Gorges *et al. Large-scale Assess Educ (2017) 5:13*
DOI 10.1186/s40536-017-0047-5

Large-scale Assessments in Education

RESEARCH

Open Access

Same but different? Measurement invariance of the PIAAC motivation-to-learn scale across key socio-demographic groups

Julia Gorges[1*], Tobias Koch[2], Débora B. Maehler[3] and Judith Offerhaus[4]

*Correspondence: julia.gorges@uni-bielefeld.de
[1] Department of Psychology, Faculty of Psychology and Sports Science, Bielefeld University, PO Box 10 01 31, 33501 Bielefeld, Germany
Full list of author information is available at the end of the article

Abstract

Background: Data from the Programme for the International Assessment of Adult Competencies (PIAAC) revealed that countries systematically differ in their respondents' literacy, numeracy, and problem solving in technology-rich environments skills; skill levels also vary by gender, age, level of education or migration background. Similarly, systematic differences have been documented with respect to adults' participation in education, which can be considered as a means to develop and maintain skills. From a psychological perspective, motivation to learn is considered a key factor associated with both skill development and participation in (further) education. In order to account for motivation when analyzing PIAAC data, four items from the PIAAC background questionnaire were recently compiled into a motivation-to-learn scale. This scale has been found to be invariant (i.e., showing full weak and partial strong measurement invariance) across 21 countries.

Methods: This paper presents further analyses using multiple-group graded response models to scrutinize the validity of the motivation-to-learn scale for group comparisons.

Results: Results indicate at least partial strong measurement invariance across gender, age groups, level of education, and migration background in most countries under study (all CFI > .95, all RMSEA < .08). Thus, the scale is suitable for comparing both means and associations across these groups.

Conclusions: Results are discussed in light of country characteristics, challenges of measurement invariance testing, and potential future research using PIAAC data.

Background

The release of the (first round of) PIAAC data in 2013 has drawn educational researchers' attention to a thus far neglected target group in education: adult learners. With respect to literacy, numeracy, and problem solving in technology-rich environments (ICT) skills, PIAAC data reveals differences across the participating countries (OECD 2013a). Going beyond cross-national comparisons, the general OECD report (OECD 2013a) and country-specific publications (e.g., Maehler et al. 2013; Statistics Canada 2013) provide in-depth analyses of specific population subgroup competencies that are relevant for researchers, educators, and policy-makers alike. These analyses show systematic skill differences across gender, age groups, level of education, and migration background (OECD 2013a). Similar group differences have been found with respect to adults' participation in further education and training (OECD 2005), which is considered a means to develop and maintain skills.

Taking on a psychological perspective, motivation to learn is a key factor for skill development and participation in (further) education (Boeren et al. 2010; Manninen 2005; Gorges 2015). In order to account for motivation when analyzing PIAAC data, four items from the PIAAC background questionnaire have recently been compiled into a motivation-to-learn scale (Gorges et al. 2016). The scale so far has been found to measure motivation to learn in an equivalent way across 21 countries and thus allows comparative analyses.

To enable in-depth analyses of different groups concerning their motivation to learn, this paper addresses the applicability of the scale for group comparisons within each of the 21 countries included in Gorges et al. (2016). In particular, measurement invariance—an important prerequisite for valid comparisons of estimates across groups—has been investigated with respect to four socio-demographic variables. Thus, the goal of this paper is to advise researchers whether they can draw valid inferences when using the motivation-to-learn scale in comparative research on the said groups of people.

Key grouping variables in PIAAC

We will use four grouping variables when analyzing measurement invariance of the motivation-to-learn scale, namely gender, age groups, level of education, and migration background. Each individual can easily be characterized by any combination of these grouping variables' subgroups. As these variables represent basic socio-demographic information, they are generally used as independent variables or standard controls in empirical research in psychology and in almost all social science disciplines. Thus, they represent the starting point of measurement invariance testing. Of course, analyzing other variables, e.g., family status or income would be possible and a promising endeavor for future research projects with a more specific focus.

