Assessing the comparability of teacher-related constructs in TIMSS 2015 across 46 education systems: an alignment optimization approach

Leah Natasha Glassow¹ · Victoria Rolfe¹ · Kajsa Yang Hansen¹

Received: 18 December 2019 / Accepted: 9 December 2020 / Published online: 2 February 2021

© The Author(s) 2021

Abstract

Research related to the “teacher characteristics” dimension of teacher quality has proven inconclusive and weakly related to student success, and addressing the teaching contexts may be crucial for furthering this line of inquiry. International large-scale assessments are well positioned to undertake such questions due to their systematic sampling of students, schools, and education systems. However, researchers are frequently prohibited from answering such questions due to measurement invariance related issues. This study uses the traditional multiple group confirmatory factor analysis (MGCFA) and an alignment optimization method to examine measurement invariance in several constructs from the teacher questionnaires in the Trends in International Mathematics and Science Study (TIMSS) 2015 across 46 education systems. Constructs included mathematics teacher’s *Job satisfaction*, *School emphasis on academic success*, *School condition and resources*, *Safe and orderly school*, and teacher’s *Self-efficacy*. The MGCFA results show that just three constructs achieve invariance at the metric level. However, an alignment optimization method is applied, and results show that all five constructs fall within the threshold of acceptable measurement non-invariance. This study therefore presents an argument that they can be validly compared across education systems, and a subsequent comparison of latent factor means compares differences across the groups. Future research may utilize the estimated factor means from the aligned models in order to further investigate the role of teacher characteristics and contexts in student outcomes.

Keywords TIMSS · Measurement invariance · Alignment optimization · Teacher characteristics

✉ Leah Natasha Glassow
leah.glassow-hill@gu.se

1 Introduction

1.1 Teacher quality: context and comparability

Internationally, teachers have been cited as the most important school-level determinant of academic success (Darling-Hammond 2000; Hattie 2003; Rivkin et al. 2005; Kyriakides et al. 2013; Nilsen and Gustafsson 2016a, 2016b). However, despite decades of research, there is still considerable debate over the importance of particular teacher characteristics. Research on teacher characteristics varies widely, and ranges from beliefs about intelligence and learning, self-efficacy, job satisfaction and motivation, to workload, and stress (Goe 2007). This study will operate from the theoretical framework defining teacher characteristics within the teacher quality construct outlined by Goe (2007). According to this review, changeable characteristics or teacher “attributes and attitudes” form part of the input dimension of teacher quality. While teachers are considered crucial for student outcomes, evidence on the importance of teacher characteristics is weak or conflicting. A myriad of studies conducted using international large-scale assessment data have found mixed results (Goe 2007; Nilsen and Gustafsson 2016a, 2016b; Toropova et al. 2019). With this in mind, Goe (2007) recommends that more research on teacher characteristics be conducted with a particular focus on the teaching context.

International large-scale assessments (ILSAs) such as the Trends in International Mathematics and Science Study (TIMSS) are well positioned to answer such questions through information collected in the contextual questionnaires for students, teachers, and principals. While such studies have advanced global educational accountability and contributed valuable knowledge regarding determinants of student outcomes, they have also sparked questioning over the validity of cross-national comparison (Oliveri and von Davier 2011; Biemer and Lyberg 2003). Underlying the contextual questionnaires is the much-debated assumption of scale score equivalence or measurement invariance (MI). Issues related to MI often prevent researchers from answering important substantive questions, which entail comparing latent factor means and the relationships among latent variables across countries or time. The main reason for the concern over measurement invariance in cross-national comparison is the difficulty involved with measuring psychological traits or constructs across cultures, as cultural factors may influence how respondents interpret and answer such questions. Several scholars have argued that TIMSS is superior to other ILSAs regarding the potential to examine teacher characteristics due to the systematic collection of data directly from teachers. TIMSS is also the only ILSA to link students and teachers directly. Despite this, research on teachers in international large-scale assessments is often limited to comparisons of relationships between variables because of the failure to reach scalar invariance across countries (Nilsen and Gustafsson 2016a, 2016b). Such questions which have not yet been investigated involve comparisons of latent construct means in teacher questionnaires across education systems or their subgroups. Certain teacher characteristics may matter in some contexts and not in others (Strong 2011). For instance, teachers have been shown to be especially important for low-achieving and socioeconomically disadvantaged students, and especially in mathematics (Goe 2007; Darling-Hammond 2000; Rivkin et al. 2005). Equally, the context (i.e., school, country, or educational system) may predict the teacher characteristics themselves due to

differences in system level characteristics or educational policies. Taken together, mean comparisons and subsequent connection to student outcomes may have important insight into teacher-related policies which researchers have been largely unable to investigate.

As will be discussed in the following section, the alignment optimization method outlined by Asparouhov and Muthén (2014) provides one possible resolution to this problem as well as an empirical basis for investigating such contextual questions. This study will utilize this method and examine measurement invariance in five scales of the teacher background questionnaires in TIMSS 2015. These constructs fall under the category of “teacher characteristics, beliefs, and attributes” according to Goe’s (2007) framework, but vary in their scope. *Job satisfaction* (JS) refers to how satisfied teachers are with their employment and their plans for continuing to teach in the future. *School emphasis on academic success* (SEAS) refers to teachers’ perceptions of the academic climate and emphasis on academics of other teachers at their school, and *Safe and orderly school* (SOS) refers to the teacher’s general feelings of safety and organization at their workplace. *School condition and resources* (SCR) refers to the teacher’s perceptions of their access to teaching resources and how well the school is maintained. Last, teacher’s *Self-efficacy* (TSE) refers to the teachers’ perceptions of their confidence and ability to teach mathematics (for more on TSE, see Raudenbush et al. 1992).

The present study applies the alignment method as an exploratory tool to examine measurement invariance in the latent constructs from the teacher questionnaires in TIMSS 2015 across educational systems. Our paper is both content and method focused. Our intention is to provide researchers in comparative education—particularly those interested in teacher effectiveness—with one possible starting point for tackling questions which remain unanswered due to issues surrounding measurement invariance and cross-national comparison. The paper seeks to answer the following research questions:

- (1) What is the level of configural, metric, and scalar invariance of the teacher-related constructs in the teacher background questionnaires of TIMSS 2015 across educational systems?
- (2) Within these constructs, which indicators display the highest level of non-invariance in the teacher-related constructs? Is there a statistical basis for making comparisons of these constructs educational across systems?
- (3) Based on the newly constructed group mean values, which education systems have the lowest and highest levels of the teacher-related constructs?

1.2 Approaches to measurement invariance and a review of past literature

MI (Jöreskog 1971; Mellenbergh 1989; Meredith 1993) refers to the assumption that latent constructs and their relations should be unrelated to group membership, and is one of the main challenges of working with ILSA data (Gustafsson 2018). Within the traditional multiple group confirmatory factor analysis (MGCFA) approach, several levels of MI are tested, beginning with the configural or baseline model. In order to confirm configural invariance, factors must be equally configured under a similar variance-covariance structure across groups. Next, factor loadings (regression slopes)

are compared; if loadings are similar across groups, metric invariance is achieved. This implies that each indicator is related to its underlying latent variable with a similar gradient. Scalar invariance is the most restricted form of MI and requires regression intercepts to be equivalent, in addition to latent structures and factor loadings. In scalar invariance, the same regression line should be able to estimate the relationship between an indicator and the latent variable for all groups. The three forms of MI build successively upon each other, representing a growing degree of invariance. Violating the assumption of MI results in constraints that inherently limits how researchers may interpret and relay their findings in a comparative context. As meeting the scalar MI assumption is very rare, occasionally, “researchers just ignore MI issues and compare latent factor means across groups or measurement occasions even though the psychometric basis for such a practice does not hold” (van de Schoot et al. 2015, p 1). More cautious approaches avoid comparing constructs altogether. Either scenario may be problematic in the context of ILSA research, given its relevance and potential for educational policy and reform.

There are several conceptual and methodological recommendations for managing MI. Rutkowski and Rutkowski (2010, 2013, 2017) propose the possibility that “one size might not fit all” and that scales be constructed with differing cultural conceptions in mind. A more moderate and early solution comes from Byrne et al. (1989) in partial measurement invariance, which allows intercepts and loadings of individual items to be tested. Following this approach, the majority of scholars recommend basing the types of comparisons on the level of invariance confirmed (i.e., configural, metric, or scalar), and this undoubtedly leads to smaller number of constructs being investigated due to their failure in reaching full invariance. Schulz (2016) argues that focusing only on constructs and variables that are highly similar in terms of measurement may lead to a narrowing in the scope of international studies. Generally, partial measurement invariance is a practical assumption in ILSA research, where invariance at the scalar level is rarely confirmed. However, scholars have debated whether the traditional MGCFA approach to partial measurement invariance is the most “simple or interpretable” solution (for more detail, see Marsh et al. 2018 and Asparouhov and Muthén 2014).

