacher ratings explain more variation in student achievement than principal ratings. Since SEAS ratings diverge by informants and relations to student achievement, researchers need to craft validity arguments in their studies.

Keywords School climate, School emphasis on academic success, Student achievement, Multilevel structural equation modeling, Measurement invariance, Rater agreement

Introduction

School emphasis on academic success is an essential aspect of school climate as it is positively associated with student achievement and well-being, leading researchers to focus on understanding the role of school climate in promoting positive student outcomes (Nilsen & Gustafsson 2014). The Trends in International Mathematics and Science Study (TIMSS) measures school emphasis on academic success (SEAS) as a critical component of school climate. SEAS refers to school principals' and teachers' reports of to what extent students, parents, and teachers at their schools strive for high student achievement (Mullis & Martin 2017).

Although teachers and principals were asked to rate SEAS levels within their schools, previous studies have rarely discussed potential differences between their ratings. Studies have commonly utilized teachers' perceptions of SEAS when investigating teaching and learning in the classroom (Ker 2016). In contrast, researchers have tended to rely on principals' perceptions of SEAS when examining the school or policy level (Bellens et al., 2020). Ramsey and colleagues (2016) conducted a study to investigate the differences in perceptions of school climate dimensions, including schools' emphasis on academic success, among students, parents, and school staff (including teachers, administrators, and assistant principals). They found that school staff rated academic emphasis lower than parents and students. In contrast, parents rated parental involvement the lowest, and students reported the lowest perception of safety and connectedness, indicating that informants differed in their perception of school climate (Ramsey et al. 2016).

These findings call attention to the agreement between multiple informants and raise concerns about possible reporting bias in results on SEAS. With parallel SEAS items, TIMSS data provided an opportunity to examine rater agreement between teachers and principals. The Israeli data, which included three teachers per school and sufficient rating responses (over 97.5% for principals and 98.5% for teachers on SEAS items), was used to investigate the concordance between teachers' and principals' perceptions of the emphasis placed by teachers, parents, and students on academic success. Additionally, in an attempt to construct a valid argument for either perception, we investigated the correlation between different perceptions and the mathematics achievement of students.

Conceptualizing school emphasis on academic success

Many studies within the field of educational effectiveness research have focused on identifying malleable school factors that are associated with student achievement, with the aim of improving educational outcomes and closing achievement gaps (Scheerens 2017). Among these studies, research consistently observed that, among other factors, the school climate plays a crucial role in shaping the quality of interactions within a school (Rudasill et al. 2018). Moreover, it reflects the shared values held within a school community and has a pervasive impact on virtually every aspect of students' school experience (Thapa et al., 2013). Thus, school climate has caught wide research attention because of its potential to improve student achievement and reduce problematic behavior or student dropout (Wang & Degol 2016).

Although there is no consensus on how to conceptualize school climate in detail, the importance of a school's emphasis on academic success as one subdomain of school climate is well recognized (Thapa et al., 2013; Wang & Degol 2016). Studies have revealed positive relationships of it to student achievement across cultures (Wu et al. 2013) and subjects (Park et al. 2019). Moreover, Hoy et al. (2006) showed that academic emphasis is closely related to collective efficacy and faculty trust by defining the construct "Academic Optimism" as "a shared belief among faculty that academic achievement is important, that the faculty has the capacity to help students achieve, and that students and parents can be trusted to cooperate with them in this endeavor" (p. 204). "Academic emphasis" refers to how learning and teaching are promoted in a school—for instance, emphasizing the commitment to high academic standards, expectations that encourage students to do their best, leadership from teachers and principals directed towards improving student outcomes, and social interactions focusing on academic activities (Wang & Degol

2016). Based on this conceptualization, Martin and colleagues (2013) created the TIMSS construct called School Emphasis on Academic Success (SEAS), which was "indicated by rigorous curricular goals, effective teachers, students that desired to do well, and parent support."

In detail, SEAS is assessed in TIMSS by school principals' and teachers' ratings of students', parents', and teachers' expectations for student achievement and their commitment to supporting high achievement through various means (Mullis & Martin, 2017). The scales were implemented for the first time in TIMSS 2011 (Martin et al. 2013) and changed slightly in TIMSS 2015 and 2019 based on conceptual and empirical considerations (Mullis & Martin, 2017).

Earlier studies utilizing TIMSS data revealed that SEAS was positively related to student achievement and motivation (Wang & Degol 2016). Nilsen and Gustafsson (2014) explored the impact of SEAS as a subdomain of school climate on science performance in Norway, finding that SEAS helped to improve students' performance in all science subdomains. Other studies examining the relationship between SEAS, safety, orders in school, and instructional quality also found positive associations (Scherer & Nilsen 2016).

Dimensions of SEAS and their relation to student achievement

Discussions about academic emphasis mainly focus on three groups—teachers, students, and parents. But only a few studies have looked into whether these form different dimensions of SEAS and thus violate the unidimensionality assumption underlying many studies. In a preliminary approach, Gustafsson and Nilsen (2016) adopted five items from teacher questionnaires in TIMSS 2007 and 2011 to measure SEAS. They indeed found signs of multidimensionality. Badri (2019) used 13 SEAS items from the principal questionnaire in TIMSS 2015 and came to a similar conclusion. Thus, these studies are the first indications that one may have to distinguish how different groups' emphasis is perceived at a school.

While a positive relation between SEAS and student outcomes has been established on a general level (Martin et al. 2013), the associations between the dimensions of SEAS pertaining to parents, students, and teachers, and their impact on academic outcomes, have been largely unexplored by researchers. We found only two studies summarized above (Badri 2019; Gustafsson & Nilsen 2016). Gustafsson and Nilsen (2016) found that parental support had a strong effect, student desire had a smaller positive influence, but teacher items had no impact on student achievement. Badri (2019) found that the items related to parents and students had a significant impact on student achievement, while the items related to teachers did not demonstrate any significant influence.

Choosing the appropriate level of analysis

Teachers' data in TIMSS were collected by providing a questionnaire to individual teachers at a school and asking for their respective responses. However, the primary purpose of collecting individuals' ratings of SEAS is to assess the *school* climate (Mullis & Martin, 2017). SEAS is thus a collective property, and the appropriate level of analysis would be the school level (Marsh et al. 2012), based on aggregated individual perceptions of the teachers (Hoy et al., 2006). Ratings at the school level reflect teachers' average perception of SEAS, while individual differences are considered sources of error (Lüdtke et al.,

2009). Hence, the primary interpretation of SEAS should be based on the proper aggregation of individual responses (Marsh et al., 2012).