The following sections further elaborate on differences of our four grouping variables with respect to participation in further education and motivation to learn, and thus underline the relevance and importance of these key socio-demographic variables.

Differences by gender

Based on the large body of literature, gender is one of the most important grouping variables in educational research (for overviews see, for example, Bose and Kim 2009; Chrisler and McCreary 2010; Skelton and Francis 2006). Empirical findings drawing on large-scale assessments of adult skills (Statistics Canada and OECD 2005; OECD 2013a) and meta-analyses (for an outline e.g., Else-Quest et al. 2010) suggest that gender differences are only marginal when controlling for other relevant covariates such as education and employment. Moreover, rates of participation in non-formal education (EACEA P9 Eurydice 2012) and employer-sponsored further education (OECD 2005) are comparable for men and women in most developed countries.

Nevertheless, gender differences regarding mathematical and verbal skills, which are documented for children and adolescents in particular, may be explained by gender-specific socialization (Wigfield and Eccles 2000; Wigfield et al. 2009), identity formation processes (Eccles 2009), different career choices (Watt and Eccles 2010), or differences in motivational factors such as self-efficacy (OECD 2004, 2010). Given that motivation to learn is also strongly affected by individual experiences in different cultural and social environments (Wigfield and Eccles 2000) that have diverging gender roles, gender differences in item understanding may occur. Therefore, measurement invariance across men and women needs to be tested prior to comparing the motivation-to-learn scale or its relations to other variables.

Differences by age group

According to PIAAC, for instance, literacy skills peak between age 25 and 34 and are lowest for adults over 55 years of age across all countries (OECD 2013a). Findings from other studies also suggest that adult skills decline with age (OECD and Statistics Canada 2000; Statistics Canada and OECD 2005). Age-related differences may be attributed to cognitive maturation and decline (Baltes et al. 2006). However, because PIAAC so far is only cross-sectional in most countries, investigating individual skill development over the life course is not possible. The observed age group differences rather reflect cohort differences that result from variations in skill formation regimes and respective changes in national education systems. Higher proportions of younger cohorts have experienced the benefits of educational expansion and thus had access to extensive formal schooling compared to older cohorts (Desjardins 2003; Staudinger et al. 1995). For example, PIAAC data shows that skill differences between younger and older age cohorts are particularly pronounced in Korea, which reflects the substantial expansion of secondary schooling over the last 30 years. Hence, this case illustrates how age-related skill differences are due to the country's history (and thus cohort effects) rather than age-related cognitive decline (OECD 2013a). In addition, using longitudinal data Reder (1994) and Reder and Bynner (2009) show that socialization processes—e.g., cultural and school environments—appear to be related to skills more than biological aging processes. Yet, cohort effects cannot fully be disentangled from age effects.

Empirical findings about participation in further education show a similar pattern to PIAAC results on adult competencies: participation increases in early adulthood, is highest during the mid-life phases, and declines later in life. As most further education is job-related, skill acquisition is tied to the different stages in the individual employment career. Thus, initially adults need to acquire job-related skills and continuously expand these while building their careers. Accordingly, the need for learning is important in established career phases, especially as workers still can recoup benefits from their investments in further education for a considerable amount of time. However, when approaching retirement, workers may be less inclined to invest in skills as return periods decrease (Becker 1962).

Age-related differences in skills and participation in further education could be related to changes in motivation to learn. However, individuals may interpret a measure of motivation to learn differently depending on their age. For example, young adults may associate learning with formal schooling, whereas older adults think of company-based vocational training. Therefore, we need to test the comparability of the motivation-to-learn scale across age groups.

