

RESEARCH

Open Access



Measurement of job motivation in TEDS-M: testing for invariance across countries and cultures

Christin Laschke^{1*} and Sigrid Blömeke²

*Correspondence:
christin.laschke@gmail.com

¹ Institut für
Erziehungswissenschaften,
Humboldt-Universität zu
Berlin, Unter den Linden 6,
10099 Berlin, Germany
Full list of author information
is available at the end of the
article

Abstract

The paper presents the challenges of cross-country and cross-cultural research on the motivation to become a mathematics teacher based on data from the “Teacher Education and Development Study in Mathematics (TEDS-M)”. Referring to studies from cross-cultural psychology, measurement invariance (MI) of constructs representing different motivations to become a teacher was examined in confirmatory factor analysis (CFA) across the countries that participated in TEDS-M. The data supported metric invariance which means that comparing relationships between motivation and other constructs across countries is permitted, with the exception of extrinsic motivation in Taiwan. Scalar invariance was not supported by the data across countries but across cultures: Scale means can be compared between Germany, Switzerland and (with regard to intrinsic motivation) Norway and Poland as well as between Singapore and Taiwan (with regard to the intrinsic motivation) and Malaysia, Philippines and Thailand (again regarding intrinsic motivation).

Keywords: Measurement invariance, Cross-cultural comparison, Job motivation

Background

Many countries face difficulties in recruiting teaching candidates particularly for mathematics. To learn what motivates teacher candidates to go into teaching can therefore be useful from a policy perspective. Motivations to teach have already been assessed within many national studies. However, comparative evidence of future teachers’ job motivation is rare although countries could learn from each other about potential factors which motivate people to become a teacher, how to recruit teaching candidates or about potential outcomes of teachers’ job motivation.

The largest comparative study that provides information about mathematics future teachers’ job motivation is available from the Teacher Education and Development Study in Mathematics (TEDS-M; Tatto et al. 2008). Based on these data, we are for the first time able to examine research questions related to this construct across countries. An important challenge of such large-scale assessments though is to ensure that self-reported data collected are measuring the object of interest in the respective countries and cultures in the same way. Due to different frames of references between countries and cultures, cultural response biases, translation errors, or cultural differences in

understanding the underlying construct, the comparability of constructs could be threatened (Markus and Kitayama 1991; Rutkowski and Svetina 2014, p. 51). This is particularly true for affective constructs as motivation, which are assessed via self-reports (Van de Vijver and Tanzer 2004). For that reason, testing for different levels of equivalence of scales is required before relating motivation to teach to other constructs or even to compare scale means, in order to avoid inappropriate use of data across different groups.

The aim of the present study is therefore to examine, whether the set of motivations to become a teacher assessed by TEDS-M could be used to construct motivation scales which are related to theories of motivation and which are invariant across the different countries that participated in TEDS-M.

Theoretical framework: motivation and choices

The question what is motivating people to become a teacher can be examined within the expectancy-value framework, “the most comprehensive motivational model for explaining academic and career choices” (Watt and Richardson 2007, p. 170). Following the expectancy-value theory (Wigfield and Eccles 2000), decision making such as choosing teacher education as the field of study is determined by people’s values and expectancies, which are shaped by goals and self-schemata, which in turn are influenced by cultural and social norms and individual’s perception of it (Wigfield and Eccles 2000, p. 69).

The values component of the expectancy-value model implies, in addition to *costs*, (which refer to perceived negative consequences as effort or emotional expenditure) and *attainment value* (the sense of self and identity resulting in subjective goals, which determine the importance for the individual of doing well in specific tasks), the *intrinsic value* and the *utility value*. Intrinsic value refers to the enjoyment an individual gets from performing an activity. Utility value refers to the usefulness of doing an activity for the individual in the future, capturing extrinsic motivation (Wigfield and Eccles 2000).

One important source of the values component of the expectancy-value model is Deci and Ryan’s self-determination theory (1985), which distinguishes between intrinsic and extrinsic motivation. According to the self-determination theory, individuals have a natural need for competence, autonomy, and social relatedness. It is assumed, that individuals are intrinsically motivated to pursue a goal to satisfy their natural needs and to feel self-determined. Intrinsic motivation represents therefore the prototype of self-determined behaviour, and intrinsically motivated behaviour is strongly related to feeling competent, autonomous and socially related.

Extrinsically motivated behaviour is initially not self-determined, but can be transferred into self-determined behaviour by the processes of internalisation and integration. Through the process of internalisation, external values are taken in. By integration, the internalized values become embedded into the sense of the individual’s self and function as drivers of pursuing a goal.

Within the teacher education literature, it is common to operationalize intrinsic motivation as enjoyment in teaching or interest in a subject. Extrinsic motivation usually addresses conditions and amenities as job security or salary. In line with that, in TEDS-M future teacher’s motivation to become a teacher was assessed by different items which could be classified into intrinsic and extrinsic motivation. On the one hand, future teachers were asked, to what extent their expected talent for teaching, the wish to work

with young people and to influence the next generation and the perception of teaching as a challenging job constitute reasons to become a teacher; aspects, which represent an intrinsic pedagogical motivation. Future teachers also were asked, whether a love to mathematics is a reason to become a teacher, which is capturing an intrinsic subject specific motivation. On the other hand, they were asked, whether they are attracted by the availability of teaching positions, by teacher salaries and by the long-term security associated with being a teacher, items which represent an extrinsic motivation (Laschke and Blömeke 2014; Blömeke et al. 2010).

Theoretical framework: the role of the societal context

Social contextual conditions can catalyse or undermine the influence of intrinsic motivation on learning and achievement by meeting or rejecting the learners' needs of autonomy, relatedness, and competence. The same applies to extrinsic motivation by fostering or hindering the processes of internalization and integration (Deci and Ryan 1985; Ryan and Deci 2000).

That is particularly important for the context of education. For example, the more a learner feels autonomous, the better his or her performance is, besides other positive outcomes (see Ryan and Deci 2000, p. 63). Intrinsic motivation is fostering high-quality learning and creativity (Ryan and Deci 2000, p. 55). Feeling socially related to teachers and parents is facilitating the willingness to accept their values. And learners' need to feel competent can be satisfied by providing a goal which the individual understands and is able to succeed at (Ryan and Deci 2000).