A more recent approach, an alignment optimization method, has been proposed (Asparouhov and Muthén 2014). Alignment optimization allows for invariance of individual items to be tested, for scales to be reformulated in order to take non-invariance into consideration, and to create a more flexible threshold for measurement invariance. Schulz (2016) writes, “the question is also at what point lack of measurement invariance becomes problematic and leads to problematic bias in cross-national surveys” (p. 15). The alignment method (Asparouhov and Muthén 2014) undertakes this question. This method has certain advantages over other approaches to MI. Traditionally, MI is tested using MGCFA at each constraint of the latent factor model, with groups defined by unordered categorical variables (van de Schoot et al. 2015). This approach requires that invariance levels be tested sequentially and for each item, which can result in hundreds of tests. Moreover, such tests can result in inaccurate results if multiple groups are present or if sample sizes are large (Asparouhov and Muthén 2014; Rutkowski and Svetina 2014). The traditional approach to MI also assumes that full measurement invariance can be achieved, which may be an “unachievable ideal” when the number of groups is large (Marsh et al. 2018; Asparouhov and Muthén 2014). Unlike MGCFA, alignment as outlined by

Asparouhov and Muthén (2014) does not assume MI, but identifies a result which minimizes parameter invariance across groups through an iterative process analogous to the rotation in an exploratory factor analysis. Several studies have investigated measurement invariance using the alignment method with promising results as an alternative to MGCFA. Munck et al. (2018) investigated MI across 92 groups by country, cycle, and gender using civic education data and found that despite significant non-invariance in some groups, comparison of group mean scores had a statistical basis, and that attitudes toward civic engagement across countries and time could be validly compared. Similarly, both Marsh et al. (2018) and Lomazzi (2018) employ the alignment method to test MI of gender role attitudes across countries.

Much attention has been paid to the phenomenon of MI in the student background questionnaires, but much less in teacher-related constructs (Caro et al. 2014; Schulz 2016; Segeritz and Pant 2013; He et al. 2018; Rutkowski and Svetina 2014). Nevertheless, some studies have investigated measurement invariance in teacher background questionnaires using traditional approaches. Examining teacher self-efficacy, Vieluf et al. (2013) find evidence supporting metric equivalence, while Scherer et al. (2016) also find evidence for metric but not scalar invariance. Taking a different approach, Zieger et al. (2019) use multiple pairwise mean comparison to teacher job satisfaction in TALIS, whereby they identify the comparability of countries based on such pairs. Similarly to MGCFA, this approach grows in cumbrousness alongside the number of groups in focus. Despite a growing awareness of the potential of the alignment method, application of this approach in investigating the measurement invariance of latent constructs related to teachers and teacher quality is still rare. Our search was able to produce a single study published just this year. Zakariya et al. (2020) examined teacher job satisfaction in TALIS, and also found no evidence for scalar invariance. Extending their analysis to include an alignment optimization approach, they found that teachers in Austria, Spain, Canada, and Chile had the highest mean job satisfaction compared to the other countries in the sample. Our analysis does not use the same sampling procedure as TALIS, as TIMSS focuses on teachers as they represent *students* in a country. Additionally, our results apply only to mathematics teachers, unlike the results of studies looking at *all* teachers using TALIS data. As such, it will be especially interesting to compare our results to those of Zakariya et al. (2020) and other past studies.

2 Methods

2.1 Data and measurement

TIMSS is a curriculum-based survey, which tests mathematics and science achievement for students in grades 4 and 8 around the world. TIMSS employs a two-stage stratified sampling procedure and samples whole classrooms as well as schools. Additionally, responding to the teacher context questionnaires is mandatory. Student data can therefore be aggregated to the teacher level (Eriksson et al. 2019). TIMSS uses a cross-sectional design and is conducted every 4 years. This study consisted of 46 education systems included in the TIMSS 2015 survey. There was a total sample size of 13,508 grade 8 (or equivalent) mathematics teachers. In the total sample, 36.8% of

teachers were male and 56.6% female; while 2.7% were under 25, 12.5% were between 25 and 29, 29.9% were between 30 and 39, 24.7% were between 40 and 49, 18.9% were between 50 and 59, and 4.7% were above the age of 60 (6.6% had no response). In total, 42 separate countries participated, but in some cases, sub-regions of countries were included, such as Buenos Aires in Argentina, Ontario and Quebec in Canada, and Dubai and Abu Dhabi in the United Arab Emirates (UAE), the term “education system” will be used interchangeably with country, system or group. Norway included cohorts from two grades. However, aside from the regions previously listed, the majority of the groups are representative of countries. Table 1 describes each education system and its respective sample size.

Several teacher-related constructs from the teacher questionnaire were included in the analysis: Teacher *Job satisfaction* and *Self-efficacy*, teacher perception of *School emphasis on academic success*, *School condition and resources*, and *Safe and orderly school*. Indicators and coding for each construct can be seen in Table 2.

Each of the constructs included a varying number of indicators. For *School emphasis on academic success*, only 5 out of a total of 17 indicators were used; as the remaining indicators did not relate to teachers, they were excluded. All indicators were included for each of the other 5 constructs. Coding varied from frequency-dimensions (i.e., “Very often” to “Never or almost never”) to agreement (i.e., “Agree a lot” to “Disagree a lot”) and more general ratings (i.e., “Very high” to “Low”).

2.2 Alignment optimization

As we have previously discussed, there are three levels of measurement invariance: configural, metric, and scalar. In order to compare latent variable means and variances across subgroups, scalar invariance is required (Millsap 2011). However, this assumption (i.e., equal factor loadings and indicator intercepts across subgroups) often fails. Moreover, the likelihood ratio chi-square testing for each parameter very quickly becomes cumbersome, especially when many subgroups are being compared. The alignment approach does not assume MI and “can estimate the factor mean and variance parameters in each group while discovering the most optimal measurement invariance pattern. The method incorporates a simplicity function similar to the rotation criteria used with exploratory factor analysis” (Asparouhov and Muthén 2014, p. 496). It estimates a factor score for all individuals despite the presence of significant non-invariance in some groups. Alignment starts with estimating such a configural model with group-varying factor loadings and intercepts to latent variable indicators and the factor mean and variance. Consider a configural MGCFA model, written as:

$$Y_{pj} = v_{pj} + \lambda_{pj}\eta_j + \varepsilon_{pj}, \quad (1)$$

Here, v_{pj} is the intercept of an indicator p in a group j , λ_{pj} is the factor loading of the indicator i in the group j , η_j is the latent variable for group j , and ε_{pj} is the residual for indicator p in group j . In this model, the latent variable mean is fixed to zero and the latent variable variance to 1:

$$E(\eta_j) = \alpha_j = 0; V(\eta_j) = \psi_j = 1 \quad (2)$$

Table 1 Education systems and number of teachers sampled

Country	Abbreviation	<i>N</i>
Australia	AUS	941
Bahrain	BHR	201
Armenia	ARM	210
Botswana	BWA	169
Canada	CAN	409
Chile	CHL	173
Chinese Taipei	TWN	216
Georgia	GEO	188
Hong Kong	HGK	175
Hungary	HUN	278
Iran	IRN	251
Ireland	IRE	526
Israel	ISR	603
Italy	ITA	230
Japan	JPN	231
Kazakhstan	KAZ	239
Jordan	JOR	260
South Korea	KOR	317
Kuwait	KWT	191
Lebanon	LBN	185
Lithuania	LTU	269
Malaysia	MYS	326
Malta	MLT	224
Morocco	MOR	373
Oman	OMN	356
New Zealand	NWZ	489
Norway	NOR	239
Qatar	QAT	250
Russia	RUS	226
Saudi Arabia	SAU	149
Singapore	SNG	334
Slovenia	SLV	467
South Africa	ZAF	334
Sweden	SWE	206
Thailand	THL	205
UAE	UAE	746
Turkey	TUR	220
Egypt	EGY	215
United States	USA	429
England	ENG	215
Norway 8	NOR8	239
UAE Dubai	UAED	267

Table 1 (continued)

Country	Abbreviation	<i>N</i>
UAE Abu Dhabi	UAEAD	207
Canada Ontario	CANON	217
Canada Quebec	CANQU	175
Argentina (BA)	ARGBA	138

As a second step, the fixed factor mean and variance are set free. Normally, this model would be unidentified. The alignment method, however, constrains the parameter estimation through imposing restrictions to optimize the simplicity function F . As is shown in Eq. 3, the sum of the component loss function for the factor loadings and intercepts of every latent variable indicator p between any pair of groups weighted by their group sizes¹ should be minimal.

$$F = \sum_p \sum_{j_1 < j_2} w_{j_1, j_2} f(\lambda_{pj1} - \lambda_{pj2}) + \sum_p \sum_{j_1 < j_2} w_{j_1, j_2} f(v_{pj1} - v_{pj2}) \quad (3)$$

The alignment approach estimates the latent variable mean and variance for each pair of groups in such a way that the parameter estimates are optimized to produce the minimal total amount of non-invariance across groups. This procedure leads to a great number of parameters that have no significant non-invariance across groups and a few being largely non-invariant. Significant differences are tested by z-statistics (for a more detailed description of the algorithm, see Asparouhov and Muthén 2014; Muthén and Asparouhov 2018).

The aligned model produces an alignment optimization metric (A-metric) with some useful statistical information for determining measurement invariance of the latent variable across groups. The first important piece of information is the amount of groups that has no significant differences in each intercept and factor loading. The order of the factor mean and the groups who hold the minimum and maximum intercept and factor loading for each factor indicator are also given in the alignment results. In addition, an R -square, measuring the degree of invariance of the intercept and factor loading of each factor indicator is estimated in the model.

$$R_{intercept}^2 = 1 - \frac{V(v_0 - v - \alpha_j \lambda)}{V(v_0)} \quad (4)$$

$$R_{factor\ loading}^2 = 1 - \frac{V(\lambda_0 - \sqrt{\psi_j} \lambda)}{V(\lambda_0)} \quad (5)$$

¹ $w_{j_1, j_2} = \sqrt{N_{j1} N_{j2}}$, N_{j1} and N_{j2} is the sample size of group j_1 and j_2 .