Differences by level of education

As education directly affects skill acquisition and development (Kirsch et al. 2002; OECD and Statistics Canada 2000), individual level of education is strongly associated with skill levels across all countries (OECD 2013a). In addition, level of education also relates to occupational status, income, and participation in further education (Desjardins et al. 2006; OECD 2005). Thus, people with higher levels of education have access to and use more opportunities to maintain and develop their skills. As the socio-economic background of the family strongly impacts individual level of education (Breen and Jonsson 2005; Ishida et al. 1995), individuals with different levels of education have experienced different upbringings that may translate into differences in self-concepts, the value attached to education, and, consequently, motivation to learn. As quantity and quality of motivation as well as individual understanding of motivation to learn may vary depending on their socio-economic and educational background, we need to ensure that the scale shows measurement invariance across levels of education.

Differences by migration background

Previous studies on the relationship between skills and migration background (e.g. PISA, International Adult Literacy Survey [IALS]) show significantly higher skills for native speakers compared to non-native speakers (OECD and Statistics Canada 2000; Stanat et al. 2010). PIAAC countries differ considerably regarding migrants' languages and cultures of origin. For example, while some countries such as Australia and Spain have many immigrants with the countries' official language as their mother tongue, in most countries (e.g., Germany or Sweden) immigrants are non-native speakers of the countries' official language(s). With respect to using the motivation-to-learn scale for comparative research on migrants versus non-migrants, potential divergence between the test language and the native language of the test-taker may bias comparability of item understandings within the participating countries (Hambleton 2005; Maehler et al. 2017).

Motivation to learn in PIAAC

Previous research identifies an invariant measure of adult motivation to learn using items from the PIAAC background questionnaire (Gorges et al. 2016). Motivation generally pertains to "*the process whereby goal-directed activities are instigated and sustained*" (Schunk et al. 2014, p. 5, italics in original). Motivation to learn in educational psychology mainly focuses on children and adolescents, while research on adult motivation to learn is rare despite its importance as a predictor of adult learning (Courtney 1992; Gorges 2015).

The four items used by Gorges et al. (2016) primarily tap enjoyment of learning and goals of knowledge expansion. These aspects are commonly referred to as intrinsic forms of motivation (Ryan and Deci 2000) and mastery goal orientation (Maehr and Zusho 2009). Within educational psychology research shows that intrinsic motivation and mastery goal orientation predict voluntary engagement in learning activities, the use of deep learning strategies, positive affect experienced during learning, and positive learning outcomes (cf. Wigfield et al. 2006). Hence, adults scoring high on the motivation-to-learn scale are assumed to readily engage in and gain as much as possible from learning activities. In particular, we find that motivation to learn is significantly related to participation in further education in most PIAAC countries even after controlling for level of education (Gorges et al. 2016).

Measurement invariance of the motivation-to-learn scale across key socio-demographic groups

Measurement invariance (MI, also called measurement equivalence) means that a theoretical construct is measured in the same—i.e., equivalent—way in two or more groups. As such, MI is a necessary prerequisite for valid comparative research (Chen 2008). MI is typically employed when a measure of several items (e.g., single tasks in a test or agree-disagree statements) is used to represent a latent construct. For example, while age and level of education are directly reported by the PIAAC participants (or can easily be inferred from the provided information), unobservable constructs like motivation to learn are estimated based on participants' responses to four items. When comparing across groups, it is important to ensure that the items used to reflect a latent construct are understood in a similar way across these groups. For instance, when we want to compare how motivation affects participation in further education for men versus women, we can only draw valid conclusions when potential differences in motivation are not attributable to the measurement instrument, that is, when men and women attribute the same meaning to the items and we can assume measurement invariance across gender (Chen 2008; Vandenberg and Lance 2000).

Advanced statistical procedures allow for MI tests of the assumption that the measurement instruments are invariant across groups, and that group differences of, for example, latent means are thus attributable to the grouping variable. Several levels of MI can be established (Meredith 1993). The most basic level, *configural MI*, concerns the factor structure of the measurement instrument. The next level, *weak or metric MI*, refers to the factor loadings of the indicators being equivalent across groups. The third level, *strong or scalar MI*, specifies that the intercepts of the indicators are equivalent across groups. Weak MI is sufficient to compare associations between variables, whereas strong MI is necessary to compare latent means. Thus, testing the MI of the motivation-to-learn scale is a necessary prerequisite for using it in comparative research across gender, age groups, level of education and migration background; if the assumption of MI does not hold, group comparisons may be invalid.