However, few studies have examined SEAS at the school level; most have adopted teachers' ratings and focused on the classroom level (Nilsen & Gustafsson 2014). These studies did not deal with potential differences between school environments but with individual differences in teacher perceptions (Lüdtke et al. 2009). One possible explanation for this practice is that using teachers' individual ratings may be easier methodologically and more convenient to handle in the modeling. However, this method raises conceptual and methodological concerns. Among others, measurement invariance and multilevel reliability issues need to be considered. It is worth noting that, in most case, there was only one math teacher per classroom in each school in TIMSS data, which may lead to the practice ignoring the difference between school and classroom levels. Besides, since teachers were not chosen randomly but rather selected based on the random selection of the classroom, an individual teacher's assessment of SEAS may not accurately reflect the entire school.

Measuring SEAS from multiple informants

Besides taking into account the potential multidimensionality of SEAS and choosing the appropriate level of analysis, there is another consideration in measuring SEAS: Whose ratings should be used to receive a valid picture? In TIMSS, both teachers and principals are asked to provide information about the implementation of curricula, teachers' ability to inspire students, parents' involvement, and their expectations for student achievement (Mullis & Martin, 2017). Since SEAS is collective school property, both principals' and teachers' ratings may matter in its assessment.

Moreover, multi-informant ratings may reflect various layers and dimensions of a context, in our case, SEAS, and thus help reduce the impact of biases and random errors, which is at the heart of crafting a validity argument for a construct (Wagner et al. 2010). Teachers and principals play varying roles in the school, and their perceptions of SEAS may differ due to their differential interactions with students and the environment. For instance, teachers are nested in schools, and their perceptions of SEAS may be affected by principals to a certain degree (Hallinger 2018). It is well known that principals play a crucial role in shaping the learning environment in schools (Hallinger 2018). Consequently, how principals perceive SEAS within the school may also affect teachers' perceptions (Sanchez et al. 2020). Meanwhile, teachers' perceptions of principals' behaviors also depend on their interactions with the principal (Hallinger 2018). Therefore, a multilevel framework of teachers' perceptions, both nesting and interacting with principals' perceptions, may illuminate construct validity and further our understanding of SEAS.

Although both teachers' and—rarely—principals' perceptions were used to explore the impact of SEAS, their congruence has rarely been discussed. However, the importance of gathering data from multiple informants—and being aware of potential discrepancies—has been highlighted in the literature about aggregated characteristics. Kuger and Klieme (2016) emphasized the importance of assessing school climate using various sources, such as teachers' and principals' ratings. Holzberger and Schiepe-Tiska (2021) used both teacher and principal ratings to evaluate the relationship between school context and instructional quality. Finding that teachers' perceptions of school context were more relevant to instructional quality than those of principals, they recommended

measuring school characteristics with various sources on both classroom and school levels. Discrepancies in ratings across groups have been found despite seemingly similar experiences, suggesting that respondents may have disparate perceptions (Mitchell et al., 2010).

The present study

The present study examines teachers' and principals' perceptions of SEAS in a multilevel framework and investigates their relationship to student achievement in mathematics, aiming to contribute to crafting a validity argument and understanding the role of SEAS in school improvement. In this context, we address methodological challenges associated with school research to provide reliable and valid findings.

Taking advantage of the parallel SEAS measures for teachers and principals in the TIMSS 2019 data, we examined teachers' and principals' perceptions of parents', students', and teachers' emphasis on academic success in terms of their invariance, agreement, and relations to students' mathematics achievement. Specifically, we address the following research questions (RQs):

RQ1. To what extent are the measures of SEAS invariant between teachers and principals?

RQ2. To what extent are teachers' and principals' perceptions of parents', students', and teachers' emphasis on academic success correlated?

RQ3. To what extent are teachers' and principals' perceptions of parents', students', and teachers' emphasis on academic success related to student achievement?

Methods

Sample and procedure

TIMSS 2019 followed a two-stage stratified cluster sampling design in which schools were randomly sampled. Within schools, one or more intact classes were then selected from the target grades (Mullis & Martin, 2017). The sample utilized for the present study comprises Israeli eighth-grade students who participated in TIMSS 2019 ($N=3582$), along with their teachers ($n=418$) and principals ($k=156$). We removed student data without teachers' or principals' ratings of SEAS from the sample (0.5–3.0%). On average, nine students were nested in one classroom, and three classrooms were nested in one school. In addition, 47 students with two or more mathematics teachers were removed from the data to secure a clear-cut student-teacher link. On average, teachers had 15.7 years of experience in teaching ($SD=10.1$), and principals had nine years of experience in their jobs ($SD=7.6$).

Measures

School emphasis on academic success (SEAS)

Teachers and principals were asked to respond to the same 11 items measuring SEAS (see Table 1). Items 1 to 4 referred to teachers, items 5 to 8 to parents, and 9 to 11 to students. Teachers and principals rated the SEAS items in five categories, from 1 (*very high*) to 5 (*very low*). Response codes were reversed for the present study so that a higher score represented a higher level of perceived SEAS. For the measurement properties of the scales, please refer to the Results section.

Table 1 Descriptive statistics for items assessing "Schools' Emphasis on Academic Success (SEAS)"

	SEAS items	Variable	Teachers (N = 418)			Variable	Principals (N = 156)		
			M	SD			M	SD	
1	Teachers' understanding of the school's curricular goals	BTBG06A	4.23	0.71		BCBG14A	4.17	0.65	
2	Teachers' degree of success in implementing the school's curriculum	BTBG06B	3.95	0.75		BCBG14B	3.96	0.58	
3	Teachers' expectations for student achievement	BTBG06C	3.86	0.79		BCBG14C	4.01	0.66	
4	Teachers' ability to inspire students	BTBG06D	3.82	0.80		BCBG14D	3.58	0.64	
5	Parental involvement in school activities	BTBG06E	3.24	0.97		BCBG14E	3.24	0.85	
6	Parental commitment to ensure that students are ready to learn	BTBG06F	3.04	0.90		BCBG14F	3.22	0.93	
7	Parental expectations for student achievement	BTBG06G	3.84	0.78		BCBG14G	3.95	0.73	
8	Parental support for student achievement	BTBG06H	3.41	0.80		BCBG14H	3.55	0.80	
9	Students' desire to do well in school	BTBG06I	3.57	0.76		BCBG14I	3.82	0.60	
10	Students' ability to reach school's academic goals	BTBG06J	3.47	0.71		BCBG14J	3.66	0.62	
11	Students' respect for classmates who excel academically	BTBG06K	3.67	0.77		BCBG14K	3.97	0.61	

Note Ratings were recoded to (1) very low – (5) very high

Students' mathematics achievement

TIMSS 2019 included more than 200 items requiring students to use knowing, application, and reasoning skills in mathematics (Mullis & Martin, 2017). The eighth-grade mathematics test had four content domains, with 30% of the assessment devoted to the number domain (i.e., ratios), 30% to algebra (i.e., equations), 20% to geometry (i.e., geometric shapes), and 20% to data and probability.