Table 2 Constructs, indicators, and coding of teacher-related constructs

Construct	Indicators	Coding
Job satisfaction (JS)	I am content with my profession as a teacher	4 = “Very often”
	I am satisfied with being a teacher at this school	3 = “Often”
	I find my work full of meaning and purpose	2 = “Sometimes”
	I am enthusiastic about my job	1 = “Never or almost never”
	My work inspires me	
	I am proud of the work I do	
	I am going to continue teaching for as long as I can	
School emphasis on academic success (SEAS)	Teachers’ understanding of the school’s curricular goals	5 = “Very low”
	Teachers’ degree of success in implementing the school’s curriculum	4 = “Low”
	Teachers’ expectations for student achievement	3 = “Medium”
	Teachers working together to improve student achievement	2 = “High”
	Teachers’ ability to inspire students	1 = “Very high”
School condition and resources (SCR)	The school building needs significant repair	4 = “Not a problem”
	Teachers do not have adequate workspace (for preparation, collaboration, or meeting with students)	3 = “Minor problem”
	Teachers do not have adequate instructional materials and supplies	2 = “Moderate problem”
	The school classrooms are not cleaned often enough	1 = “Serious problem”
	The school classrooms need maintenance work	
	Teachers do not have adequate technological resources	
Safe and orderly school (SOS)	Teachers do not have adequate support for using technology	
	This school is located in a safe neighborhood	4 = “Agree a lot”
	I feel safe at this school	3 = “Agree a little”
	This school’s security policies and practices are sufficient	2 = “Disagree a little”
	The students behave in an orderly manner	1 = “Disagree a lot”
	The students are respectful of the teachers	
	The students respect school property	
	This school has clear rules about student conduct	
Self-efficacy (TSE)	This school’s rules are enforced in a fair and consistent manner	
	Inspiring students to learn mathematics	4 = “Very high”
	Showing students a variety of problem-solving strategies	3 = “High”
	Providing challenging tasks for the highest achieving students	2 = “Medium”
	Adapting my teaching to engage students’ interest	1 = “Low”
	Helping students appreciate the value of learning mathematics	
	Assessing student comprehension of mathematics	
	Improving the understanding of struggling students	
	Making mathematics relevant to students	
	Developing students’ higher-order thinking skills	

As is shown in Eqs. 4 and 5, ν_0 and λ_0 are the intercept and factor loading estimates from the configural model and ν and λ are the average intercept and factor loading estimated from the aligned model. The R^2 “tells us how much of the configural parameter variation across groups can be explained by variation in the factor means and factor variances (Muthén and Asparouhov 2018, p. 643). An R^2 value close to one indicates a high degree of measurement invariance and close to zero indicates high non-invariance.

Mplus detects missing patterns in the data sets and provides full information maximum likelihood (FIML) estimates of the missing data through the EM algorithm. It also should be noted that all models in the study were estimated with the COMPLEX option implemented in Mplus to account for the non-independency of the students and teachers caused by the cluster sampling design in TIMSS (Muthén and Muthén 1998-2017).

2.3 Analytical process

The current analysis was done stepwise. All analyses were conducted using Mplus software 8.3 (Muthén and Muthén 1998-2017). In the first step, a single-factor measurement model was estimated for each of the teacher-related constructs with pooled data. These single-factor measurement models were modified by adding correlated residual terms suggested by the modification indices to get acceptable model fit. The significantly correlated residuals indicate that there are common variances between the pairs of residuals, suggesting some narrow dimensions in addition to the single latent factor. In the current study, we are only interested in precisely measuring the general factor, with no narrow residual factors being specified. With the pooled model structure as the point of departure, the conventional MGCFA models of different teacher-related factors were conducted, and model fit indices of the configural, metric, and scalar invariance models were compared for each of the constructs. Based on these comparisons, conclusions of the MI were reached. In the next step, the alignment approach was tested to the degree of measurement invariance of the teacher-related constructs, as mentioned above. The results of the two MI approaches are compared, and the advantages and disadvantages of the two are discussed. In order to check the reliability, a Monte Carlo simulation was done to further test whether the conclusion about measurement invariance based on the aligned model results of the constructs is trustworthy.

3 Results

3.1 Results from the MGCFA approach

A single-factor measurement model was fitted to each of the teacher-related constructs with the pooled data of all 46 education systems. These single-factor measurement models, however, did not fit the data well. Modification indices suggested the inclusion of one or more correlated residuals to improve the model fit. These modified single-factor model structures were used to test the measurement invariance across the 46 groups in the conventional approach. Table 3 presents the model fit indices of the configural, metric, and scalar MI models for all teacher-related constructs.

Table 3 Conventional measurement invariance model fit and model comparisons for all teacher-related constructs

	No of parameters	χ^2	df	χ^2_{diff}	Δdf	RMSEA (90% CI)	SRMR	CFI	TLI
Configural	1104	1143.283*	506			.068 [†] (.063–.073)	.030 [†]	.978 [†]	.959 [†]
Metric	834	2041.882*	776	898.599	270	.077 (.073–.081)	.140	.957 [†]	.946
Scalar	564	4411.334*	1046	2369.452	270	.109 (.105–.112)	.184	.886	.894
School emphasis on academic success									
Job satisfaction									
Configural	782	214.61*	138			.045 [†] (.033–.056)	.018 [†]	.995 [†]	.983 [†]
Metric	602	482.951*	318	268.341	180	.043 [†] (.035–.051)	.055 [†]	.989 [†]	.984 [†]
Scalar	422	3242.335*	498	2759.384	180	.142 (.137–.146)	.148	.815	.829
School condition and resources									
Configural	1104	1353.168*	506			.078 [†] (.073–.083)	.040 [†]	.968 [†]	.939
Metric	834	1997.683*	776	644.515	270	.076 [†] (.072–.080)	.074 [†]	.954 [†]	.943
Scalar	564	479.880*	1046	2793.197	270	.114 (.111–.118)	.119	.859	.870
Safe and orderly school									
Configural	1288	1732.337*	736			.070 [†] (.066–.075)	.047 [†]	.972 [†]	.950 [†]
Metric	973	2819.242*	1051	1086.905	315	.078 [†] (.075–.082)	.203	.950 [†]	.938
Scalar	658	6239.447*	1366	342.205	315	.114 (.111–.117)	.302	.861	.869
Self-efficacy									
Configural	1426	2144.005*	1058			.062 [†] (.058–.065)	.034 [†]	.972 [†]	.956 [†]
Metric	1066	2918.397*	1418	774.392	360	.063 [†] (.059–.066)	.069 [†]	.961 [†]	.955 [†]
Scalar	706	7042.664*	1778	4124.267	360	.105 (.102–.107)	.098	.863	.873

* $p < .000$ [†] Values demonstrate acceptable fit (Hu and Bentler 1999)

The configural models of all the latent constructs in Table 3 show acceptable or close model fit, with the Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) being below .08, and comparative fit index (CFI) and Tucker-Lewis index (TLI) being greater than .95 (see, e.g., Hu and Bentler 1999). Three out of the seven teacher-related factors (teacher perception of *School emphasis on academic success*, *School Condition and Resources*, and teacher's *Self-efficacy*) reached metric invariance, which implied that the factor loadings of each of the three latent constructs were equal across all educational systems, but not the intercepts of the latent construct indicators. It may also be observed that none of the scalar MI models fits the data, indicating that the assumption that both intercepts and factor loadings be equal across the 46 systems cannot be held true.

With the traditional measurement invariance approach, the restricted MI assumption (scalar invariance) has been proven false. Additionally, metric invariance was only found in three latent constructs. Consequently, cross-country comparisons cannot be made with the latent variable means as well as the relationships among the latent variables. Given these results, the next section will aim for an approximate partial measurement invariance (e.g., Millsap and Kwok 2004) by using the alignment approach (Muthén and Asparouhov 2014).

3.2 Results from alignment optimization

Alignment optimization explores partial (approximate) measurement invariance by starting out with a well-fitting configural model. It then adjusts the factor loadings and intercepts of the factor indicators in such a way that these parameter estimates should be as similar as possible across groups without compromising the model fit. Essentially, the fit for the aligned model stays the same as the configural invariance model. In this section, the aligned model results for each of the seven teacher-related factors will be presented.

3.2.1 Job satisfaction

Table 4 presents the results from the aligned modeling approach for the latent construct JS. The highest *R*-square of the intercept estimate is observed for the variable *My work inspires me*. About 87% of the variation in the intercept observed in the configural model can be explained by the variation in latent variable mean and variance in the aligned model, indicating a high degree of invariance. Morocco is the only non-invariant country in the intercept estimate of the indicator *I am proud of the work I do*. This variable together with the indicator *I am enthusiastic about my job* also displayed a rather high *R*-square. *I am content with my profession as a teacher* and *My work inspires me* hold completely invariant factor loading estimates across all systems. For the variables *I am enthusiastic about my job*, and *I find my work full of meaning and purpose*, a large number of groups with invariance in the intercept estimates are also observed, ranging from 44 to 46 educational systems. The variable *I am going to continue teaching as long as I can* holds the least invariant intercept with the *R*-square being the lowest, 44%. For the factor loadings, the indicator *I am proud of what I do* is the least invariant, with an *R*-square of 23%.