Among our four grouping variables, gender differences received most attention in empirical educational research. Results for MI testing generally support the assumption of at least weak—mostly strong—MI of motivational measures across gender (e.g., Choy et al. 2016; Freund et al. 2011; Gaspard et al. 2015; Grouzet et al. 2006; Kosovich et al. 2015; Litalien et al. 2015; Marsh 1993; Su et al. 2015).

Turning to age-related differences, studies in educational psychology typically address young age groups in the context of primary and/or secondary schooling. Findings based on longitudinal datasets support assumptions of MI across age groups ranging from elementary to upper secondary school students (e.g., grades 7, 8, 9, and 10; Grouzet et al. 2006, Marsh 1993; elementary and middle-school students, Choy et al. 2016; Zhu et al. 2012). Woo et al. (2007) examined latent mean differences in facets of achievement motivation in a sample of students (mean age 20.74, SD = 4.43, 58% female) and adult workers (mean age 42.82, SD = 9.89, 34% female). Their results show MI for these age groups. However, the age groups tested in the literature only cover a very limited age range and, therefore, do not allow for generalization across entire adult populations.

Less researched is MI for motivational measures with respect to level of education and migration background. Because educational psychological research is heavily focused on young learners, samples typically do not differ in educational attainment. Nevertheless, a study by Gorges and Hollmann (2015) using the German sample of the Adult Education Survey (AES) found weak MI for motivation to participate in further education across levels of education. Finally, considering migration background, Segeritz and Pant (2013) report at least weak MI for motivational aspects from the PISA 'Students' Approaches to Learning Instrument' across different ethnic/cultural groups within a country.

In sum, the current literature on MI across these key socio-demographic groups would benefit from further investigating MI by providing a comprehensive picture of potential group differences.

Methods

Data and sample restrictions

We analyzed PIAAC data from the 21 countries that met the analytic prerequisites and provided representative samples (OECD 2013b).[1] In some countries, a very low share of the population (less than 5%) has a native language that differs from the respective official language(s) (Maehler et al. 2014). These countries (Japan, Czech Republic, Estonia, Finland, Poland, and Korea) are excluded from the MI analyses regarding migration background.

In this study, we used a multiple-group graded response model (GRM; Samejima 1969). The GRM belongs to the family of item response models and is equivalent to a confirmatory factor model with categorical observed variables (Takane and de Leeuw 1987). Many statistical software programs (e.g., *Mplus*, Muthén and Muthén 1998–2012) require that the number of response categories is equal in all groups. For this purpose, researchers can either collapse adjacent categories with no or low case numbers or exclude the respective groups from the analysis. In this study, we decided to exclude three countries (i.e., Finland and Norway in the age group analyses; Slovak Republic in the analyses of level of education); in order to include all countries, it would have been necessary to combine adjacent response categories in all countries, which seemed inappropriate. By excluding at maximum three countries from the analyses, it was possible to use the original response categories.

Measures

The motivation-to-learn scale and all relevant socio-demographic information are part of the PIAAC background questionnaire. Descriptive statistics for our sample are displayed in Table 1.

[1] Cyprus, the Russian Federation, and Belgium (Flanders) were excluded. The PIAAC net sample includes literacy-related non-respondents (LRNR), for whom age and gender were collected by the interviewer (see the guidelines for completed cases in PIAAC as defined by an international consortium on standards and guidelines; OECD 2010). However, these respondents comprise less than 5% of the population in the countries considered in our analyses. For further details on the data collection procedure, see the PIAAC Technical report (OECD 2013b).