Within the self-determination theory as well as the expectancy-value theory, the social surrounding and the concept of the self are thus important factors. The concept of the self has been expanded and refined by addressing the social environment (Ryan and Deci 2000). However, these characteristics can vary between countries with different cultural orientations (Markus and Kitayama 1991; Hofstede 1986; Triandis 1995). This applies especially strongly to individualistic versus collectivistic orientated cultures (Hofstede 1986). Markus and Kitayama (1991) discriminate therefore between the *independent* and the *interdependent* self. The distinction is made due to the different roles of an individual within different societies and the differences in individual self-conception. In cultures with an individualistic orientation, the individual and its personal fulfilment and independence is more strongly emphasized than in collectivist orientated cultures, where the group and the relationship of the individual to group members is most important.

These differences may lead to differences in the importance of aspects such as the need of autonomy and relatedness, which are emphasized in the self-determination theory. As Triandis (1995) pointed out, "Individualists focus on the achievement of personal goals, by themselves, for the purpose of pleasure, autonomy, and self-realization. Collectivists focus on the achievement of group goals, by the group, for the purpose of group well-being, relationships, togetherness, the common good, and collective utility." (Triandis 1995, p. 1). According to that, the factors which catalyse or undermine intrinsic motivation and the processes of internalization and integration may differ between individualistic and collectivistic cultures.

Moreover, the concepts of intrinsic and extrinsic motivation and their respective importance and acceptance can differ. For example, the literature points out that in East

Asian collectivistic cultures extrinsic pressure is an important factor in education which results in the inner will to fulfill the expectations of the group and of teachers (Leung 2001, pp. 42–43). Extrinsic motivation is therefore well-accepted as important driving force within education whereas in Western individualistic cultures aptitude and enjoyment, which constitute indicators of an intrinsic motivation, are the preferable form of motivation. Extrinsic motivation is associated with not desirable, pragmatic reasons (Vollstedt 2011, p. 76; Leung 2001, pp. 41–42).

These differences in normative preferences of types of motivation are reflected in results of cross-cultural studies. Extrinsic motivation to learn mathematics is according to these negatively related to mathematics achievement in Western countries, whereas in East Asian countries extrinsic motivation to learn mathematics is positively related to mathematics achievement (Zhu and Leung 2011). This is in line with the view, that extrinsic motivation is supportive for achievement in East Asia but not in the West (Leung 2001; Watkins and Biggs 1996). However, there are also contradictory results revealed for example by Shin et al. (2009), who found a larger positive effect of extrinsic motivation to mathematics achievement for American students than for East Asians. Nevertheless, the results point to differences between collectivistic orientated East Asian countries and individualistic orientated Western countries.

Differences in conceptualizations and importance of affective constructs could not only exist between collectivistic and individualistic orientated countries as defined by Hofstede (1986), caused by the more social orientation in collectivistic and stronger individual orientation in individualistic countries, but also between groups of countries contrasted by global region and its cultural and educational tradition. In case of Asia for example a group of Singapore and Taiwan and a group of Malaysia, Philippines and Thailand can be discriminated by region and particularly by cultural and educational roots. The first group is located in the same region and share its cultural and educational tradition, in the sense that their culture is deeply rooted in the Confucian heritage (Leung et al. 2006). Following the ideas of Confucius, education and learning is playing a key role for individual and its contribution to the society in Taiwan and Singapore (Salili 1995). This does not apply to Malaysia, Philippines and Thailand. This group of countries is not shaped by the Confucian culture, but closer to the culture of South Asian societies (House et al. 2004), where the belief systems of Christianity, as in the Philippines, Islam, as in Malaysia and Buddhism, as in Thailand, are more represented (Banks 2012, p. 369). In none of these belief systems education is as much emphasized as in the Confucian tradition (Zhao 2011). Therefore, education and academic achievement should not to the extend valued in this countries as it applies in the Confucian Taiwan and Singapore.

Cross-cultural studies assessing affective constructs

Despite of much effort to study motivations to become a teacher across countries, there is a lack of comparative evidences, caused by the variety of instruments, which differ substantially between the different studies. Exceptions are studies using the FIT-Choice scale,¹ which was developed in Australia and applied in different countries. According to Watt et al. (2012), the FIT-Choice scale is invariant with regard to the loading patterns

¹ Factors influencing teaching choice (Watt and Richardson 2007).

and intercepts across the USA, Australia, Germany and Norway. However, specific motivation to become a teacher in the fields of science, technology, engineering or mathematics was studied in Australia only (Watt et al. 2009, 2013). Thus, the comparative studies applying the FIT-Choice scale provide valuable information about teachers across all subjects but not specifically for mathematics teachers.

In contrast, TEDS-M provides a database to compare future mathematics teachers' motivation to become a teacher in different countries and cultures. The lower-secondary TEDS-M study included future teachers who were prepared to teach mathematics in grade 8 (Tatto et al. 2008). Assessing professional knowledge and beliefs of student teachers was the main objective of the study but also background characteristics and the motivation to become a teacher were surveyed. The TEDS-M instruments resulted from a collaborative process of careful development and translation accomplished by the national research coordinators of each participating country and other experts under supervision of the International Association for the Evaluation of Educational Achievement (IEA) (Tatto 2013). Nevertheless, testing for comparability of the measurement instruments is required to ensure meaningful comparisons, since construct equivalence across countries and cultures is not guaranteed albeit careful and elaborative scale construction (Nagengast and Marsh 2014).

For TEDS-M instruments assessing professional knowledge, cross-country measurement invariance and item functioning were examined as presented in Blömeke et al. (2011, 2013) and Tatto (2012, 2013). However, this is not only relevant for achievement tests but becomes particularly important if data are collected by self-reports, which are more likely vulnerable to biases caused by different meanings of constructs or different response styles (Van de Vijver and Tanzer 2004). An incongruity of self-reported data with test results has already been shown based on the TEDS-M data by König et al. (2012), who found a low correlation between future teachers' pedagogical knowledge and their sense of preparedness for the teaching profession based on the German TEDS-M data, and by Blömeke (2014), who showed that future teacher's evaluations of teacher education quality and effectiveness are only weakly correlated with their professional knowledge.