Table 4 Results from the aligned model of job satisfaction (JS)

How often do you feel the following way about being a teacher? R^2	Item parameter estimates in the alignment optimization metric	Countries with non-invariance			
		Minimum		Maximum	
		Est.	Country	Est.	Country
Intercept					
I am content with my profession as a teacher	.770	-.747	.155	-1.071	TWN GEO
I am satisfied with being a teacher at this school	.542	-.748	.316	-1.203	LTU BWA
I find my work full of meaning and purpose	.670	-.751	.134	-1.045	SWE EGY
I am enthusiastic about my job	.808	-.751	.087	-.909	ENG THA
My work inspires me	.870	-.745	.086	-.930	KOR EGY
I am proud of the work I do	.749	-.760	.122	-.944	ITA HGK
I am going to continue teaching for as long as I can	.442	-.742	.262	-1.063	KOR MAR
Loading					
I am content with my profession as a teacher	.619	.993	.094	.680	ARGBA 1.178 JOR
I am satisfied with being a teacher at this school	.357	.994	.151	.523	IRN 1.243 LBN
I find my work full of meaning and purpose	.540	.990	.133	.589	GEO 1.207 ARGBA
I am enthusiastic about my job	.458	.993	.127	.525	ARGBA 1.280 TUR
My work inspires me	.683	.994	.075	.798	EGY 1.187 QAT
I am proud of the work I do	.226	.996	.154	.618	CANON 1.428 GEO
I am going to continue teaching for as long as I can	.315	.995	.157	.377	ARGBA 1.378 KWT
Average invariance index	.575				
Total non-invariance	8.85%				

Countries with extreme parameter estimates can be found in columns 4 to 7. For example, South Korea holds the lowest intercept estimates in *My work inspires me*, while Canada-Ontario has the lowest factor loading estimate. In general, the overall degree of invariance of the construct JS is rather high, with few education systems showing measurement non-invariance in the factor loadings, complying with the close fit for the metric invariance model in Table 3. The average invariance index is 58% for JS. The percentage of significant non-invariance groups is 8.9%, much lower than the limit of 25% suggested by Muthén and Asparouhov 2014. A higher number of groups show invariance in the factor loadings of each of the indicators as compared to the intercepts.

3.2.2 Teacher perception of school emphasis on academic success

Five indicators are used to identify the latent construct of school emphasis on academic success, and the results from the aligned model of SEAS are presented in Table 5.

For factor loading estimates, all five indicators to the construct *School emphasis on academic success* showed complete invariance over the 46 countries. This agrees with the model fit indices for the metric invariance model in Table 3. For the intercepts, only two countries are non-invariant for the indicator *Teachers' degree of success in implementing the school's curriculum*, corresponding with the high *R*-square estimate 73%. The intercept of *Teachers' expectations for student achievement* holds the most variation, with only half of the countries being invariant. The minimum and maximum estimates of the intercept and factor loadings can be found in columns 4 to 7. Only 7.8% of groups have been observed with significant non-invariance. In general, the high degree of confidence indicated by the average invariance index of .65 implies that the mean of the construct SEAS can be compared meaningfully across the different groups.

3.2.3 Teacher perception of school conditions and resources

Table 6 shows the results of approximate invariance from the aligned model of the school condition and resources.

As revealed in Tables 6 and 4 indicators, *The school building needs significant repair*, *Teachers do not have adequate instructional materials and supplies*, *The school classroom needs maintenance work*, and *Teachers do not have adequate support for using technology* have invariant factor loadings across all education systems. Only Lithuania is non-invariant in the factor loadings for the variables *Teachers do not have adequate workplace* and *Teachers do not have adequate technological resources*. The *R*-square for these indicators also showed a high degree of invariance, being above 60%. However, one exception can be observed for the variable *The school building needs significant repair*, for which the *R*-square is 29%, despite showing complete invariance across all groups. For the intercept estimates, the number of non-invariant systems in each indicator ranges from 4 for the variable *The school classroom needs maintenance work* (*R*-square = 82%) to 10 for the variable *Teachers do not have*

Table 5 Results from the aligned model of school emphasis on academic success (SEAS)

How would you characterize each of the following within your school?	R^2	Item parameter estimates in the alignment optimization metric				Countries with non-invariance			
		Mean	Std. Deviation	Minimum	Maximum				
				Est.	Country	Est.	Country		
Intercept									
Teachers' understanding of the school's curricular goals	.720	-.646	.278	-1.432	JPN	-.070	RUS	JPN. MOR. RUS	
Teachers' degree of success in implementing the school's curriculum	.734	-.619	.275	-.985	MYS	.410	TWN	ARM. TWN	
Teachers' expectations for student achievement	.431	-.683	.442	-1.634	MAR	.001	NZL	AUS. CAN. CHL. HUN. MOR. OMN. NZL. NOR. RUS. SGP. ZAF. SWE. THA. USA. ENG. UAEAD. CANON. CANQU	
Teachers working together to improve student achievement	.708	-.694	.285	-1.473	MAR	-.164	KWT	CHL. TWN. ISR. JPN. MOR. SGP	
Teachers' ability to inspire students	.718	-.690	.274	-1.472	QAT	-.028	GEO	ARM. GEO. HUN. ITA. JPN. KOR	
Loading									
Teachers' understanding of the school's curricular goals	.630	.984	.131	.717	JOR	1.257	MAR		
Teachers' degree of success in implementing the school's curriculum	.756	.982	.105	.736	TWN	1.313	JPN		
Teachers' expectations for student achievement	.605	.986	.129	.623	JOR	1.163	AUS		
Teachers working together to improve student achievement	.636	.989	.116	.766	ITA	1.336	KAZ		
Teachers' ability to inspire students	.536	.987	.142	.701	ITA	1.314	ARGBA		
Average invariance index	.647								
Total non-invariance	7.83%								

Table 6 Results from the aligned model of school condition and resources (SCR)

In your current school, how severe is each problem?	R ²	Item parameter estimates in the alignment optimization metric				Countries with non-invariance	
		Mean	Std. Deviation	Minimum			Maximum
				Est.	Country		
Intercept							
The school building needs significant repair	.531	-.764	.327	-1.704	SAU	.435 RUS AUS. KAZ. MAR. NOR. RUS. SAU	
Teachers do not have adequate workspace (e.g., for preparation, collaboration, or meeting with students)	.567	-.770	.303	-1.288	ZAF	.004 ISR CHL. IRE. ISR. ITA. KOR. QAT. SVN. UAE. USA. UAED	
Teachers do not have adequate instructional materials and supplies	.791	-.768	.166	-1.126	MYS	-.476 UAED SVN. UAE. UAED. UAEAD. CANQU	
The school classrooms are not cleaned often enough	.635	-.769	.262	-1.163	ARM	.227 MYS AUS. KAZ. LTU. MYS. RUS. SGP. SVN. ENG	
The school classrooms need maintenance work	.821	-.781	.224	-1.108	CANQU	.137 RUS IRE. RUS. SVN. USA	
Teachers do not have adequate technological resources	.888	-.768	.243	-1.245	SWE	.123 MAR CHL. IRE. MAR. EGY	
Teachers do not have adequate support for using technology	.793	-.757	.339	-1.358	GEO	.461 BWA BWA. GEO. JPN. KAZ. KWT. LTU. MAR. SVN. UAE. EGY	
Loading							
The school building needs significant repair	.285	.996	.149	.712	ITA	1.541 JPN	
Teachers do not have adequate workspace (e.g., for preparation, collaboration, or meeting with students)	.501	.989	.162	.531	LTU	1.51 JPN LTU	
Teachers do not have adequate instructional materials and supplies	.661	.990	.092	.301	ARM	1.133 HUN	
The school classrooms are not cleaned often enough	.247	.987	.291	.717	ARGBA	1.737 CANQU ARM. HUN. KWT. LTU CANQU	
The school classrooms need maintenance work	.657	.987	.118	.672	ITA	1.234 ARM/NOR	
Teachers do not have adequate technological resources	.696	.987	.109	.784	JPN	1.281 LTU LTU	
Teachers do not have adequate support for using technology	.622	.984	.133	.583	BWA	1.284 ARGAB	
Average invariance index	.621						
Total non-invariance	8.39%						

adequate workplace (R -square = 57%). These results were also confirmed by the conventional measurement invariance results, where metric invariance was achieved for the SCR construct but not scalar invariance (see Table 3).

The average invariance index for the construct SCR was 62%, indicating 62% confidence to carry out trustworthy cross-system comparisons. The total non-invariance measure is 8.39%, below the limit of 25%.

3.2.4 Teacher perception of safe and orderly school

Among the 8 indicators of the latent construct *Safe and orderly school* (Table 7), *The students behave in an orderly manner*, *The students respect school property*, and *The students are respectful of the teachers* are completely invariant in the factor loadings over the 46 countries. The R -square estimate for the factor loading of these three variables is around or above 70%, implying that approximately 70% or above of the variation in the factor loadings estimated in the configural model can be explained by the factor mean and variance across the groups. For these three variables, the standard deviation of the parameter mean is also smaller, compared to those of other indicators. The lowest R -square for the factor loading is observed in the indicator *The school is located in a safe neighborhood* (29%), relating to a larger variation (see column 3 under SD).