Israel participated in TIMSS 2019 in a digital format (eTIMSS) for the eighth-grade population and ranked 9th among the 39 participating countries (average score of 519). The Israeli education system is characterized by its complexity, with schools catering to diverse racial and religious groups and resource allocation varying across schools in response to disparities in economic status (Agbaria 2018). Due to the relatively lower requirement for specialized resources (e.g., lab equipment) in mathematics compared to subjects like science, we opted to utilize data from the TIMSS mathematics assessment for our study.

To estimate students' proficiency, TIMSS created plausible values (PVs) based on students' responses to the items and conditioned on all available background data (Laukaityte & Wiberg 2017). Specifically, TIMSS used population models to estimate distributions of proficiencies and a latent regression of the proficiency on contextual data, and the estimation of the regression is conducted separately for each country (Martin et al. 2020). We used all five plausible values, conducted the analyses for each of the five PVs, and combined the results via Rubin's (1987) combination rules.

Statistical analyses

Intraclass correlations and reliability

First, we computed the intraclass correlations ICC(1) and ICC(2) to determine the proportion of variance at the school level and, respectively, to evaluate whether the aggregated ratings were reliable indicators of group-level constructs (Lüdtke et al. 2009).

Second, we estimated level-specific reliability coefficients for each SEAS dimension and the entire scale. With teachers nested in schools, overall reliability estimates may be biased due to violations of the independent residual assumption, and hence, level-specific reliability estimates are preferred over single-level reliability estimates (Geldhof et al. 2014). In our study, we therefore estimated level-specific Cronbach's Alpha and McDonald's Omega. To estimate the latter, we used Lai's (2021) corrected formulas for individual-level constructs.

Measurement models and model fit

Testing invariance is vital in evaluating whether items measure the same attributes for different respondent groups (Putnick & Bornstein, 2016). Since teachers and principals are groups at distinct levels, measurement invariance testing in this study includes two parts—the invariance of the SEAS measurement models across levels (i.e., classroom vs. school level) and across groups (i.e., teachers vs. principals). Because the traditional multi-group confirmatory factor analysis method is not feasible for this purpose, multi-level structural equation modeling provides a solution to separate the levels of analysis and evaluate within- and between-group differences, respectively (Muthén, 2011).

For the between-level construct SEAS, representing the aggregate of teachers' ratings within each school, cross-level constraints are required (Stapleton, Yang, & Hancock, 2016). The within and between levels should share the same factor structure and equivalent factor loadings (Jak & Jorgensen, 2017). Therefore, cross-level invariance for teachers' perceptions of SEAS was established as the first step. Subsequently, measurement invariance across informants was tested at the school level, as suggested by Kim et al. (2018). Specifically, we estimated a series of five multi-group confirmative analysis models (MCFA) with equality constraints and compared the deterioration in model fit. The first model (MCFA Model 1) assumed configural invariance across levels for teacher ratings and configural cross-group invariance without any parameter equality constraints. The second model (MCFA Model 2) assumed cross-level metric invariance with equal factor loadings of teacher ratings across levels, yet freely estimated parameters of the measurement model describing principal ratings. The third model (MCFA Model 3) contained metric cross-level and cross-group invariance constraints. The fourth model added equal intercepts at the school level between teacher and principal ratings and thus assumed cross-group scalar invariance. Finally, the fifth model was based on MCFA Model 3 and added equal factor covariances across informants (see Fig. 1).

To evaluate the model fit, we used relative fit indices: the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA), and the Standardized Root Mean Square Residual (SRMR). This study followed the following common guidelines: CFI≥0.90, RMSEA≤0.05, and SRMR<0.08 for an acceptable overall fit of the model (Marsh, Hau, & Grayson 2005). Moreover, ΔCFI<0.01 and ΔRMSEA<0.015 were adopted as the criteria to test measurement invariance (Kim et al. 2018). However, we did not consider these guidelines as fixed rules but applied some flexibility, as they depend on, for instance, the number of factors, the size of factor loadings, the levels of analysis, the sample-size ratios, and the complexity of the analytic models (Rappaport et al., 2020).

As Asparouhov and Muthén (2018) noted, if the number of clusters is less than 200, SRMR values at the between level may be high, even when the model fits reasonably
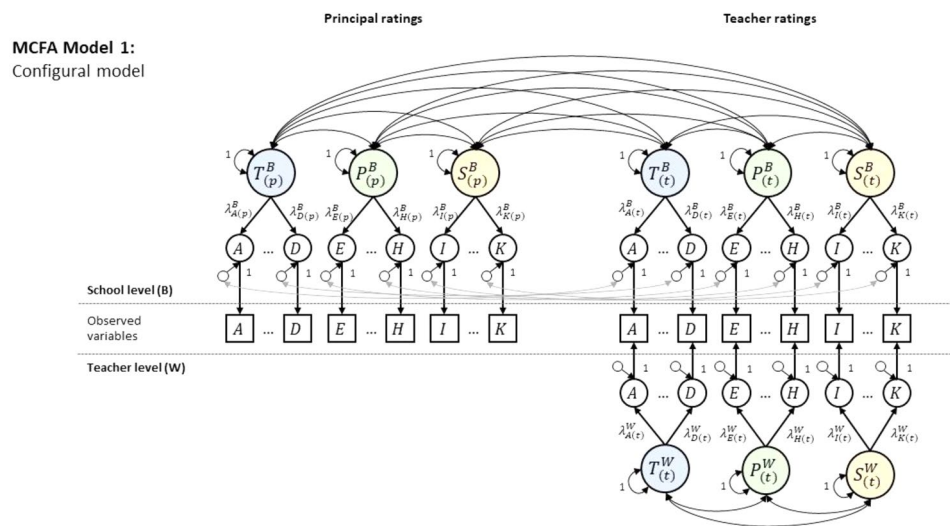
Fig. 1 MCFA Model 1 Teachers' and Principals' Perceptions of SEAS. *Note* Circles represent latent variables; boxes represent manifest variables; *(t)* and *(p)* refers to teachers' and principals' ratings; superscripts *B* and *W* refer to the between or within level; *T*, *P*, and *S* refer to the teacher, parent, and student sub-dimension of SEAS; $T_{(p)}^B$ = principals' ratings of the Teacher Sub-dimension on the between level; $T_{(t)}^B$ = teachers' ratings of the Teacher Sub-dimension on the between level; $T_{(t)}^W$ = teachers' ratings of the Teacher Sub-dimension on the within level; A = item BTBG06A in teachers' rating or BCBG14A in principals' rating

well. Considering the number of schools in our analysis, the 0.08 threshold for SRMR at the between level is too strict. We followed the suggestion of Asparouhov and Muthén (2018) and conducted chi-square tests to confirm the model fitting at the between level.

Multilevel structural equation modeling (MSEM)

This study also used the MSEM framework for evaluating the relations between SEAS and student achievement via three-level models, with student data at level 1, teacher ratings at levels 2 and 3 (aggregated), and principal ratings at level 3. We extended the multilevel CFA models with acceptable model fit and parameter constraints by structural models at the teacher and the school level using students' mathematics achievement as an outcome variable (see Fig. 2).