Whether instruments assessing affective constructs such as motivation are measuring in the same way across different groups has also been examined in other large-scale assessments. Artelt (2005) showed based on data of PISA 2000 that the scales assessing intrinsic and extrinsic motivation of students were only metrically equivalent across the 26 participating countries which means that factor loadings were invariant so that the relations of constructs can be compared across countries but not the means (Artelt 2005, p. 249). Similar results of metric but not scalar measurement invariance were revealed by Segeritz and Pant (2013) who examined scales assessing preferences to learn mathematics, beliefs and self-related cognitions used in the PISA study 2003 with respect to different ethnic groups within Germany.

Levels of measurement invariance (MI) and sources of measurement non-invariance

Comparing scale scores of constructs across groups produces meaningful results only if the scales measure the same construct in all of the groups (Van de Vijver and Leung 2000). In order to ascertain such equivalence, MI is to be established by examining the

interrelations between items and the scale representing the underlying trait (Chen 2008, p. 1006). It is common to test for MI by using multiple-group confirmatory factor analysis (MG-CFA). According to the bottom-up approach of Brown (2006), at first configural invariance, the basic level of measurement invariance, is to be examined. If in each of the groups the same items are associated with the same latent factors, configural invariance is established. As a second step, it is to be tested whether the factor loadings are invariant, to ensure that the unit of measurement is identical across the groups. Invariance of factor loadings allows to compare relationships between the construct assessed and others across groups. The third step is to test for scalar invariance. If the intercepts are invariant, the items have the same origins in all groups. Only if a scale consists of the same units of measurement and the same origin, it is allowed to compare factor means across groups (Chen 2008).

Country and culture specific concepts and conditions can cause a lack in MI. For example, if a construct is more complex in one country or culture than in another, the number of items underlying a latent factor could vary and therefore no configural invariance exists (Kwan et al. 2002). If the conceptual framework such as the definitions and meanings of a construct are not congruent in all of the countries or cultures of interest, loading invariance can be threatened (Cheung and Rensvold 2000). Another source of measurement non-equivalence is constituted if response styles vary by culture. According to cross-cultural research, East Asians as Taiwanese tend to avoid extreme response categories and are more likely to use middle response categories compared to Western respondents (Chen et al. 1995). According to Hui and Triandis (1989), differences in response styles appear between cultures if four-point rating scales are used but not with ten-point rating scales. Whereas the tendency to use extreme or neutral responses affects the invariance of factor loadings, an acquiescence response style could result in a lack of invariance of loadings as well as a lack of invariance of intercepts (Cheung and Rensvold 2000). A response style which is adopting social desirability could result in a lack of invariance of intercepts (Chen and West 2008). Thus, a wide range of sources which could harm the comparability of constructs between countries and cultures exists.

Study objectives/research purpose

The present study examines whether the instruments applied in TEDS-M to assess future mathematics teachers' job motivation could be used to construct cross-country equivalent motivation scales.

The data of lower-secondary future teachers from Chile, Germany, Malaysia, Norway, Oman, Philippines, Poland, Russia, Singapore, Switzerland, Taiwan, Thailand and USA are used. As mentioned above, culture specific differences in the concepts of intrinsic and extrinsic motivation may exist besides cultural differences in response styles. Invariance of the factor structure (configural measurement invariance), of the factor loadings (metric invariance) and of the intercepts (scalar invariance) was tested.

Methods

Sample

In 15 countries TEDS-M was conducted in 2008 to test student teachers who were intending to teach mathematics in lower-secondary schools, identified by the criterion

Table 1 Sample sizes (cases with missing values on all variables and teacher preparation units with less than four future teachers were excluded)

Sample	N	Sample	N
Chile	701	Russia	2105
Germany	765	Singapore	389
Malaysia	382	Switzerland	140
Norway	540	Taiwan	365
Oman	237	Thailand	650
Philippines	632	USA	443
Poland	288		

of gaining a license to teach mathematics in grade 8 (Tatto et al. 2008). Botswana and Georgia were excluded from the analysis because of their sample sizes smaller than $N = 100$, in order to meet the requirement of sufficiently large sample sizes in MG-CFA. The remaining samples are varying between 140 and 2105 participants (see Table 1).

The international sampling of TEDS-M followed a stratified multistage probability sampling design. Randomly selected institutions, preparing student teachers, were divided into subgroups by level (preparation for primary and/or lower secondary school), route (consecutive vs. concurrent program) and program-type (preparation for primary and/or secondary level with/without focus on mathematics) called teacher preparation units (TPUs). Within the TPUs the student teachers were selected randomly if the number of future teachers was higher than 30, if there were less than 30 future teachers within a TPU all of them were surveyed. The latter applies to Oman, Norway,² Switzerland, Singapore and Taiwan (Tatto 2013, p. 90). In order to obtain robust estimates, teacher preparation units with less than four student teachers were excluded. The cluster structure could not be taken into account because the number of clusters was in some countries smaller than the number of parameters to estimate. Neglecting the cluster structure may affect the estimation of standard errors, a constraint important to recognize when interpreting the results.

Instruments

In order to assess the motivation to become a teacher, TEDS-M participants were asked to rate the following statements on four-point rating scales (1: “not a reason” through 4: “a major reason”).³

- A I am attracted by the availability of teaching positions
- B I believe I have a talent for teaching
- C I like working with young people
- D I am attracted by teacher salaries
- E I want to have an influence on the next generation
- F I see teaching as a challenging job
- G I seek the long-term security associated with being a teacher

² Norway did not meet the sample requirements of TEDS-M, the response rate was less than 60 %.

³ Domain-specific motivation was also assessed in TEDS-M, by asking whether loving mathematics is a reason to become a teacher. That item doesn't match the situation in every country, caused by different roles of mathematics in teacher education or schooling in the respective countries. In some countries generalist teachers, in other countries specialist for the subject mathematics were prepared. The domain-specific motivation was therefore excluded from the analyses.

Psychometric analysis including all TEDS-M countries confirmed as expected two latent factors of motivation, namely intrinsic pedagogical motivation (“I believe I have talent for teaching”, “I like working with young people” “I see teaching as a challenging job”, “I want to have an influence on the next generation.”) and extrinsic motivation (“I’m attracted by the availability of teaching positions”, “I’m attracted by teacher salaries”, “I seek the long-term security”) (Laschke and Blömeke 2014).