Students respect school property holds the highest R -square (i.e., 83%) for its intercept estimate, only Lebanon is non-variant. The lowest R -square is found in the indicator *The school's rules are enforced in a fair and consistent manner* (35%). The number of countries with non-invariance intercept ranges from 1 and 13. From the model fit indices of the conventional measurement invariance model, metric invariance is supported and was confirmed by the aligned model.

In sum, the parameter estimates of the latent variable model reached 58% confidence to make reliable across-country comparison and the percent of significant non-invariance for education systems is only 9.8% over all estimated parameters.

3.2.5 Teacher's self-efficacy

Aligned model results for self-efficacy can be seen in Table 8. The intercept estimates show the indicator *Developing students' higher-order thinking skills* as the most invariant, with an R -square of about 90%. Here, only four educational systems show measurement non-invariance and the variance in the estimated mean intercept is rather small. The intercept estimate for indicator *Making mathematics relevant to students* also holds a high R -square (86%). *Improving the understanding of struggling students* and *Assessing student comprehension of mathematics* show the lowest R -square values, implying a high degree of non-invariance. This is also confirmed by the higher standard deviations in column 3. Over ten educational systems show non-invariance for these two indicators. Columns 4 to 7 present the education system with the minimum or maximum estimate of the intercepts.

The number of educational systems with invariant factor loadings for the TSE constructs is higher than that of the intercepts. *Developing students' higher-order thinking skills*, *Improving the understanding of struggling students*, *Providing challenging tasks for the highest achieving students*, and *Adapting my teaching to engage*

Table 7 Results from the aligned model of safe and orderly school (SOS)

Thinking about your current school. Indicate the extent to which you agree or disagree with each of the following statements.	R ²	Item parameter estimates in the alignment optimization metric				Countries with non-invariance	
		Deviation					
		Mean	Std.	Minimum	Maximum	Est.	Country
Intercept							
This school is located in a safe neighborhood	.486	– 1.151	.554	– 1.810	BHR	1.105	ARGBA BHR, CHL, NOR, UAE, UAED, ARGBA
I feel safe at this school	.607	– 1.144	.292	– 1.797	SWE	.178	BWA ARM
This school's security policies and practices are sufficient	.663	– 1.092	.416	– 1.543	SWE	.982	BWA BWA, IRE, LTU
The students behave in an orderly manner	.752	– 1.037	.231	– 1.319	ZAF	– .359	KWT ARM, ISR, KWT, LBN, LTU, OMN, NZL, NOR, THA, UAE, EGY, UAEAD
The students are respectful of the teachers	.726	– 1.031	.182	– 1.580	BWA	– .511	JPN ARM, JPN, KWT, LTU
The students respect school property	.832	– 1.030	.155	– 1.420	LBN	– .708	NOR LBN
This school has clear rules about student conduct	.499	– 1.031	.186	– 1.469	RUS	– .537	JPN RUS
This school's rules are enforced in a fair and consistent manner	.369	– 1.039	.284	– 1.568	BWA	– .479	CANQU AUS, CAN, CHL, HUN, IRE, NZL, NOR, RUS, USA, CANQU
Loading							
This school is located in a safe neighborhood	.294	.967	.412	.051	ARGBA	1.830	TUR KWT, MAR, OMN, QAT, UAE, TUR, UAED
I feel safe at this school	.365	.966	.356	.233	NOR	1.749	JPN TWN, ISR, ITA, JPN, KWT, MAR, OMN, NOR, ZAF, UAE, TUR, UAEAD
This school's security policies and practices are sufficient	.476	.973	.248	.455	ARM	1.513	TUR KOR, LTU, MAR, SAU, SVN, SWE, TUR
The students behave in an orderly manner	.734	.988	.102	.774	KWT	1.254	ARGBA
The students are respectful of the teachers	.689	.989	.114	.776	EGY	1.531	ARGBA
The students respect school property	.73	.987	.099	.747	KAZ	1.314	ARM
This school has clear rules about student conduct	.45	.973	.226	.506	RUS	1.366	KWT KWT, OMN, RUS, ZAF, EGY
This school's rules are enforced in a fair and consistent manner	.53	.982	.167	.526	ARM	1.243	RUS IRN
Average invariance index	.575						
Total non-invariance	9.78%						

Table 8 Results from the aligned model of teacher's self-efficacy (TSE)

In teaching mathematics to this class, how would you characterize your confidence in doing the following?	R^2	Item parameter estimates in the alignment optimization metric				Countries with non-invariance	
		Mean	Std. Deviation	Minimum			Maximum
				Est.	Country		
Intercept							
Inspiring students to learn mathematics	.668	-.919	.274	-1.570	SAU	-.378	LTN BHR, BWA, JOR, KWT, LBN, LTU, MAR, SAU, ZAF, EGY, NOR
Showing students a variety of problem-solving strategies	.817	-.909	.189	-1.453	SWE	-.583	IRN SWE, UAE, USA
Providing challenging tasks for the highest achieving students	.591	-.897	.335	-1.490	NOR8	.008	TUR LBN, LTU, MYS, OMN, SVN, TUR, ENG, NOR, UAED
Adapting my teaching to engage students' interest	.809	-.909	.209	-1.869	JPN	-.453	RUS JPN, RUS
Helping students appreciate the value of learning mathematics	.811	-.928	.173	-1.520	JPN	-.635	SVN GEO, IRN, SVN
Assessing student comprehension of mathematics	.520	-.882	.261	-1.469	CANQU	-.248	SAU AUS, CAN, GEO, IRE, ISR, JOR, SAU, SWE, USA, CANQU, ARGBA
Improving the understanding of struggling students	.161	-.880	.331	-1.474	IRN	-.069	EGY AUS, CHL, IRN, JOR, KWT, MYS, MAR, OMN, TUR, EGY
Making mathematics relevant to students	.863	-.901	.171	-1.380	LTU	-.198	TWN BWA, TWN, LTU, MYS, TUR
Developing students' higher-order thinking skills	.897	-.897	.163	-1.392	ARM	-.495	RUS ARM, OMN, RUS, SVN
Loading							
Inspiring students to learn mathematics	.387	.992	.156	.603	EGY	1.319	SWE AUS, MAR, ZAF, SWE, EGY
Showing students a variety of problem solving strategies	.641	.993	.090	.787	BHR	1.231	IRE IRE
Providing challenging tasks for the highest achieving students	.470	.994	.116	.552	LBN	1.184	ZAF
Adapting my teaching to engage students' interest	.489	.996	.091	.785	KWT	1.22	SWE
Helping students appreciate the value of learning mathematics	.302	.997	.122	.701	KWT	1.302	HKG IRN, KWT, MAR

Table 8 (continued)

In teaching mathematics to this class, how would you characterize your confidence in doing the following?	R^2	Item parameter estimates in the alignment optimization metric		Countries with non-invariance			
		Mean	Std. Deviation	Minimum		Maximum	
				Est.	Country	Est.	Country
Assessing student comprehension of mathematics	.292	.998	.121	.772	IRE	1.257	ARM IRE
Improving the understanding of struggling students	.315	.998	.114	.757	TUR	1.25	LBN
Making mathematics relevant to students	.607	.995	.083	.728	LTU	1.189	ARGBA LTU
Developing students' higher-order thinking skills	.645	.996	.072	.790	SAU	1.196	BWA
Average invariance index	.571						
Total non-invariance	8.58%						

students' interest are completely invariant over all 46 education systems. The factor loading estimate for *Inspiring students to learn mathematics* has the highest number of non-invariant systems (5).

In general, the average invariance index was rather high for all estimated parameters in the aligned model and a low proportion of significantly non-invariant groups. We, therefore, have 57% confidence to make meaningful comparisons of the means and variances of teacher self-efficacy.

3.3 Monte Carlo simulation

As recommended by Asparouhov and Muthén (2014), Monte Carlo simulations were conducted in order to check the quality of the alignment results of the five teacher-related factors. These simulations used parameter estimates from the alignment models as data-generated population values. For each of the teacher-related factors, two sets of simulations were run with 100 replications, 46 groups, and two different group sample sizes (500 vs. 1000). Table 9 shows the correction between the generated population values and estimated parameters.

The correlations in Table 9 are the average of the correlation between the population factor mean (or factor variance) and model estimated factor mean (or factor variance) of the 100 replications. These correlations generally are very high, most of which are .98 or above, with the average correlation higher than the factor variance. However, relatively low correlations also are observed for the simulations based on 500 group sample size, for example, .95 for the average correlation of the factor variance in *Job satisfaction* and .96 in teacher perception of *School emphasis on academic success*.