Missing data and sampling weights

We used maximum likelihood estimation with robust standard errors (MLR) and handled missing data with the full-information-maximum-likelihood procedure. TIMSS data have ready-to-use scaled weights prepared for a single-level analysis but are not suitable for multilevel modeling (Rutkowski et al. 2010). Therefore, we followed the method proposed by Rutkowski et al. (2010) and used the manually rescaled student, classroom, and school weights in our analyses. The Israeli TIMSS 2019 data are openly available and can be downloaded from the International Association for the Evaluation of Educational Achievement (IEA) international TIMSS database at https://timss2019.org/international-database/. We prepared the data using the R software Version 4.0.2 (R Core Team, 2019) and the R package *intsvy* (Caro & Biecek, 2017) and conducted all analyses using M*plus* Version 8.8 (Muthén & Muthén, 2017).
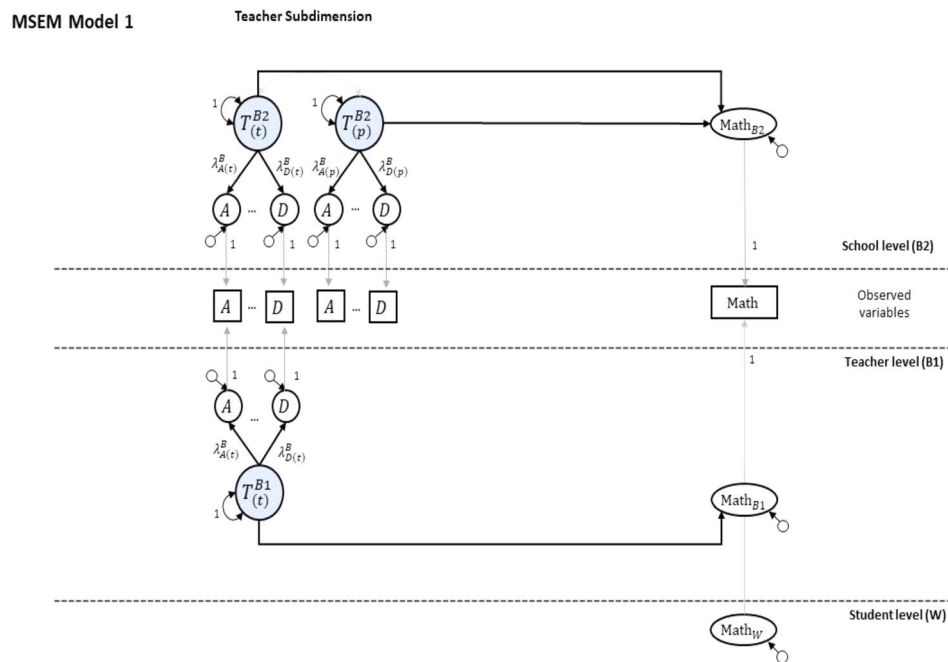
Fig. 2 MSEM Model 1 Relation Between the Teacher Sub-dimension and Students' Mathematics Achievement. *Note* Circles represent latent variables; boxes represent manifest variables. $T^{B2}_{(p)}$ = Principals' ratings of the Teacher Sub-dimension at the school level; $T^{B2}_{(t)}$ = Teachers' ratings of the Teacher Sub-dimension at the school level; $T^{B1}_{(t)}$ = Teachers' ratings of the Teacher Sub-dimension at the classroom level; A = item BTBG06A in teachers' rating or BCBG14A in principals' rating; $Math_W$ = Math achievement at the individual level, B1 = classroom level, B2 = school level

Results

Variability and reliability

Despite minor variations in the descriptive statistics, there was a notable similarity between the ratings of SEAS items provided by both teachers and principals (see Table 1). Analysis of the SEAS assessment data revealed that the teacher-related dimension consistently contained the two highest-rated items, while the parent-related dimension consistently contained the three lowest-rated items. Notably, principals assigned higher ratings to items within the student-related dimension compared to teachers.

A necessary precondition for multilevel analyses is sufficient variability at different levels, which can be expressed in the intraclass correlation ICC(1). In line with Hedges and Hedberg (2007), this study found ICC(1) values sufficiently variably since they are in the range of 0.15 to 0.32 (see Table 2). Item 5 from the parent dimension exhibited the highest proportion of between-school variance in teacher ratings (ICC[1] = 0.32). Item 3 from the teacher dimension showed the lowest between-school variability (ICC[1] = 0.14). Based on the relative agreement among teachers and an average of 2.7 teacher ratings per school, ICC(2) values were above 0.75, indicating excellent reliability of average ratings (LeBreton & Senter, 2008).

Next, we examined the reliabilities of SEAS dimensions. In principals' ratings, Cronbach's α estimates were acceptable for the dimensions related to teachers and parents, but only barely for the students (α = 0.69). If the more unbiased indicator, McDonald's Omega, was considered, the reliability for the student dimension was acceptable (ω = 0.72). In teachers' ratings, between-level reliabilities were higher than the within-level estimates for dimensions related to parents and students (see Table 3).

Table 2 Intraclass correlations of teachers' perceptions of SEAS dimensions

	Teacher dimension				Parent dimension				Student dimension		
	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11
ICC(1)	0.19	0.16	0.14	0.16	0.32	0.26	0.29	0.22	0.24	0.25	0.15
ICC(2)	0.84	0.81	0.78	0.81	0.91	0.89	0.90	0.87	0.88	0.88	0.80

Note On average, 2.7 teachers per school were included in the data set

Research question 1: Measurement invariance between informants

As noted earlier, we estimated a series of MCFA models testing different invariance constraints. The fit indices of these models and the respective model comparisons are shown in Table 4. Adding cross-level metric invariance constraints to the teacher ratings did not deteriorate the model fit significantly (MCFA Models 1 vs. 2). Similarly, the fit indices remained stable when incorporating the cross-group metric invariance constraints (MCFA Model 3). Hence, metric invariance across levels and informants could be assumed.

However, when imposing the quality of intercepts to the school-level variables (scalar measurement invariance; MCFA Model 4), the model fit deteriorated, $\Delta\chi^2(5)=116.9$, $p<.001$, Δ CFI = -0.016. As a consequence, we cannot assume that average ratings of SEAS were invariant between teachers and principals, and latent mean differences may thus not be reported.