Table 1 Descriptive statistics for each country using sampling weights

Country	N	Female (%)	Level of education (%)			Age group (%)			Migration background (%)
			High	Intermediate	Low	16–29	30–49	50–65	
Australia	7430	50	33	39	28	30	42	29	17
Austria	5130	50	17	60	23	26	44	31	14
Canada	26,683	50	46	39	15	27	41	33	22
Czech Republic	6102	50	18	67	16	25	43	32	2
Denmark	7328	50	34	40	26	26	42	32	11
Estonia	7632	52	37	45	18	28	41	30	4
Finland	5464	50	36	44	20	26	39	35	4
France	6993	51	27	45	28	27	42	32	9
Germany	5465	50	30	53	17	25	44	31	13
Ireland	5983	51	32	40	28	28	47	26	10
Italy	4621	50	12	34	54	23	47	30	9
Japan	5278	50	42	44	15	23	44	33	–[a]
Korea	6667	50	35	43	22	26	46	28	1
The Netherlands	5169	50	31	38	31	26	42	32	10
Norway	5128	49	35	38	27	27	43	30	13
Poland	9366	51	26	59	15	30	39	31	1
Slovak Republic	5723	50	19	60	21	29	42	29	6
Spain	6055	50	29	23	47	21	48	30	8
Sweden	4469	49	28	48	24	28	41	32	17
United Kingdom	8892	50	36	40	24	29	42	29	11
United States	5010	51	36	50	15	29	41	30	15
OECD total/average	150,588	50	30	45	24	27	43	31	9

[a] Japan did not collect data on the migrant population

In countries with multiple official languages like Canada (English and French) and Spain (Spanish, Catalan, Galician, Valencian and Basque), the background questionnaire was provided in all these languages. Additionally, the background questionnaire was provided in multiple languages in Austria (German, Turkish, and Serbo-Croatian), Finland (Finnish and Swedish), Norway (Norwegian and English), the Slovak Republic (Slovak and Hungarian) and the United States (English and Spanish) to accommodate larger shares of non-native speakers (OECD 2013b). A high quality translation process headed by cApStAn has been implemented to ensure comparability of these questionnaires across the participating countries (OECD 2013b).

The *motivation-to-learn scale* consists of four items: "*I like learning new things*" (I_Q04d), "*I like to get to the bottom of difficult things*" (I_Q04j), "*I like to figure out how different ideas fit together*" (I_Q04l), and "*If I don't understand something, I look for additional information to make it clearer*" (I_Q04m). The internal consistency of the scale ranges between .75 and .89 (for details see Table 2).

Level of education (based on the variable EDCAT6) is measured according to the International Classification of Educational Attainment (ISCED; UNESCO 2011). A *low* level of education reflects completed primary and lower secondary education (ISCED 1 and

Table 2 Overview of the internal consistency of the scale and the highest level of measurement invariance per country and grouping variable

Country	α	Gender	Age groups	Level of education	Migration background
Australia	.837	Strong	Strong	Partial strong	Strong
Austria	.811	Strong	Partial strong	Strong	Strong
Canada	.812	Strong	Strong	Strong	Strong
Czech Republic	.755	Partial strong	Weak	Weak	n/a
Denmark	.785	Strong	Partial strong	Partial strong	Partial strong
Estonia	.825	Strong	Strong	Strong	n/a
Finland	.745	Strong	n/a	Weak	n/a
France	.780	Strong	Strong	Strong	Strong
Germany	.790	Strong	Partial strong	Strong	Strong
Ireland	.820	Strong	Strong	Partial strong	Strong
Italy	.836	Strong	Strong	Strong	Strong
Japan	.804	Strong	Strong	Strong	n/a
Korea	.844	Strong	Partial strong	Partial strong	n/a
The Netherlands	.824	Strong	Strong	Partial Strong	Strong
Norway	.761	Strong	n/a	Partial Strong	Weak
Poland	.840	Strong	Strong	Strong	n/a
Slovak Republic	.886	Strong	Strong	n/a	Strong
Spain	.773	Strong	Strong	Strong	Strong
Sweden	.784	Strong	Partial strong	Strong	Weak
United Kingdom	.837	Strong	Partial strong	Partial strong	Strong
United States	.822	Strong	Strong	Strong	Strong

n/a not available because case numbers in at least one response category for at least one group were too small to fit the model

2), an *intermediate* level completed upper secondary education (ISCED 3 and 4), and a *high* level completed tertiary education (ISCED 5 and 6).