Table 9 Correlations between the generated population and aligned estimated values

Monte Carlo Design	<i>n</i> = 500	<i>nrep</i> = 100	<i>n</i> = 1000	<i>nrep</i> = 100
Estimated statistics	Average	SD	Average	SD
Job satisfaction				
Factor mean	.99	.003	.99	.001
Factor variance	.95	.013	.97	.006
School emphasis on academic success				
Factor mean	.98	.007	.99	.005
Factor variance	.96	.012	.98	.006
School condition and resources				
Factor mean	.99	.002	1.00	.001
Factor variance	.97	.008	.98	.004
Safe and orderly school				
Factor mean	.99	.003	.99	.088
Factor variance	.98	.007	.98	.094
Teacher's self-efficacy				
Factor mean	1.00	.002	1.00	.001
Factor variance	.98	.007	.97	.007

n, group sample size; *nrep*, number of replications; *SD*, standard deviation

These correlations tend to get higher when the group sample size is increased to 1000. Asparouhov and Muthén (2014) suggested a level of .98 for these correlations to be able to confirm reliable alignment estimates, and a correlation below .95 may be cause for concern. The current simulations therefore suggest that to a great extent the aligned results for the teacher-related constructs are highly reliable for cross-country comparison, despite some non-invariance among education systems. It can be noted that the aligned models work better when the group sample size is higher, implying an asymptotic accuracy in the alignment results under maximum likelihood estimation.

3.4 Average estimates of intercepts and factor loadings across invariant groups

Table 10 presents the weighted average estimates of factor loadings and intercepts across all invariant groups in each teacher-related construct. These weighted mean values are common for the invariance education systems, and only apply to those invariance systems. The number of such systems can be found in the column next to the weighted mean of intercepts and factor loadings.

As is shown in Table 10, the highest average intercepts for teacher's *Self-efficacy*, for example, is observed on its indicator *Providing challenging tasks for the highest achieving students* ($v = 1.616$)—and the lowest on *Helping students appreciate the value of learning mathematics* ($v = 1.375$). The average factor loading was highest for *Developing students' higher-order thinking skills* ($\lambda = .495$), indicating that this indicator forms an important part of the construct of self-efficacy in teaching mathematics.

3.5 Comparing estimated latent variable means of the teacher-related constructs

Latent variable means of all teacher-related latent constructs that were estimated for the 46 education systems by the aligned model (see Appendix Table 11). Groups can be compared based on these factor means.

3.5.1 Teacher job satisfaction

The latent variable mean of teacher job satisfaction is based on indicators concerning teachers' feelings of contentment with the profession as a whole, their current school, their enthusiasm and pride in their work, and their intention to continue teaching. According to the estimated mean of JS in Fig. 1, students in Japan, Singapore, England, Hong Kong, and Hungary have mathematics teachers with the highest level of job satisfaction as compared to other education systems in TIMSS 2015. Students in Italy, Lithuania, Sweden, South Korea, and New Zealand also have mathematics teachers with relatively low levels of job satisfaction. By contrast, in Chile, Qatar, Thailand, Argentina (Buenos Aires), Kuwait, Oman, Israel, Lebanon, Malaysia, and the United Arab Emirates, students have mathematics teachers who are the least satisfied with their job.

3.5.2 Teacher perception of safe and orderly school

Broadly, SOS refers to whether teachers feel the schools are located in a safe neighborhood and feel the students are respectful. The latent variable mean of

Table 10 Weighted average estimates across invariant groups

	ν	$N\nu$	λ	$N\lambda$
Job satisfaction				
I am content with my profession as a teacher	1.259	39	.423	46
I am satisfied with being a teacher at this school	1.323	40	.376	45
I find my work full of meaning and purpose	1.157	38	.395	42
I am enthusiastic about my job	1.138	44	.466	43
My work inspires me	1.221	42	.516	46
I am proud of the work I do	1.133	45	.401	41
I am going to continue teaching for as long as I can	1.301	33	.468	43
School emphasis on academic success (teachers)				
Teachers' understanding of the school's curricular goals	3.929	42	.459	46
Teachers' degree of success in implementing the school's curriculum	3.682	44	.530	46
Teachers' expectations for student achievement	3.526	28	.533	46
Teachers working together to improve student achievement	3.646	40	.589	46
Teachers' ability to inspire students	3.578	40	.540	46
School condition and resources				
The school building needs significant repair	1.566	40	.540	46
Teachers do not have adequate workspace (e.g., for preparation, collaboration, or meeting with students)	1.384	36	.611	45
Teachers do not have adequate instructional materials and supplies	1.364	41	.660	46
The school classrooms are not cleaned often enough	1.231	38	.380	41
The school classrooms need maintenance work	1.471	42	.529	46
Teachers do not have adequate technological resources	1.516	42	.609	45
Teachers do not have adequate support for using technology	1.517	36	.557	46
Safe and orderly school				
This school is located in a safe neighborhood	1.165	40	.191	39
I feel safe at this school	1.037	45	.184	34
This school's security policies and practices are sufficient	1.166	43	.213	39
The students behave in an orderly manner	1.222	33	.454	46
The students are respectful of the teachers	1.180	42	.446	46
The students respect school property	1.427	45	.457	46
This school has clear rules about student conduct	1.165	45	.271	41
This school's rules are enforced in a fair and consistent manner	1.211	35	.331	45
Self-efficacy				
Inspiring students to learn mathematics	1.416	35	.436	41
Showing students a variety of problem-solving strategies	1.388	43	.398	45
Providing challenging tasks for the highest achieving students	1.616	36	.407	46
Adapting my teaching to engage students' interest	1.463	44	.452	46
Helping students appreciate the value of learning mathematics	1.375	43	.488	43
Assessing student comprehension of mathematics	1.508	34	.399	45
Improving the understanding of struggling students	1.550	36	.445	46
Making mathematics relevant to students	1.483	41	.493	45
Developing students' higher-order thinking skills	1.598	42	.495	46

Note: ν = intercept; $N\nu$ = number of countries are invariance in intercept; λ = factor loading; $N\lambda$ = number of countries are invariance in factor loadings

Fig. 1 Estimated latent variable means for mathematics teachers' job satisfaction

SOS is shown in Fig. 2. The results indicated that students in Botswana, South Africa, Morocco, Turkey, Japan, Italy, Slovenia, South Korea, Sweden, and Jordan had mathematics teachers with the highest levels of perceived school safety. In Argentina (Buenos Aires), Ireland, Kazakhstan, Norway, UAE, Lebanon, Qatar, Singapore, Hong Kong, and Lithuania, students had mathematics teachers with the lowest levels of feeling as though the school was orderly and safe.

3.5.3 Teacher perception of school conditions and resources

SCR refers to school infrastructure, whether teachers have adequate workspace and instructional materials, and whether the school environment is well taken care of. Results for latent mean comparisons can be found in Fig. 3. Students' mathematics teachers in Botswana, South Africa, Turkey, Morocco, Saudi Arabia, Egypt, Jordan, Armenia, Malaysia, and Iran reported the highest levels of satisfaction with school conditions and resources. In UAE, Singapore, and Bahrain, students' mathematics teachers reported the lowest perceptions of SCR.

3.5.4 Teacher perception of school emphasis on academic success

SEAS is indicated by teachers' perceptions of whether teachers understand schools' curricular goals, their success in implementing the curriculum, their expectations for student achievement, and their ability to inspire students. Latent variable means are presented in Fig. 4. Recall that SEAS is reverse coded so countries with the lowest levels show the highest mathematics teacher perceptions of SEAS. Students in Italy, Japan, Russia, Hong Kong, Chile, Hungary, Sweden, Norway, Turkey, and Thailand have mathematics teachers who report the highest levels of SEAS. In Qatar, Malaysia, Oman, Ireland, Canada, South Korea, UAE, Bahrain, and Kazakhstan, students generally have mathematics teachers who report the lowest levels of school emphasis on academic success.

Fig. 2 Estimated latent variable means for mathematics teachers' perceptions of safe and orderly school

3.5.5 Teacher self-efficacy

Latent variable means for TSE are found in Fig. 5. Teacher self-efficacy is measured by teachers' feelings of capacity to inspire students in mathematics, show students a variety of problem-solving strategies, adapt their teaching to engage students, make mathematic relevant, and develop higher-order thinking skills. In Japan, Hong Kong, Singapore, Chinese Taipei, Thailand, Iran, Morocco, New Zealand, Sweden, and England, students have mathematics teachers who report the highest levels of self-efficacy in teaching mathematics. In Qatar, UAE, Bahrain, Lebanon, Oman, Argentina (Buenos Aires), Slovenia, Kazakhstan, and Botswana, students have mathematics teachers with the lowest levels of self-efficacy to teach mathematics.

4 Discussion and concluding remarks

Seeking an optimal alternative to assess measurement invariance of the teacher-related constructs across multiple countries, the current study compared the more restricted

Fig. 3 Estimated latent variable means for mathematics teachers' perceptions of school condition and resources

Fig. 4 Estimated latent variable means for mathematics teachers' perceptions of school emphasis of academic success

traditional MI approach with an alignment optimization method. With TIMSS 2015 data from 46 countries as the empirical basis, the results confirm the initial position of this study. In the traditional MI approach, the level of metric invariance was only reached for three constructs, namely, teacher perception of *School emphasis on academic success*, *School condition and resources*, and teacher *Self-efficacy*. This result implied a limited comparability across countries restricted to the associations between these constructs and other variables being studied. The quest for furthering cross-national comparability is a worthwhile and essential endeavor in the large-scale international studies.