Research question 2: Agreement between teachers and principals

Regarding the agreement between teachers and principals, we examined the factor correlations at the school level, utilizing the MCFA Model 3. Table 5 shows the resultant correlation matrices. First, we investigated the correlations among the three SEAS dimensions for the teacher and the principal ratings, respectively. The findings indicated strong positive correlations among the three dimensions of teachers' ratings, with coefficients ranging from 0.86 to 0.96. Of note, the highest correlation was observed between teachers' perceptions of parents' and students' prioritization of academic success ($r=.96$), whereas the lowest correlation was identified between teachers' perceptions of their own and students' emphasis on academic success ($r=.86$). For principals, the correlations among the three dimensions were generally lower ($0.60 \leq r \leq .87$). In line with teachers' perceptions, the strongest relationship was found between the student and parent dimensions ($r=.87$). The weakest, however, was between the teachers' and parents' dimensions ($r=.60$).

To further test if these factor correlations among the SEAS dimensions were the same for teachers and principals, we constrained the factor correlations in MCFA Model 3 to be equal across informants (see Table 4). The resultant MCFA Model 5 assumed structural invariance and deteriorated the model fit, $\Delta\chi^2(3)=12.7$, $p<.01$, Δ SRMR-level 3=0.049. Hence, factor correlation matrices were not the same for the two groups.

Second, we examined the correlations for each dimension across teachers' and principals' ratings at the school level. The correlation between teachers' and principals' ratings was the highest for their perception of parents' emphasis on academic success ($r=.72$). A weaker correlation was found for the perception of students' emphasis ($r=.53$), and only a very low correlation was found for the teacher dimension ($r=.17$). In summary, teachers and principals shared the highest agreement on the parent dimension and the lowest for the teacher dimension.

Research question 3: Relations to student achievement

To evaluate the relationship between SEAS and students' mathematics achievement, we used MSEM Model 2 to ensure the same interpretation of SEAS for teachers and principals. In the following section, the school-level results are reported. Given the strong

Table 3 Cronbach's alpha, and composite reliability for SEAS and its three dimensions

	Principals' Perceptions (Between)		Teachers' Perceptions (Within)		Teachers' Perceptions (Between)	
	α	ω	α	ω	α	ω
TSUB	0.78	0.78	0.73	0.72	0.71	0.74
PSUB	0.84	0.85	0.80	0.81	0.85	0.82
SSUB	0.69	0.72	0.73	0.74	0.85	0.86

Note α=Cronbach's Alpha, ω=McDonald's Omega, T/P/SSUB=teacher/parent/student dimension

Table 4 Model Fit indices for SEAS in three-level CFA models testing the invariance across informants

Three-level CFA	Cross-level Invariance	Cross-group invariance	X^2	df	p	RMSEA	CFI	SRMR-L1	SRMR-L2	SRMR-L3
MCFA Model 1	Configural	Configural	311.377	219	0.000	0.011	0.962	0.000	0.062	0.140
MCFA Model 2	Metric	Configural	329.081	227	0.000	0.011	0.958	0.000	0.053	0.119
MCFA Model 3	Metric	Metric	336.961	235	0.000	0.011	0.958	0.000	0.054	0.124
MCFA Model 4	Metric	Scalar	380.425	240	0.000	0.013	0.942	0.000	0.053	0.121
MCFA Model 5	Metric	Structural	348.215	238	0.000	0.011	0.954	0.000	0.056	0.173

Note CFI=Comparative Fit Index; RMSEA=Root Mean Square Error of Approximation; SRMR-L1, -L2 and -L3=Standardized Root Mean Square Residual at the student, classroom, and school levels, MCFA=Multilevel Confirmatory Factor Analysis; Chi-square tests supported that MCFA models were fitting well despite higher SRMR values at Level 3

Table 5 Correlation matrices for SEAS dimensions at the School Level

		Teacher ratings			Principal ratings		
	Dimensions	Teachers	Parents	Students	Teachers	Parents	Students
Teacher ratings	Teachers	1					
	Parents	0.87*	1				
	Students	0.86*	0.96*	1			
Principal ratings	Teachers	0.17	0.47*	0.49*	1		
	Parents	0.24	0.72*	0.66*	0.60*	1	
	Students	0.18	0.59*	0.53*	0.69*	0.87*	1

Note *. Correlation is significant at the 0.05 level (2-tailed)

correlations among some of the SEAS dimensions, we examined the relations between student achievement and the three SEAS dimensions separately.

The strongest association between SEAS and student outcomes was found for teachers' rating of students' emphasis on academic success (see Table 6). Similarly, principals' rating of students' emphasis on academic success was the strongest predictor for student achievement.

However, teachers' and principals' ratings differ for the other two dimensions. In the case of teachers' ratings, their perception of parents' emphasis on academic success was the weakest predictor for student outcomes, whereas principals' perception of teachers' emphasis was the lowest predictor for students' achievement. Both SEAS ratings were

Table 6 Relations between SEAS and student achievement at the school level

MSEM Model 2	Teacher dimension		Parent dimension		Student dimension	
	Estimate (*SE*)	*p*	Estimate (*SE*)	*p*	Estimate (*SE*)	*p*
BSMMAT ON TSEAS(T/P/S)	0.73* (*0.28*)	0.01	0.70* (*.18*)	0.00	0.73* (*.17*)	0.00
BSMMAT ON PSEAS(T/P/S)	0.00 (*0.18*)	0.98	0.03 (*.16*)	0.86	0.06 (*.19*)	0.75
TSEAS(T/P/S) WITH PSEAS(T/P/S)	0.41* (*.18*)	0.00	0.62* (*.13*)	0.00	0.53* (*.11*)	0.00

Note BSMMAT=Student mathematics achievement (five plausible values), TSEAS=Teachers' perception of SEAS, PSEAS=Principals' perception of SEAS, (T/P/S)=teacher, parent, student dimensions of SEAS, *. Coefficient is significant at the 0.05 level (2-tailed)

significantly related to student achievement only if cross-level and cross-group metric invariance were not assumed. However, in MSEM Model 2, when metric invariance was established, only teachers' ratings of SEAS were significantly associated with the outcome, not the principals'. In summary, teachers' yet not principals' ratings were related to student achievement.

Discussion

Previous studies have revealed a positive relationship between SEAS and student achievement. However, little was known about the validity and reliability of the construct, depending on who provided the information (i.e., teachers or principal), which group at a school the informants were thinking about (i.e., parents or students), and the influence of the assorted SEAS dimensions on educational outcomes. This study aimed to investigate, firstly, to what extent teachers and principals agree in their perceptions of SEAS, secondly, whether there are differences across target groups, and thirdly how the SEAS ratings are related to students' mathematics achievement.

Moreover, using a large-scale data set obtained from TIMSS 2019, this study introduced multilevel confirmatory factor analysis models that allowed measurement invariance testing. Only with the establishment of measurement invariance across levels and groups, it is possible to make meaningful and valid comparisons between teachers' and principals' perceptions. As a component of school climate, SEAS is a school-level construct. However, teachers at a school may vary in their SEAS perceptions. To appropriately model the source of variation, this study followed the suggestion of Stapleton et al. (2016)—treating SEAS as a shared construct at the school level and SEAS as a configural construct at the individual level. Factor loadings were constrained to be equal across levels.