Procedure

MI was tested by using MG-CFA (Vandenberg and Lance 2000). Starting from the psychometric analysis, which confirmed the two latent factors intrinsic pedagogical motivation and extrinsic motivation the model in Fig. 1 was tested.

Following the approach of Brown (2006), at first the instruments were tested for configural invariance, second for metric invariance and finally for scalar invariance. The analyses were carried out by using the robust maximum likelihood (MLR) estimator (Satorra and Bentler 2001) and a sandwich-type covariance matrix to compute standard errors and Chi square statistics robust to non-normality of the data (Yuan and Bentler 2000). Although the WLSMV estimator is required, if responses have to be given on rating scales with four or fewer points (Sass et al. 2014; Rhemtulla et al. 2012; Flora and Curran 2004), the MLR estimator was applied to avoid the necessity to collapse meaningful categories. In the Swiss sample none of the future teachers rated the category “not a reason” for the statements “I believe I have talent for teaching” and “I like working with young people.” The estimator WLSMV is not able to handle categories without observations. Nevertheless, the results obtained by using the MLR estimator were validated by estimations with WLSMV whenever possible. Full information maximum likelihood (FIML) estimation, integrating missing data analyses and parameter estimation under the missing at random assumption, was used to handle partially missing data (Little and Rubin 2014). All analyses were conducted in the software package Mplus 7.4 (Muthén and Muthén 1998–2015).

To evaluate to what extent the models specified fit the data, absolute and incremental fit indices were used. X^2 is testing the null hypothesis that the covariance matrix implied by the model is equal to the population covariance matrix. Since X^2 test is sensitive to the sample size and the complexity of a model, the ratio of X^2 and the

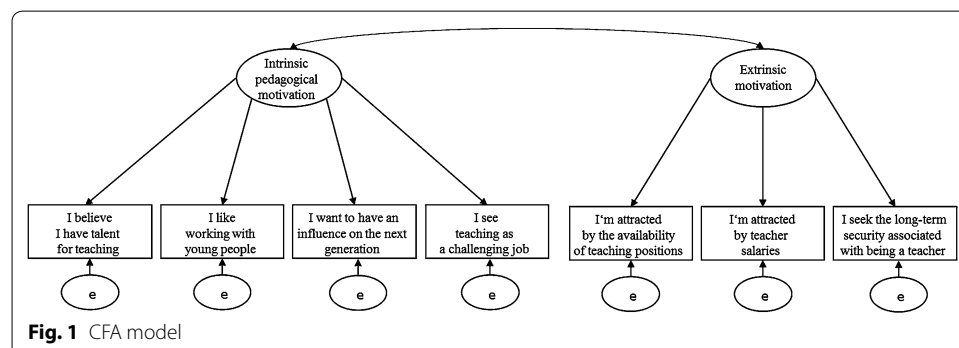


Fig. 1 CFA model

degrees of freedom (df) was computed. X^2/df should be small, an estimate of $X^2/df \leq 3 > 2$ indicates an acceptable, an estimate of $X^2/df \leq 2$ a good model fit (Schermelleh-Engel et al. 2003). RSMEA and SRMR are measuring whether the estimated model reproduces well the observed covariance matrix. For RSMEA and SRMR the following values are recommended: RSMEA and SRMR $< .08$ point to an acceptable model fit, RSMEA and SRMR $< .05$ point to a good fit (Hu and Bentler 1999). The formulas to compute RSMEA and SRMR contain the X^2 value. Both indices are therefore sensitive to sample size.

The comparative fit index (CFI) and the Tucker Lewis index (TLI) assess to what extent the model estimated reproduces the observed covariance matrix better than a baseline model that is assuming all observed variables are uncorrelated. According to Hu and Bentler (1999), CFI and TLI $> .90$ point to an acceptable fit, CFI and TLI $> .95$ indicate a good fit of the model. CFI's performance is relatively unaffected by sample size (Hu and Bentler 1998).

For evaluating the significance of changes of model fit after restricting models within the MI procedure ΔCFI was used. Results of simulation studies provided by Rutkowski and Svetina (2014) suggest the following cutoffs: To determine loading invariance a change of CFI $< .020$ besides a change in RMSEA or SRMR $< .010$ is recommended. Regarding to the determination of equivalence of intercepts a change in CFI, RMSEA and SRMR $< .010$ indicate invariance. During the process of testing for MI, model modifications were conducted post hoc if modification indices matched theoretical justifications.

If the requirements of MI did not apply to all countries, MI was examined within groups of countries. The selection of the groups was theoretically driven. Countries which share a cultural tradition with respect to an individualistic versus collectivistic orientation (Hofstede 1986) were combined in one group. According to Hofstede's individualism scale (IDV), which is ranging from 0 (strongly collectivistic orientated) to 100 (strongly individualistic orientated), Norway (IDV = 69), Switzerland (IDV = 68) Germany (IDV = 67) and Poland (IDV = 60) belong to the more individualistic orientated group. Philippines (IDV = 32), Malaysia (IDV = 26), Chile (IDV = 23) Singapore (IDV = 20), Thailand (IDV = 20), and Taiwan (IDV = 17) belong to a more collectivistic orientated group.⁴ If MI could not be established within these two groups, the analyses was carried out in subgroups more narrowly defined through shared cultures and regions.

Results

Testing for configural invariance

As a first step, the measurement model of intrinsic and extrinsic motivation was tested separately in each country to confirm within-country model fit. According to Hu and Bentler (1999), the measurement model fits well in nearly all countries (see Table 2). However, in the USA and Russia model fit could not be confirmed which means that the theoretical model of intrinsic and extrinsic motivation does not fit to the data of

⁴ For Oman there is no IDV value available since the scale has not been used in this country. For that reason Oman was not added to a group.