In this study, the purpose of the alignment optimization method is to justify previously unanswerable questions related to group mean comparisons. Scalar invariance was not reached for any of the teacher-related constructs, signifying that under the traditional MI framework, latent factor means could not be validly compared in any case. The results from the alignment optimization approach, however, have demonstrated a different picture, since it takes into account the partial invariance in the parameters of each latent variable indicator and identifies the most optimal

Fig. 5 Estimated latent variable means for mathematics teachers' self-efficacy

measurement invariance pattern when assessing comparability (Asparouhov and Muthén 2014). Departing from the configural invariance models, the current study found a low number of indicators in each construct and country with significant non-invariance. Despite this, all five constructs fell below the non-invariance threshold of 25% suggested by Asparouhov and Muthén (2014). In general, the Monte Carlo simulations confirm the reliability of the majority of the alignment results, with some caution around *Job satisfaction* and *School emphasis on academic success*. These results give valuable information about the specifics of what contributes most to scalar non-invariance. Indeed, the indicator-by-indicator results may be more informative of cultural and societal differences across the constructs than traditional MI approaches.

It was noteworthy that the teacher *Self-efficacy* construct in particular reached acceptable invariance level, as the cultural comparability of self-efficacy has long been the subject of inquiry in teacher quality literature (see Scherer et al. 2016; Vieluf et al. 2013). The current findings support those of Scherer et al. (2016) in suggesting that teacher *Self-efficacy* is a construct that can be generalized across cultures. The results for teacher *Job satisfaction* are more difficult to compare with previous research, as the construct for teacher job satisfaction in TIMSS greatly differs from that in TALIS. In TALIS (2013), the construct includes regretting becoming a teacher, whether teachers would make the same decision if they could decide again, whether they wonder if it would have been better to choose another profession, and the advantages of being teacher outweigh the disadvantages (Zakariya et al. 2020). By contrast, the TIMSS teacher job satisfaction construct includes pride and enthusiasm for the job, ability to feel inspired, intention to continue teaching, and satisfaction with the profession as a whole and with working at the current school. However, both Zakariya et al. (2020) and Zieger et al. (2019) found statistical grounds to compare the construct across some countries. Zieger et al. (2019) present a more conservative approach, however, recommending that comparisons with Chile, Shanghai, Mexico, and Portugal were unreliable. In the current study, only Chile overlaps as an education system with these countries. Interestingly, this is the country in our research which differs the most with previous research. Zakariya et al. (2020) found that teachers in Chile reported among highest levels of job satisfaction compared to other countries, while we find that students in Chile have mathematics teachers with some of the lowest levels of job satisfaction. Perhaps this is a reflection of math teachers differing from other teachers, or perhaps this is a reflection of a more serious issue of comparability. As mentioned JS was a construct that displayed some reliability concerns in the Monte Carlo simulation. This caution may be reflected by other investigation recommendation caution around cross-cultural comparisons of teacher JS (Pepe et al. 2017; Zieger et al. 2019). There is little empirical research on MI and the other constructs, including teacher's perceptions of *School emphasis on academic success*, *Safe and orderly school*, and *School conditions/resources*. The results of this study, therefore, provide the first evidence for the potential of comparability for the majority of these constructs.

Several insights came out of simple observations of the resulting factor mean scores. First, it was possible to detect which countries are on the higher or lower ends of the constructs. As mentioned, Japan, Singapore, England, Hong Kong, and Hungary had the highest levels of mathematics teacher job satisfaction, with Qatar, Chile, Kuwait, Thailand, and Argentina (Buenos Aires) reporting the lowest level of mathematics teacher JS. Interestingly, countries with students with mathematics teachers who

reported the highest levels of job satisfaction also tend to be among the top performers in mathematics in 2015 (for Singapore, Japan, and Hong Kong in particular). However, more research needs to be done to investigate the relationship between these newly constructed means and student outcomes. Our results for teacher job satisfaction differ vastly from those of Zakariya et al. (2020). Given the differing sample (we are focused on mathematics teachers only), as well as entirely different indicators of job satisfaction, as well as different countries included, however, this is not so surprising. In addition, recall that TIMSS samples teachers as representative of students in a country, while TALIS samples teachers as representative of teachers in a country. We are more interested in the former for this paper, as our ultimate interest in cross-national comparison is the comparison of educational contexts of students. For TSE, similar patterns emerged, with top mathematics performers Japan, Singapore, Hong Kong, and Chinese Taipei taking the top ranking positions. Middle Eastern countries such as Qatar, UAE, and Bahrain reported the lowest levels of TSE. The Japanese sample also displayed the highest level of self-efficacy, contradicting the oft discussed cultural tendency in Japan to avoid self-enhancement (Takata 2003). The other constructs did not have such similarly evident clusterings of countries, such as the contrast between East Asian countries (who tended to report higher levels of job satisfaction and self-efficacy) and those situated in the Middle East (who tended to report lower levels of most constructs). It was possible to detect a small group of African countries (Botswana, Morocco, and South Africa) which tended to report high levels of both satisfaction with school conditions and resources as well as perceptions of safety and orderliness in the school. Future research can investigate these differences in more detail and investigate potential hypotheses as to why they exist.

This study has some limitations. As mentioned by Munck et al. (2018), differentiating sources of bias from each other (i.e., method bias related to the instrument versus construct bias, see Schulz 2016) is not possible with this method. Next, interpreting the importance of the non-invariance of individual indicators (as compared to the final average invariance index) is not straightforward. Determining the ultimate degree of comparability rests on the total alignment score. In our study, the Monte Carlo results for JS and SEAS fell below the recommended threshold when the N was reduced to 500, indicating potentially unresolvable issues with the comparability of these constructs. Last, there are some important potential limitations of the alignment optimization method itself which call into question its usefulness as an alternative to the traditional MGCFA approach. Svetina et al. (2016) write that “this sort of latent variable standardization implies that the latent variables are not on the same scale, and as a result, cannot be compared” (p. 128). They and other authors argue that it should be used as primarily an exploratory approach. We believe such an exploratory approach is extremely useful in the context of international comparison. Particularly in the case of research on teacher characteristics, where certain questions continue to be ignored because of obstacles related to the issue of MI.

We believe the significance of the present study outweighs its limitations. First, it demonstrates and supports the possibilities of applying the proposed method to the field of comparative psychological and educational research. Next, as mention extensively throughout this paper, it presents ways for ILSA

researchers to investigate previously unanswered questions related to group mean comparisons of latent constructs. Alignment can be applied to assess the comparability of a myriad of other student-related or school-related constructs. It has implications for policy-related research, given that system level factors may be related to group mean scores. Last, it has important implications for future research investigating the importance of teacher characteristics for student outcomes.

We have several recommendations for future research regarding this method. First, as mentioned, differences in group mean scores and those in the individual indicators can give us important information about cultural differences which should not necessarily exclude their comparison. Future research can investigate such differences with potential cultural conceptions in mind. Next, policy-makers should pay attention to countries which consistently score high on constructs reflecting teacher job satisfaction, self-efficacy, and their working environments. Such countries include Japan, Singapore, Hong Kong, and Chinese Taipei. Similarly, there is much to be learned about countries which consistently score low, such as many countries in the Middle East. Such differences may be attributable to differences in teacher resources and teacher-focused policies. We are also interested in particular in the question of the role of teacher characteristics in student outcomes. Researchers may also use this method to examine first whether JS, SEAS, SCR, SOS, and TSE are comparable across TIMSS surveys, and then to examine changes in teacher characteristics across the last two decades for countries. We can also recommend more in depth comparisons of teacher characteristics across subgroups in participating countries, such as student from disadvantaged socioeconomic backgrounds. Ultimately, further investigations of such questions would yield more insight into the potentially context-dependent aspect of teacher characteristics as they relate to student achievement.

The purpose of international large-scale assessments is to examine differences in educational systems across countries. However, as noted by Scherer et al. (2016) in much public policy research “there is a pre-occupation with cross-cultural differences rather than of cross-cultural generalizability” (p. 4). Herein lies the paradox of research with international large-scale assessments. ILSA and comparative education research necessitate that education systems have differences—but their differences almost never comply with the restrictive statistical rules necessary for cross-country comparison. Although not without limitation, the method outlined in this paper provides one way forward. The growing number of studies using this method suggests possible changes in the future of large-scale assessment research, and scholars are extending its capacity (Marsh et al. 2018). According to Munck et al. (2018), the alignment optimization method can “update existing databases for more efficient further secondary analysis and with meta-information concerning measurement invariance” (p. 687). Measurement invariance has become a problem that all comparative education researchers must eventually face, either by making ill-founded comparisons or avoiding latent factor mean comparisons altogether. This method exists as one promising way that large-scale assessment research may reach its full potential for influencing policy and educational reform.

Funding Open Access funding provided by University of Gothenburg. This project has received funding from the European Union’s Framework Programme for Research and Innovation Horizon 2020 (2014–2020) under the Marie Skłodowska-Curie Grant Agreement No. 765400.