Agreement between perceptions of teachers and principals

SEAS can be evaluated based on different informants; in TIMSS, this is done based on principal and teacher reports. The starting point for our study was that data from multi-informants, especially structurally different groups such as teachers and principals, may provide a more complete perspective of SEAS than just asking one group. Multi-informant data may thus also provide comprehensive insights into the validity of the different group reports (Konold & Cornell 2015). However, a precondition for this would be that cross-group measurement invariance exists, meaning that disparate groups understand the construct similarly. We could not identify any study that had previously examined this question with respect to SEAS and TIMSS data. Previous multi-informant studies

(Konold & Cornell, 2015; Schwab et al. 2020; Wagner et al. 2016) concerned cross-level invariance or cross-group invariance of other school or classroom constructs.

Our study revealed that metric invariance held between teacher and principal ratings—hence, teachers and principals seemed to *understand and interpret* the SEAS similarly. Thus, factor means could be compared in a meaningful way by constraining the factor loadings across the two groups equally (Jak & Jorgensen 2017). However, a similar interpretation or understanding of a construct does not mean that various groups agree on the level of emphasis on academic success in a school. Such an agreement would require scalar measurement invariance which we were not able to confirm based on the data at hand. To phrase it bluntly, teachers and principals may agree on what they are talking about when they think about their school's emphasis on academic success, but not on the extent to which different group actually emphasize it. Indeed, our results indicate that teachers rated SEAS at their schools lower than principals with respect to students' emphasis. This result concurs with Schwab et al. (2020), who concluded that it matters who rates a construct.

Our study also investigated the correlations between teachers' and principals' ratings of the three SEAS dimensions as a way to validate agreements and disagreements. The highest correlation was found for the parent dimension ($r=.72$). Although this can be evaluated as a strong relationship that indicates some agreement, the effect size leaves substantial room for distinctions, given that only half of the variance is explained. While a relatively weaker correlation was observed for the student dimension ($r=.53$), which accounted for only about 30% of the variance, additional inquiries must be made concerning the level of agreement between teachers' and principals' perceptions with regards to the teacher-related dimension. The relationship between principals' and teachers' ratings of teachers' emphasis on academic success was found to be non-significant ($r=.17$, $p=.24$).

We interpret the result with respect to parents and students as an indicator of the different roles played by teachers and principals at a school. For example, teachers are closer to students, and their perceptions of students and parents are based on more direct observations than principals'. On the one hand, these may provide teachers with more accurate information about students' and parents' emphasis on academic success. On the other hand, being closer to students may go along with higher expectations of teachers compared to principals when it comes to parental involvement, commitment, and support or students' desire to do well and to respect their classmates. These differences between principals and teachers could explain that there is some unexplained variance in the correlations.

It is more difficult to understand and explain why principals' and teachers' ratings of teachers' emphasis on academic success are not related at all, in particular, given that no mean differences exist and that the two groups interpret the construct in the same way. Differential expectations and accuracy in observations could play a role here as well. In addition, the reliability of teachers' ratings of their own emphasis is lower than that of other groups' emphasis. However, further research is needed here to fully understand the differences. In any case, the differences not only highlight the unique roles and perspectives of informants but also methodological and practical questions concerning the measurement of SEAS. Similar to the findings of studies exploring the dimensions of other constructs (Konold & Cornell 2015; Wagner et al., 2016), this study found that

agreement varied by dimension, which does not suggest treating SEAS as a unidimensional construct.

While no further studies exist that examine the present research question, our results can be compared with multi-informant studies on other school or classroom constructs. Konold and Cornell (2015) compared teachers' and students' school climate ratings with respect to perceived student support, student engagement, and the prevalence of teasing and bullying. Similar to one of our results, with respect to principals' and teachers' ratings of parents' emphasis on academic success, they found rather strong evidence of convergent validity across the two groups ($r \sim .80$). Wagner et al. (2016) used teachers' and students' ratings to examine dimensions of instructional quality. In line with the wide variety of our results depending on the target group that had to be rated, they found also that the strength of correlations varied by dimension and time point between not significant at all and a maximum of $r = .77$, with higher agreements between teacher and student ratings on classroom management and goal clarity but lower on student support.

Relationship to student achievement

To validate the types of ratings conducted in TIMSS, we investigated the relationship between SEAS and students' mathematics achievement at the school level. Previous studies examined the influence of SEAS at the classroom level (Nilsen & Gustafsson 2014) or in a single-level analysis (Badri, 2019). Unlike these studies, we pointed out the importance of choosing the appropriate analysis level for SEAS. As a school climate characteristic, it is preferred to analyze SEAS at the school level, with teacher ratings aggregated to this level. This also makes it possible to use principals as informants.

Our data revealed that the relationship of SEAS to student outcomes varied by the source of information (teacher or principal) and the dimension (parents', students', or teachers' emphasis on academic success). With respect to all three SEAS dimensions, teacher ratings were significantly and positively related to students' mathematics achievement, while this did not apply to principals. Since predictive validity is a cornerstone of empirical research, we can confirm that teachers provide valid information on SEAS whereas the validity of principals' ratings has to be questioned. More research on this relation is advised.

Notably, at the classroom level, the relationship between teachers' ratings of their own emphasis on academic success and student outcomes was negative ($\beta = -0.02$, $SE = 0.08$, $p = .75$). Although not significant, the estimate points on the one hand to the importance of selecting the appropriate level of analysis. It underscores on the other hand the importance of examining potentially non-linear impacts of SEAS in the future. To some degree, this result corresponds with the finding that the interaction between teachers and students exhibits a non-linear relationship with student outcomes (Aldrup et al. 2022). We would like to point out in this context that the present study in all cases examined linear relations only. It is well possible that in cases where we did not find such relations, non-linear relations in terms of quadratic ("u" curve), cubic (two curves), or other shapes exist.

Previous studies utilizing TIMSS data found teacher ratings of the parent and student dimensions to be the strongest and second-strongest predictors of student outcome (Badri 2019; Nilsen & Gustafsson 2014). Our study not only added principals' perceptions of SEAS but also revealed different results with respect to the relevance of teachers'

perceptions of SEAS for student outcomes. The relationship between teachers' perception of students' emphasis on academic success and student achievement was by far the strongest ($\beta=0.73$), while the impact of the parent dimension was weaker ($\beta=0.70$). Moreover, unlike previous findings, this study found that teachers' rating of their own emphasis also was significantly related to student outcomes ($\beta=0.73$), underscoring the importance of examining SEAS at the appropriate level.