Table 2 Fit of the country specific measurement models

Sample	CFI	TLI	RMSEA	SRMR	χ^2/df
Chile	0.92	0.86	0.06	0.04	3.6***
Germany	0.96	0.93	0.03	0.05	1.5
Malaysia	0.92	0.86	0.06	0.04	3.6***
Norway	0.91	0.85	0.06	0.04	2.6***
Oman	0.96	0.94	0.05	0.04	1.6*
Philippines	1.00	0.99	0.01	0.03	1.1
Poland	0.93	0.89	0.06	0.04	2.2***
Russia	0.84	0.74	0.12	0.06	31.2***
Singapore	0.95	0.93	0.07	0.05	2.7***
Switzerland	1.00	0.99	0.02	0.05	1.0
Taiwan	0.95	0.91	0.07	0.05	3.0***
Thailand	0.97	0.95	0.04	0.03	2.0**
USA	0.86	0.77	0.06	0.05	2.9***

CFI comparative fit index, TLI Tucker Lewis index, RMSEA root mean square of approximation, SRMR standardized root mean square residual, *df* degrees of freedom

* $p < .05$, ** $p < .01$, *** $p < .001$

future teachers in these two countries. Consequently, configural invariance across countries could not be confirmed if the USA and Russia were included. They were therefore excluded from further analyses. Without USA and Russia, configural invariance could be established (see Table 3). According to the CFI, that is insensitive to sample size, the model fits particularly well for Germany, Oman, Philippines, Singapore, Switzerland, Taiwan and Thailand and in an acceptable way for Chile, Malaysia, Norway and Poland (Hu and Bentler 1999).

Testing for metric invariance

Since configural invariance of the model was supported by the data for Chile, Germany, Malaysia, Norway, Oman, Philippines, Poland, Singapore, Switzerland, Taiwan, and Thailand, as a next step metric invariance can be examined. For that purpose, the factor loadings are constrained to be equal across the countries in the model. To decide whether metric invariance exists, the fit of the constrained model is to be compared with the fit of the unconstrained baseline model.

As shown in Table 3 (model 3 in comparison to model 2) the fit of the model is declining after constraining the model. The change of CFI indicates a substantial discrepancy between the two models (Rutkowski and Svetina 2014). Following the information revealed by modifications indices, freeing the factor loading of item A of the extrinsic-motivation scale (“I am attracted by the availability of teaching positions”) for Taiwan would substantially improve the model fit. From the information available about employment conditions, freeing the loading of this particular item is in line with the Taiwanese situation compared to other countries. Whereas typically a strong need of mathematics teachers exists, it is difficult to find a teaching position in Taiwan because of the high number of graduates applying for one teaching job (Li et al. 2011). With the factor loading of one item freed up in one country, the fit of the partially metric model does not differ substantially from the fully unconstrained baseline model anymore (see model 3a

Table 3 Goodness of fit indices for measurement invariance between countries

Invariance model	χ^2/df	CFI	TLI	RMSEA	SRMR	Compare with	ΔCFI
(2) Fully unconstrained baseline model	2.1***	0.95	0.92	0.05	0.04		
(3) Metric invariance	2.4***	0.92	0.90	0.05	0.07	(2)	0.03
(3a) Metric invariance (Item A for TW free) ^a	2.2***	0.93	0.92	0.05	0.06	(2)	0.02
(4) Scalar invariance (across all countries)	9.8***	0.33	0.36	0.14	0.14	(3a)	0.06
(5) Scalar invariance (across individualistic countries)	3.5***	0.84	0.82	0.07	0.09	(3a)	0.09
(5a) Scalar invariance (Item D for NW and I for PL free) ^b	2.2***	0.92	0.91	0.05	0.07	(3a)	0.01
(6) Scalar invariance (across collectivistic countries)	5.3***	0.70	0.69	0.10	0.10	(3a)	0.02
(7) Scalar invariance across Philippines, Malaysia and Thailand	2.8***	0.89	0.87	0.06	0.07	(3a)	0.04
(7a) Scalar invariance across Philippines, Malaysia (G free) and Thailand (G free) ^c	2.3***	0.92	0.91	0.05	0.06	(3a)	0.01
(8) Scalar invariance across Taiwan and Singapore	2.5***	0.91	0.89	0.06	0.07	(3a)	0.02
(8) Scalar invariance across Taiwan (B free) ^d and Singapore	2.2***	0.92	0.91	0.05	0.06	(3a)	0.01

df degrees of freedom, *CFI* comparative fit Index, *TLI* Tucker Lewis index, *RMSEA* root mean square of approximation, *SRMR* standardized root mean square residual, χ^2 and *df* are estimated by MLR, χ^2 -difference cannot be compared by subtracting χ^2 and *df*, $\Delta\chi^2$ and Δdf were estimated considering the scaling correction factor (Satorra and Bentler 2001; Muthén and Muthén 2010)

* $p < .05$, ** $p < .01$, *** $p < .001$

^a Factor loadings of item A "I am attracted by the availability of teaching positions" for Taiwan were set free to vary

^b Intercepts of item D "I am attracted by teacher salaries" for Norway and item G "I seek the long-term security associated with being a teacher" for Poland were set free to vary

^c Intercepts of item G "I seek the long-term security associated with being a teacher" for Thailand and Malaysia were set free to vary

^d Intercepts of item A "I am attracted by the availability of teaching positions" for Taiwan were set free to vary

in Table 3). The $\Delta CFI < 0.02$ is in line with the cut off value provided by Rutkowski and Svetina (2014). Therefore, comparing relationships between TEDS-M is permitted.

Testing for scalar invariance

To test for scalar invariance, the intercepts of the items were set equal over the countries. The constrained model does not fit to the data at all (model 4 in Table 3). The fit indices decline correspondingly beyond acceptable thresholds. Therefore, it is to conclude, that the point of origin of the items is not the same across all the countries. Relaxing restrictions did not increase the model fit sufficiently.

As pointed out in the framework, this results does not come unexpected. So, the next step is to test invariance of intercepts separately by groups of countries, defined by Hofstede's individualism scale. In each model, the intercepts over the countries in the respective subgroup were set equal, while freely estimating the intercepts for the other countries. The fit of the models were compared with the fit of the overall country reference model (model 3a). The fit indices of model 5 for the individualistic orientated countries Norway, Switzerland, Germany and Poland are missing the cut off criteria. But free estimation of item D of the extrinsic-motivation scale ("I am attracted by teacher salaries") for Norway and item G "I seek the long-term security associated with being a teacher" in Poland results in a model fit (model 5a) which is not substantially different from the reference model anymore, as the ΔCFI , which should be smaller than 0.01

(Rutkowski and Svetina 2014), points to. Student teachers in Norway tended to rate item D lower than German, Swiss and Polish student teachers which is in line with the working conditions of teachers in Norway. Norwegian teachers' salary and life-span income is significantly lower in comparison to similarly educated professionals. In Poland the professional advancement is defined by different stages, while the teachers in the lower stages are employed on the basis of an ordinary employment agreement. Teachers at every stage has to provide evidences of their development, a procedure, that feels as a burden for most teachers (Carnoy et al. 2009; Schwille and Ingvarson 2013; OECD 2014). Thus, a conceptual justification exists for freeing up the estimation of the salary item's intercept in Norway and the long-term security item's intercept in Poland.