Appendix

Table 11 Estimated factor means of the teacher-related constructs in 46 countries by the aligned model

Country	JS	SEAS	SCR	SOS	TSE
Australia	1.120	.459	.375	1.378	1.162
Bahrain	.704	.976	.286	1.435	.377
Armenia	.701	.466	1.301	.000	1.172
Botswana	1.050	.757	2.195	2.891	.607
Canada	.787	.794	.469	1.334	.917
Chile	.440	.259	.497	1.684	.000
Chinese Taipei	1.117	.441	.583	1.568	1.557
Georgia	.642	.442	1.145	1.290	.989
Hong Kong	1.388	.202	.399	1.089	1.777
Hungary	1.350	.267	1.110	1.395	1.012
Iran	.676	.358	1.229	1.332	1.325
Ireland	.815	1.106	.494	.676	1.044
Israel	.548	.655	.971	1.274	.648
Italy	1.283	– .193	1.184	2.235	1.161
Japan	1.671	– .105	.835	2.303	2.639
Kazakhstan	.571	.829	.762	.691	.594
Jordan	.980	.427	1.364	1.709	.771
Korea	1.164	.995	.691	1.917	1.142
Kuwait	.513	.574	.808	1.345	.761
Lebanon	.568	.440	.569	.833	.393
Lithuania	1.244	.650	.783	1.094	.806
Malaysia	.568	1.232	1.288	1.485	.788
Malta	1.068	.489	.580	1.631	.967
Morocco	.895	.000	1.517	2.623	1.298
Oman	.546	1.139	.549	1.249	.459
New Zealand	1.161	.697	.521	1.223	1.247
Norway	.854	.295	.672	.742	1.148
Qatar	.315	1.299	.000	.862	.352
Russia	1.060	.110	.536	1.248	1.152
Saudi Arabia	.657	.644	1.463	1.570	.913
Singapore	1.481	.322	.280	1.012	1.652
Slovenia	1.095	.461	.340	2.104	.476
South Africa	1.096	.683	1.836	2.837	.648
Sweden	1.228	.282	1.113	1.789	1.231
Thailand	.451	.313	.889	1.455	1.364
UAE	.570	.992	.202	.906	.352
Turkey	1.075	.298	1.774	2.376	.856
Egypt	.000	.410	1.386	1.489	.739
USA	1.156	.535	.388	1.666	1.077

Table 11 (continued)

Country	JS	SEAS	SCR	SOS	TSE
England	1.438	.732	.368	1.328	1.188
UAE-Dubai	.669	.912	.068	.753	.373
UAE-Abu Dhabi	.625	.895	.226	1.051	.479
Canada-Ontario	.724	.668	.610	1.310	.766
Canada-Quebec	.892	1.002	.381	1.365	1.063
Argentina-Buenos Aires	.477	.381	.905	.560	.461

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 21, 495–508. <https://doi.org/10.1080/10705511.2014.919210>.
- Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. Hoboken, NJ: John Wiley & Sons.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin*, 105, 456–466. <https://doi.org/10.1037/0033-2909.105.3.456>.
- Caro, D., Sandoval-Hernandez, A., & Lüdtke, O. (2014). Cultural, social and economic capital constructs in international assessments: an evaluation using structural equation modelling. *School Effectiveness and School Improvement*, 25, 433–450.
- Eriksson, K., Helenius, O., & Ryve, A. (2019). Using TIMSS items to evaluate the effectiveness of different instructional practices. *Instructional Science*, 47, 1–18.
- Goe, L. (2007). The link between teacher quality and student outcomes: a research synthesis. *National comprehensive center for the teacher quality*.
- Gustafsson, J. E. (2018). International large-scale assessments: current status and ways forward. *Scandinavian Journal of Educational Research*, 62, 328–332.
- Hattie, J. (2003). Teachers make a difference: what is the research evidence? In *Paper presented at the Australian Council for Educational Research Annual Conference on Building Teacher Quality, Melbourne*.
- He, J., Barrera-Pedemonte, F., & Bucholz, J. (2018). Cross-cultural comparability of noncognitive constructs in TIMSS and PISA. *Assessment in Education: Principles, Policy & Practice*, 26, 369–385.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. <https://doi.org/10.1080/10705519909540118>.
- Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109–133.
- Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: a meta analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education*, 36(143), 152.

- Lomazzi, V. (2018). Using alignment optimization to test the measurement invariance of gender role attitudes in 59 countries. *Methods, Data, Analyses*, 12, 77–104.
- Marsh, H., et al. (2018). What to do when scalar invariance fails: the extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychological Methods*, 23, 524–545.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13, 127–143.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525–543. <https://doi.org/10.1007/BF02294825>.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Taylor & Francis Group.
- Millsap, R. E., & Kwok, O. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*, 9(1), 93–115. <https://doi.org/10.1037/1082989X.9.1.93>.
- Munck, I. M., Barber, C. H., & Torney-Purta, J. V. (2018). Measurement invariance in comparing attitudes toward immigrants among youth across Europe in 1999 and 2009: The alignment method applied to IEA CIVED and ICCS. *Sociological Methods & Research*, 47, 687–728.
- Muthén, B., & Asparouhov, T. (2014). IRT studies of many groups: the alignment method. *Frontiers in Psychology*, 5, 978. <https://doi.org/10.3389/fpsyg.2014.00978>.
- Muthén, B., & Asparouhov, T. (2018). Recent methods for the study of measurement invariance with many groups: alignment and random effects. *Sociological Methods & Research*, 47, 637–664. <https://doi.org/10.1177/0049124117701488>.
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.
- Nilsen, T., and Gustafsson, J.E. (2016a). The impact of school climate and teacher quality on mathematics achievements: a difference-in-differences approach. (pg 81–95). In *Teacher quality, instructional quality and student outcomes: relationships across countries, cohorts and time*. Springer International Publishing.
- Nilsen, T., and Gustafsson, J.E. (2016b). *Teacher quality, instructional quality and student outcomes: relationships across countries, cohorts and time*. Springer International Publishing.
- Oliveri, M. E., & von Davier, M. (2011). Investigation of model fit and score scale comparability in international assessments. *Psychological Test and Assessment Modeling*, 53, 315–333.
- Pepe, A., Addimando, L., & Veronese, G. (2017). Measuring teacher job-satisfaction: assessing invariance in the teacher job satisfaction scale (TJSS) across six countries. *European Journal of Psychology*, 13, 396–416.
- Raudenbush, S. W., Rowan, B., & Fai Cheong, Y. (1992). Contextual effects on the self-perceived efficacy of high school teachers. *Sociology of Education*, 65, 160–167.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. (2005). Teachers, schools and academic achievement. *Econometrica*, 73, 417–458.
- Rutkowski, L., & Rutkowski, D. (2010). Getting it ‘better’: the importance of improving background questionnaires in international large-scale assessment. *Journal of Curriculum Studies*, 42, 411–430.
- Rutkowski, D., & Rutkowski, L. (2013). Measuring socioeconomic background in PISA: one size might not fit all. *Research in Comparative and International Education*, 8, 259–278.
- Rutkowski, D., & Rutkowski, L. (2017). Improving the comparability and local usefulness of international assessments: a look back and a way forward. *Scandinavian Journal of Educational Research*, 62, 354–367.
- Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large scale international surveys. *Educational and Psychological Measurement*, 74, 31–57.
- Scherer, R., Jansen, M., Nilsen, T., Areepattamannil, S., & Marsh, H. W. (2016). The quest for comparability: studying the invariance of the teacher’s sense of self-efficacy (TSES) measure across countries. *PLoS One*, 11, 1–29.
- Schulz, W. (2016). Reviewing measurement invariance of questionnaire constructs in cross-national research: examples from ICCS 2016. *Australian Council for Educational Research*. Paper prepared for the Annual Meeting of the American Educational Research Association, Washington D.C.
- Segeritz, M., & Pant, H. A. (2013). Do they feel the same way about math? Testing measurement invariance of the PISA “Students’ Approaches to Learning” instrument across immigrant groups within Germany. *Educational and Psychological Measurement*, 73, 601–630.
- Strong, M. (2011). *The highly qualified teacher: what is teacher quality and how do we measure it?* New York, NY: Teachers College Press.

- Svetina, D., Rutkowski, L., & Rutkowski, D. (2016). Multiple group invariance with categorical outcomes using updated guidelines: an illustration using Mplus and the lavaan/semTools packages. *Teacher's Corner*, 111–130.
- Takata, T. (2003). Self-enhancement and self-criticism in Japanese culture: an experimental analysis. *Journal of Cross-Cultural Psychology*, 34, 542–551.
- Teaching and Learning International Survey (TALIS). (2013). *Technical report*. Paris: OECD Publishing.
- Toropova, A., Johansson, S., & Myrberg, E. (2019). The role of teacher characteristics for student achievement in mathematics and student perceptions of instructional quality. *Education Inquiry*, 10, 1–25.
- van de Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Editorial: measurement invariance. *Frontiers in Psychology*, 6, 1–5. <http://dx.doi.org/10.3389/fpsyg.2015.01064>.
- Vieluf, S., Kunter, M., & van de Vijver, F. J. (2013). Teacher self-efficacy in cross-national perspective. *Teaching and Teacher Education*, 35, 92–103.
- Zakariya, Y. F., Bjorkestol, K., & Nilsen, H. K. (2020). Teacher job satisfaction across 38 countries and economies: an alignment optimization approach to a cross-cultural mean comparison. *International Journal of Educational Research*, 101, 1–10.
- Zieger, L., Sims, S., & Jerrim, J. P. (2019). Comparing teachers' job satisfaction across countries: multiple pairwise measurement approach. *Educational Measurement: Issues and Practice*, 38, 75–85.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Leah Natasha Glassow¹ • Victoria Rolfe¹ • Kajsa Yang Hansen¹

Victoria Rolfe
victoria.rolfe@gu.se

Kajsa Yang Hansen
kajsa.yang-hansen@ped.gu.se

¹ Department of Education and Special Education, University of Gothenburg, Gothenburg, Sweden