An increasing number of studies investigate the validity of informants' ratings concerning other school or classroom constructs by relating these to external variables, such as student outcomes (Wagner et al. 2010). The results are inconsistent with respect to the validity of students' or teachers' perceptions of school climate respectively. Mitchell et al. (2010) found that classroom-level factors were more closely associated with teachers' perceptions, but school-level factors were more closely associated with students' perceptions. They also found an inverse association between teachers' and students' ratings of academic emphasis. In contrast, Wagner et al. (2016) related teachers' and students' ratings of classroom-level instructional quality to student achievement and found that students' perceptions were more closely associated with the outcomes than teachers'. Maxwell et al. (2017) found that teachers' ratings of school climates were significant predictors of students' numeracy, writing, and reading achievement, while students' ratings were not significant in the case of reading. Overall, there is a need for more research on the validity of multi-informant data on different levels.

Limitations and future directions

Utilizing a large-scale assessment dataset benefited this study in many ways (i.e., multilevel data, parallel scales given to different groups of informants, etc.), but it also limited our research. First, the SEAS assessment in TIMSS 2019 may not have captured all possible aspects of the construct to a sufficient degree. For instance, the assessment included the student, teacher, and parent dimensions yet did not include their interactions. However, current research on school climate points to the importance of these interactions for student learning (Thapa et al., 2013). Moreover, while the items related to parents and students reflect the intended construct in a convincing way so that content validity can be confirmed, some doubts remain regarding the assessment of teachers' emphasis on academic success. While two items are clearly targeting the construct, namely those about expectations and inspiration, the other two that describe the understanding and implementation of the curriculum, may not necessarily be in line. TIMSS is therefore advised to check carefully whether one or two of these items can be changed or at least adjusted a bit.

Second, since this study used only Israeli data, the findings may not be applicable to other contexts. We selected the Israelian sample due to their teacher numbers per school, which is more than most TIMSS 2019 participants. Whether the generalizability of our results across other countries exists, has to be tested systematically. Moreover, the Israeli TIMSS 2019 data contained only three teachers per school on average. Such a limited number of teachers per school is not ideal for analyzing group-level disparities. Future studies comparing teachers' and principals' perceptions of school characteristics, such as the SEAS, are encouraged to use data with larger teacher samples. Additionally, we encourage further studies to investigate the perceptions of teachers and principals, taking into account the size of their schools, as in smaller schools, teachers may

have closer proximity to principals. However, our results served their purpose: Making researchers aware of validity challenges in the case of multi-informant data and potential multi-dimensionality of the SEAS construct.

Third, we drew on teachers' and principals' perceptions of SEAS, including parents' emphasis, yet we could not draw directly on parents' reports because the corresponding TIMSS questionnaire did not include equivalent items. Including such a perspective could have shed further light on the agreements among the different perceptions and ultimately extended the evidence on the scale's validity.

Fourthly, we acknowledge that extraneous factors such as school resources may potentially influence the relationship between SEAS and student achievement. Nonetheless, when considering the Home Educational Resources (HER) index as a potential covariate, which reflects the socioeconomic status (SES) of students' families and is typically aggregated at the classroom or school level, a notable proportion of students (25.88%) were found to have missing data. Given the potential for biased results with the use of a control variable that contains a high proportion of missing data, we elected to analyze the influence of SEAS on student achievement without controlling for school SES.

Conclusion

The findings of this study highlight the importance of choosing appropriate informants for drawing valid inferences. We could confirm metric measurement invariance, which means that teachers and principals agreed on the meaning of SEAS, but the two groups rated the level of parents', students', and teachers' emphasis on academic success differently. Conclusions drawn from the different results could therefore be very different. Readers need to be informed about this challenge.

Moreover, studies should include a validity argument for selecting one informant group above the other depending on the research question at hand. In our case, teachers' perceptions were found to be more relevant to student outcomes than those of principals. Future studies are advised to select the source of their data carefully. If the relation of SEAS to educational outcomes is the main research question, since one wants to improve these, for example, teachers' perspective may be more relevant than principals'.

Finally, since we could not confirm the one-dimensionality of SEAS, and our models assumed three sub-dimensions from the beginning, it is crucial to direct respondents very precisely to which group they shall provide information. This is an important finding with respect to those creating a SEAS assessment. Otherwise, a risk exists of collecting invalid data.

Declarations

Ethics approval and consent to participate

Consent for publication

Competing interests

References

Agbaria, A. K. (2018). The 'right'education in Israel: Segregation, religious ethnonationalism, and depoliticized professionalism. *Critical Studies in Education*, *59*(1), 18–34. https://doi.org/10.1080/17508487.2016.1185642

Aldrup, K., Carstensen, B., & Klusmann, U. (2022). Is empathy the key to effective teaching? A systematic review of its association with teacher-student interactions and student outcomes. *Educational Psychology Review*, *34*(3), 1177–1216. https://doi.org/10.1007/s10648-021-09649-y

Asparouhov, T., & Muthén, B. (2018). SRMR in Mplus. Retrieved from http://www.statmodel.com/download/SRMR2.pdf

Badri, M. (2019). School emphasis on academic success and TIMSS Science/Math achievements. *International Journal of Research in Education and Science*, *5*(1), 176–189.

Bellens, K., Van den Noortgate, W., & Van Damme, J. (2020). The informed choice: Mathematics textbook assessment in light of educational freedom, effectiveness, and improvement in primary education. *School Effectiveness and School Improvement*, *31*(2), 192–211. https://doi.org/10.1080/09243453.2019.1642215

Caro, D. H., & Biecek, P. (2017). Intsvy: An R Package for analyzing International large-scale Assessment Data. *Journal of Statistical Software*, *81*(7), 1. https://doi.org/10.18637/jss.v081.i07

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

Gustafsson, J. E., & Nilsen, T. (2016). The impact of school climate and teacher quality on mathematics achievement: A difference-in-differences approach. *Teacher Quality Instructional Quality and Student Outcomes*, *2*, 81–95.

Hallinger, P. (2018). Principal instructional leadership: From prescription to theory to practice. *The Wiley Handbook of Teaching and Learning*, 505–528. https://doi.org/10.1002/9781118955901.ch21

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*(1), 60–87. https://doi.org/10.3102/0162373707299706

Holzberger, D., & Schiepe-Tiska, A. (2021). Is the school context associated with instructional quality? The effects of social composition, leadership, teacher collaboration, and school climate. *School Effectiveness and School Improvement*, 1–21. https://doi.org/10.1080/09243453.2021.1913190

Hoy, W. K., Tarter, C. J., & Hoy, A. W. (2006). Academic optimism of schools: A force for student achievement. *American Educational Research Journal*, *43*(3), 425–446. https://doi.org/10.3102/00028312043003425

Jak, S., & Jorgensen, T. D. (2017). Relating measurement invariance, cross-level invariance, and multilevel reliability. *Frontiers in Psychology*, *8*, 1640. https://doi.org/10.3389/fpsyg.2017.01640

Ker, H. (2016). The impacts of student-, teacher-and school-level factors on mathematics achievement: An exploratory comparative investigation of Singaporean students and the USA students. *Educational Psychology*, *36*(2), 254–276. https://doi.org/10.1080/01443410.2015.1026801

Kim, E. S., Wang, Y., & Kiefer, S. M. (2018). Cross-level group measurement invariance when groups are at different levels of multilevel data. *Educational and Psychological Measurement*, *78*(6), 973–997. https://doi.org/10.1177/0013164417739062

Konold, T., & Cornell, D. (2015). Multilevel multitrait–multimethod latent analysis of structurally different and interchangeable raters of school climate. *Psychological Assessment*, *27*(3), 1097. https://doi.org/10.1037/pas0000098

Kuger, S., & Klieme, E. (2016). Dimensions of context assessment. *Assessing contexts of learning* (pp. 3–37). Springer.