For the group of countries with collectivistic orientation, namely Philippines, Malaysia, Chile, Singapore, Thailand, and Taiwan, the fit indices are far away the acceptable cut off criteria (model 6). Relaxing restrictions could not enhance them.

Hence, as a final step, scalar measurement invariance is to be examined for subgroups, and these are collectivistic countries of the same global region that share in addition to societal communalities a common cultural and educational tradition. The data revealed that for these groups partial scalar invariance can in fact be established.

As hypothesized a subgroup consists of the countries Philippines, Malaysia and Thailand. According to the fit indices, the model with equal intercepts across the three countries (model 7) is missing the cut off criteria provided by Rutkowski and Svetina (2014). Inspecting the modification indices point to relaxing the conditions with respect to item G ("I seek the long-term security associated with being a teacher") for Malaysia and Thailand. In Malaysia the teachers are government servants, which enjoy different amenities including job security. Since the number of enrollments of school students is expanding every year, the demand of teachers is increasing (Schwille and Ingvarson 2013). In Thailand civil service teachers are promoted from one qualification level to the next higher one automatically by working a required period as a teacher (Schwille and Ingvarson 2013). Hence, around 90 % of the future secondary mathematics teachers, who participated in TEDS-M, agreed that teachers have a secure job (Laschke and Blömeke 2014). Freeing the intercept estimation of this item resulted in a model (model 7a) which is not substantially different from the reference model 3a, according to the $\Delta CFI < 0.01$ (Rutkowski and Svetina 2014).

For a second subgroup consisting of Taiwan and Singapore, scalar invariance could be established after freeing the intercepts for item A "I'm attracted by the availability of teaching positions" in Taiwan (model 8). Freeing the parameter is in line with the conditions in Taiwan as pointed out before.

Discussion and conclusion

The equivalence of loading patterns and intercepts can be affected by incongruent definitions and meanings of a construct or by different response styles across groups (Cheung and Rensvold 2000; Chen and West 2008). That is particularly to be expected when comparing cultures, due to the fact the response style could differ between cultures, as it applies for instance to East Asians which tend to avoid extreme categories in contrast to Western respondents (Chen et al. 1995). Furthermore, a lack of invariance of loading patterns can be caused by different construal of the self and different beliefs and values, which shape the

motivation of the individual (Markus and Kitayama 1991; Hofstede 1986; Chen and Stevenson 1995). For that reason, the comparability of results revealed by large scale studies cannot be taken for granted but has to be scrutinized. This is particularly important if a motivation scale is constructed, which could be sensitive to country-specific conditions.

The current paper intended to examine whether the TEDS-M items can be used to develop scales of teachers' job motivation in line with the theory about motivation that is invariant across countries. Such scales would be very useful from a policy perspective because they would make it possible to examine predictors and outcomes of teacher motivation as well as to learn from other countries. For our constructed scales of intrinsic and extrinsic motivation an examination of MI revealed as hypothesized that full or partial *scalar* invariance did not exist across all countries. Based on the state of research, we had hypothesized that it would be possible to confirm scalar MI of the intrinsic- and extrinsic-motivation scales across subsets of countries that share societal and educational traditions such as individualism vs. collectivism. The data supported this hypothesis for the group of individualistic countries but not for the group of the collectivistic orientated countries. Only if societal and educational traditions matched each other in subgroups of these countries, comparisons of scale means are permitted.

The good news is that partial metric invariance could be established for most TEDS-M countries which means that it is at least possible to compare relationships across countries—besides the comparisons of means another important objective of international large-scale assessments. The TEDS-M instruments used to construct scales of intrinsic and extrinsic motivation to become a teacher have the same loading patterns across Chile, Germany, Malaysia, Norway, Oman, Philippines, Poland, Singapore, Switzerland, and Thailand. Taiwan can be included in this list if the analyses are restricted to the intrinsic-motivation scale.

With respect to one item of the extrinsic scale, differences in country specific working conditions of the teaching profession turned out to constitute bias. This seems to be another important characteristic to consider in comparative research in addition to culture-specific meaning of a construct, response styles or translation errors when it comes to the comparability of results. Current international large-scale assessments attempt to collect data on a "common core" of all countries participating—in achievement as well as in opportunities to learn or context conditions. The IEA has a systematic approach to ensure this. All items that go onto the instruments are consensually agreed upon by representatives from all participating countries. There are, for example, also curriculum-test matching questionnaires that a representative from each country fills out, indicating whether an item is or is not on the country's curriculum (Hencke et al. 2009). Nevertheless, cross-country comparability of the assessed data must be ensured before comparing results, this is especially required if a culturally sensitive scale is constructed based on the data. Our study does not support cross-country comparability with respect to working conditions of teachers with the result that certain aspects of job motivation are not part of a construct in single countries from an empirical perspective.

Caused by the long-term high attractiveness of the teaching profession in Taiwan and since the number of educational programs has increased substantially during the past decades, there is a remarkable oversupply of teachers. Many qualified teachers cannot move into teaching jobs because the number of positions available is much lower than

the number of graduate teachers (Li et al. 2011). Ignoring the lack of MI for Taiwan, could result to substantial bias in regression slopes. The regression slopes may be overestimated for Taiwan if the extrinsic scale predict a criterion or could be underestimated if the extrinsic scale is modelled as the criterion (Chen 2008, pp. 1010–1011). Therefore, modelling predictive relationships for the tested TEDS-M countries can be done simultaneously, for Taiwan the analysis is to be conducted separately.

Although factor means are not comparable across all TEDS-M countries, invariance of intercepts of the intrinsic scale exists at least within groups of countries, namely for Germany, Norway, Switzerland and Poland, for Taiwan and Singapore as well as for the Philippines, Malaysia and Thailand. The respective countries share not only a societal but also an educational tradition. This seems to be a sufficient precondition for mean comparisons and should be taken into account in future reports of results from large-scale assessments. However, again country-specific working conditions harm full scalar equivalence of the instruments. In Norway, the income of teachers is comparatively low compared to other professions. Also, the earning progression over the life-span is lower than in other OECD countries (OECD 2014). Therefore, in contrast to other professionals in the Norwegian public sector many eligible teachers choose another profession, leave the teaching profession in prospect of better career opportunities or choose early retirement (Carnoy et al. 2009).

With studying MI of the intrinsic and extrinsic motivation scales constructed based on the items assessing the job motivation in TEDS-M, an important first step was done in order to make meaningful cross-cultural and cross-country comparisons of teachers' job motives. However, we have to take into account that the items used to construct motivation scales are limited to three and four items in each factor, which do not represent a continuum of motivation. Therefore, as it is often the case in international large-scale assessments with given items it is not possible to construct a strong motivation scale. The result that the constructed scales do not fit empirically to the data from the USA and Russia and the fact that item estimates have to be allowed to vary for some other countries points to this. Nevertheless, it was worthwhile to construct motivation scales to compare the intrinsic and extrinsic motivation to become a teacher in different countries.

As the results of our study show, it is indispensable to test for cross-country and cross-culture equivalence of scales. For that reason, researchers conducting secondary data analyses should investigate measurement invariance before comparing results across countries and cultures. This is especially required if scales are created by items that were not intentionally designed to measure the construct of interest.

However, the question remains what to do in those cases scalar MI cannot be confirmed although the object of interest is mean comparison. An appropriate way could be to use the alignment method or Bayesian approaches in order to address a lack of MI (Asparouhov and Muthén 2014). Under the working assumption of approximate measurement equivalence, informative priors to define elastic constraints are used in these cases. In contrast to classical exact approaches, Bayesian approaches permit small differences between parameters as loadings or intercepts with the restriction that the mean of differences of loadings or intercepts is zero across groups. Given by the results of simulation studies small variations in parameters do not harm conclusions based

on comparative results (e.g. Muthén and Asparouhov 2013). The recent availability of specific Bayesian software and support in a general software package like Mplus makes Bayesian data analysis techniques accessible to a broad range of educational researchers.

Comparing the motivation to become a teacher in different countries and cultures can help to understand which mechanisms constitute choosing teaching as a career. The present study pointed out which types of analyses are permitted and which are not. Future studies should address predictors of choosing the teaching profession. Such insights could give implications for addressing and recruiting teaching candidates in an adequate way in countries which face the challenge of mathematics teacher shortage.

Authors' contributions

CL and SB contributed to the conception and the design of the paper. CL conducted the analyses and drafted the manuscript. SB made secondary contributions to it. Both authors read and approved the final manuscript.

Author details

¹ Institut für Erziehungswissenschaften, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany.

² Centre for Educational Measurement at the University of Oslo (CEMO), Niels Henrik Abels hus, Moltke Moes vei 35, 0318 Oslo, Norway.

Competing interests

We have read and understood Large-scale Assessments in Education policy on declaration of interests and declare that we have no competing interests.

Received: 4 November 2015 Accepted: 7 September 2016

Published online: 21 September 2016

References

- Artelt, C. (2005). Cross-cultural approaches to measuring motivation. *Educational Assessment*, 10(3), 231–255.
- Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 495–508. doi:10.1080/10705511.2014.919210.
- Banks, J. A. (Ed.). (2012). *Encyclopedia of diversity in education*. London: Sage Publications.
- Blömeke, S. (2014). Vorsicht bei Evaluationen und internationalen Vergleichen: Unterschiedliche Referenzrahmen bedrohen die Validität von Befragungen zur Lehrerbildung. *Zeitschrift für Pädagogik*, 60, 109–131.
- Blömeke, S., Houang, R., & Suhl, U. (2011). TEDS-M: Diagnosing teacher knowledge by applying multidimensional item response theory and multi-group models. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, 4, 109–126.
- Blömeke, S., Kaiser, G., & Lehmann, R. (Eds.). (2010). *TEDS-M 2008—Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich*. Münster: Waxmann Verlag.
- Blömeke, S., Suhl, U., & Döhrmann, M. (2013). Assessing strengths and weaknesses of teacher knowledge in Asia, Eastern Europe and Western countries: Differential item functioning in TEDS-M. *International Journal of Science and Mathematics Education*, 11, 795–817.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Carnoy, M., Betteille, T., Brodziak, I., Loyalka, P., & Luschei, T. (2009). *Teacher education and development study in mathematics (TEDS-M): Do countries paying teachers higher relative salaries have higher student mathematics achievement?*. Amsterdam: IEA.
- Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, 95(5), 1005–1018. doi:10.1037/a0013193.
- Chen, C., Lee, S. Y., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, 6, 170–175.
- Chen, C., & Stevenson, H. W. (1995). Motivation and mathematics achievement: A comparative study of Asian, American, Caucasian American, and East Asian high school students. *Child Development*, 99, 1215–1234.
- Chen, F. F., & West, S. G. (2008). Measuring individualism and collectivism: The importance of considering differential components, reference groups, and measurement invariance. *Journal of Research in Personality*, 42, 259–294. doi:10.1016/j.jrp.2007.05.006.
- Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, 31(2), 187–212.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Publishing Co.
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analyses with ordinal data. *Psychological Methods*, 9, 466–491.
- Hencke, J., Rutkowski, L., Neuschmidt, O., & Gonzalez, E. (2009). Curriculum coverage and scale correlation on TIMSS 2003. In M. von Davier & D. Hastedt (Eds.), *IERI monograph series: Issues and methodologies in large scale assessments* (pp. 85–112). New York: Springer.