Lai, M. H. (2021). Composite reliability of multilevel data: It's about observed scores and construct meanings. *Psychological Methods*, *26*(1), 90. https://doi.org/10.1037/met0000287

Laukaityte, I., & Wiberg, M. (2017). Using plausible values in secondary analysis in large-scale assessments. *Communications in Statistics - Theory and Methods*, *46*(22), 11341–11357. https://doi.org/10.1080/03610926.2016.1267764

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*(4), 815–852. https://doi.org/10.1177/1094428106296642

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modelling. *Contemporary Educational Psychology*, *34*(2), 120–131. https://doi.org/10.1016/j.cedpsych.2008.12.001

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom Climate and Contextual effects: Conceptual and methodological issues in the evaluation of Group-Level effects. *Educational Psychologist*, *47*(2), 106–124. https://doi.org/10.1080/00461520.2012.670488

Martin, M. O., Foy, P., Mullis, I. V., & O'Dwyer, L. M. (2013). Effective schools in reading, mathematics, and science at fourth grade. In M. O. Martin, & I. V. Mullis (Eds.), *TIMSS and PIRLS 2011: Relationships among reading, mathematics, and science achievement at the fourth grade –implications for early learning* (pp. 109–178). TIMSS & PIRLS International Study Center, Boston College.

Martin, M. O., von Davier, M., & Mullis, I. V. (2020). Methods and Procedures: TIMSS 2019 Technical Report. *International Association for the Evaluation of Educational Achievement*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/methods.

Maxwell, S., Reynolds, K. J., Lee, E., Subasic, E., & Bromhead, D. (2017). The Impact of School Climate and School Identification on Academic Achievement: Multilevel modeling with Student and Teacher Data. *Frontiers in Psychology*, *8*(2069). https://doi.org/10.3389/fpsyg.2017.02069

Mitchell, M. M., Bradshaw, C. P., & Leaf, P. J. (2010). Student and teacher perceptions of school climate: A multilevel exploration of patterns of discrepancy. *Journal of School Health*, *80*(6), 271–279.

Mullis, I. V. S., & Martin, M. O. (2017). TIMSS 2019 assessment frameworks. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/

Muthén, B. (2011). Mean and covariance structure analysis of hierarchical data. Retrieved from https://escholarship.org/uc/item/1vp6w4sr

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (Eighth ed.). Muthén & Muthén.

Nilsen, T., & Gustafsson, J. E. (2014). School emphasis on academic success: Exploring changes in Science Performance in Norway between 2007 and 2011 employing two-level SEM. *Educational Research and Evaluation*, *20*(4), 308–327. https://doi.org/10.1080/13803611.2014.941371

Park, J. H., Lee, I. H., & Cooc, N. (2019). The role of school-level mechanisms: How principal support, professional learning communities, collective responsibility, and group-level teacher expectations affect student achievement. *Educational Administration Quarterly*, *55*(5), 742–780. https://doi.org/10.1177/0013161X18821355

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria: https://www.R-project.org/.

Ramsey, C. M., Spira, A. P., Parisi, J. M., & Rebok, G. W. (2016). School climate: Perceptual differences between students, parents, and school staff. *School Effectiveness and School Improvement*, *27*(4), 629–641. https://doi.org/10.1080/09243453.2016.1199436

Rappaport, L. M., Amstadter, A. B., & Neale, M. C. (2020). Model Fit Estimation for Multilevel Structural equation models. *Struct Equ Modeling*, *27*(2), 318–329. https://doi.org/10.1080/10705511.2019.1620109

Rudasill, K. M., Snyder, K. E., Levinson, H., & Adelson, J. L. (2018). Systems view of school climate: A theoretical framework for research. *Educational Psychology Review*, *30*, 35–60. https://doi.org/10.1007/s10648-017-9401-y

Rutkowski, L., Gonzalez, E., Joncas, M., & Von Davier, M. (2010). International large-scale Assessment Data: Issues in secondary analysis and reporting. *Educational Researcher*, *39*(2), 142–151. https://doi.org/10.3102/0013189X10363170

Sanchez, J. E., Paul, J. M., & Thornton, B. W. (2020). Relationships among teachers' perceptions of principal leadership and teachers' perceptions of school climate in the high school setting. *International Journal of Leadership in Education*, 1–21. https://doi.org/10.1080/13603124.2019.1708471

Scheerens, J. (2017). The perspective of limited malleability in educational effectiveness: Treatment effects in schooling. *Educational Research and Evaluation*, *23*(5–6), 247–266. https://doi.org/10.1080/13803611.2017.1455286

Scherer, R., & Nilsen, T. (2016). The relations among school climate, instructional quality, and achievement motivation in mathematics. In T. Nilsen, & J. E. Gustafsson (Eds.), *Teacher quality, Instructional Quality and Student outcomes: Relationships across Countries, cohorts and Time*. Springer International Publishing.

Schwab, S., Zurbriggen, C. L., & Venetz, M. (2020). Agreement among student, parent and teacher ratings of school inclusion: A multitrait-multimethod analysis. *Journal of School Psychology*, *82*, 1–16. https://doi.org/10.1016/j.jsp.2020.07.003

Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct meaning in multilevel settings. *Journal of Educational and Behavioral Statistics*, *41*(5), 481–520. https://doi.org/10.3102/1076998616646200

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, *83*(3), 357–385. https://doi.org/10.3102/0034654313483907

Wagner, S. M., Rau, C., & Lindemann, E. (2010). Multiple informant methodology: A critical review and recommendations. *Sociological Methods & Research*, *38*(4), 582–618. https://doi.org/10.1177/0049124110366231

Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B., & Trautwein, U. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology*, *108*(5), 705.

Wang, M. T., & Degol, J. L. (2016). School Climate: A review of the Construct, Measurement, and impact on Student outcomes. *Educational Psychology Review*, *28*(2), 315–352. https://doi.org/10.1007/s10648-015-9319-1

Wu, J. H., Hoy, W. K., & Tarter, C. J. (2013). Enabling school structure, collective responsibility, and a culture of academic optimism. *Journal of Educational Administration*. https://doi.org/10.1108/09578231311304698

Publisher's Note