- Hofstede, G. (1986). Cultural differences in teaching and learning. *International Journal of Intercultural Relations*, 10, 301–320.
- House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (Eds.). (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Thousand Oaks: Sage.
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Hui, C. H., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*, 20(3), 296–309.
- König, J., Kaiser, G., & Felbrich, A. (2012). Spiegelt sich pädagogisches Wissen in den Kompetenzselbstschätzungen angehender Lehrkräfte? Zum Zusammenhang von Wissen und Überzeugungen am Ende der Lehrerbildung. *Zeitschrift für Pädagogik*, 58(4), 476–491.
- Kwan, V. S. Y., Bond, M. H., Boucher, H., Maslach, C., & Gan, Y. (2002). The construct of individuation: More complex in collectivist than in individualist cultures. *Personality and Social Psychology Bulletin*, 28, 300–310.
- Laschke, C., & Blömeke, S. (2014). *Teacher education and development study: Learning to teach mathematics (TEDS-M). Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- Leung, F. K. S. (2001). In search of an East Asian identity in mathematics education. *Educational Studies in Mathematics*, 47, 35–52.
- Leung, F. K. S., Graf, F., & Lopez-Real, F. (Eds.). (2006). *Mathematics education in different cultural traditions—a comparative study of East Asia and the West*. New ICMI Studies Series No. 9. New York: Springer.
- Li, G., He, M. F., Tsou, W., Hong, W. P., Curdt-Christiansen, X., & Huong, P. L. (2011). Teachers and teaching in sinic education. In Y. Zhao, J. Lei, G. Li, M. F. He, K. Okano, N. Megahed, D. Gamaga, & H. Ramanathan (Eds.), *Handbook of Asian education. A cultural perspective* (pp. 51–77). New York: Routledge.
- Little, R. J., & Rubin, D. B. (2014). *Statistical analysis with missing data*. Hoboken: Wiley.
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98(2), 224–253.
- Muthén, L. K., & Muthén, B. O. (2010). MPLUS user's guide, Sixth Edition. Los Angeles, CA: Muthén & Muthén.
- Muthén, B. O., & Asparouhov, T. (2013). BSEM measurement invariance analysis. Mplus Web Notes: No. 17.
- Muthén, L. K., & Muthén, B. O. (1998–2015). Mplus user's guide. Seventh Edition. Los Angeles: Muthén & Muthén.
- Nagengast, B., & Marsh, H. W. (2014). Motivation and engagement in science around the globe: Testing measurement invariance with multigroup structural equation models across 57 countries using PISA 2006. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment. Background, technical issues, and methods of data analysis* (pp. 317–344). New York: Taylor & Francis.
- OECD. (2014). Norway—Country Note—Education at a glance 2014: OECD Indicators. <http://www.oecd.org/edu/Norway-EAG2014-Country-Note.pdf>.
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354–373. doi:10.1037/a0029315.
- Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, 74(1), 31–57.
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25, 54–67.
- Salili, F. (1995). Explaining Chinese students' motivation and achievement: A sociocultural analysis. *Advances in Motivation and Achievement*, 9, 73–118.
- Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling: A Multidisciplinary Journal*, 21, 167–180. doi:10.1080/10705511.2014.882658.
- Satorra, A., & Bentler, P. M. (2001). A scaled difference Chi square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507–514.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research-Online*, 8, 23–74.
- Schwille, J., & Ingvarson, L. (Eds.) (in collaboration with other series editors M.T. Tatto, M.T., Senk, S. Peck, R., & Rowley, G.) (2013). *The TEDS-M Encyclopedia: A Guide to Teacher Education Context, Structure and Quality Assurance in the Seventeen TEDS-M Countries*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).
- 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(4), 601–630.
- Shin, J., Lee, H., & Kim, Y. (2009). Student and school factors affecting mathematics achievement international comparisons between Korea, Japan and the USA. *School Psychology International*, 30(5), 520–537.
- Tatto, M. T. (2013). *Teacher education and development study in mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics*. Technical Report. Amsterdam: International Association of the Evaluation of Student Achievement.
- Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). *Teacher education and development study in mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics*. Conceptual framework. East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University.
- Tatto, M. T., Schwille, J., Senk, S. L., Ingvarson, L., Rowley, G., Peck, R., et al. (2012). *Policy, practice, and readiness to teach primary and secondary mathematics in 17 countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M)*. Amsterdam: IEA.

- Triandis, H. C. (1995). *Individualism and collectivism*. Boulder: Westview press.
- Van de Vijver, F. J. R., & Leung, K. (2000). Methodological issues in psychological research on culture. *Journal of Cross-Cultural Psychology*, 31(1), 33–51.
- Van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, 54, 119–135. doi:10.1016/j.erap.2003.12.004.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70.
- Vollstedt, M. (2011). Sinnkonstruktion und Mathematiklernen in Deutschland und Hongkong: Eine rekonstruktiv-empirische Studie. *Perspektiven der Mathematikdidaktik*, 2, Wiesbaden, Vieweg + Teubner.
- Watkins, D. A., & Biggs, J. B. (Hrsg.) (1996). *The Chinese learner: Cultural, psychological and contextual influences*. Hong Kong: Comparative Education Research Centre and Victoria, Australia: The Australian Council for the Educational Research.
- Watt, H. M., & Richardson, P. W. (2007). Motivational factors influencing teaching as a career choice: Development and validation of the FIT-Choice scale. *The Journal of Experimental Education*, 75(3), 167–202.
- Watt, H. M. G., Richardson, P. W., & Devos, C. (2013). (How) Does gender matter in the choice of a STEM teaching career and later teaching behaviours? *International Journal of Gender, Science and Technology*, 5(3), 187–206.
- Watt, H. M., Richardson, P. W., Klusmann, U., Kunter, M., Beyer, B., Trautwein, U., et al. (2012). Motivations for choosing teaching as a career: An international comparison using the FIT-Choice scale. *Teaching and Teacher Education*, 28(6), 791–805.
- Watt, H. M. G., Richardson, P. W., & Pietsch, J. (2009). Choosing to teach in the “STEM” disciplines: Characteristics and motivations of science, technology, and mathematics teachers from Australia and the United States. In A. Selkirk & M. Tichenor (Eds.), *Teacher education: Policy, practice and research* (pp. 285–309). New York: Nova Science Publishers.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68–81.
- Yuan, K., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with non-normal missing data. *Sociological Methodology*, 30(1), 165–200.
- Zhao, Y. (2011). *Handbook of Asian education: A cultural perspective*. New York: Routledge.
- Zhu, Y., & Leung, F. K. S. (2011). Motivation and achievement: Is there an East Asian model? *International Journal of Science and Mathematics Education*, 9, 1189–1212.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