Age is available as 5- or 10-year bands. For our analyses (based on AGEG5LFS), we grouped respondents roughly based on key phases of individual employment trajectories (Heinz 2003) into early working age (16–29), career-building, mid-life working age (30–49), and approaching retirement, later working age (50–65); these phases also correlate differently with participation in further education (O'Connell 1999).

We operationalized *migration background* by whether the test language corresponds to the respondent's native language (based on NATIVELANG).

Analyses

As a recent paper of Gorges and coauthors (2016) already elaborated on the statistical details of testing MI with categorical data, this section provides only a general summary of our analytic strategy. We used multiple-group graded response models (Muthén and Asparouhov 2002; Samejima 1969) to test MI of the four-item motivation-to-learn scale. We tested configural MI by imposing the same factor structure across groups. We tested weak or metric MI by restraining factor loadings to be equal across groups. Finally, we tested strong or scalar MI by additionally constraining thresholds to be equal across groups. In addition, we tested for partial MI in cases where full MI could not be established. Partial strong MI requires that the factor loadings and the thresholds of at least two items remain invariant across all groups (Byrne et al. 1989; Steenkamp and

Baumgartner 1998). In the present study, we freed parameters that showed modification indices above 100 when testing partial MI. Although the fixed cut-off value of 100 will lead to stricter decisions on parameters in larger samples, it has worked sufficiently well according to preliminary analyses, where we have compared results from models with different MI restrictions. In order to correctly specify a multiple-group graded response model, it is essential for researchers to set the error variances equal in all groups; in our case, we fixed them at 1. Hence, we did not explicitly test for strict measurement invariance (i.e., equality of measurement error variances across groups) as these parameters had to be fixed beforehand.

All models were fitted to the data using the weighted least square mean-and-variance adjusted (WLSMV) estimation implemented in *Mplus* 7.31 (Muthén and Muthén 1998–2012). To evaluate model fit, we used the root mean square error of approximation (RMSEA; Steiger 1990) and the comparative fit index (CFI; Bentler 1990). Following Schermelleh-Engel, Moosbrugger, and Müller's (2003) account of cutoff criteria, the RMSEA should be below .06 to indicate good model fit, while values up to .10 are still acceptable. The CFI should be >.95 to indicate good fit and >.90 for acceptable fit (Hu and Bentler 1999).

To assess whether imposing MI led to a significant decline in model fit, we compared each restricted model to the respective less restricted model (i.e., weak MI to configural MI, strong MI to weak MI). Although Rutkowski and Svetina (2014) proposed more liberal cutoff values to evaluate change in model fit for large-scale data analyses, these differences are mainly justified with the larger number of groups (i.e., countries). Because we conceptualize our analyses as within-country analyses comprising only two or three groups, we used the general guidelines suggested by Cheung and Rensvold (2002) and Chen (2007) according to which a decrease in model fit is insignificant if the RMSEA drops by less than .015 and if the CFI drops by less than .01.

Results

We tested the three levels of MI across gender, age groups, level of education, and migration background within each of the 21 countries provided that the information from the datasets fulfilled the prerequisites (see "Data and sample restrictions"). We summarize results for each grouping variable in the following sections (see Table 2); in the Appendix, we provide more details concerning the specified multiple-group graded response models (see Tables 3, 4, 5, 6). In addition, the appendix contains a detailed overview of the parameters that have been freed to test partial MI (see Table 7). In sum, the thresholds between answering option 3 and 4 (on a 5-point Likert-type scale) of Item I_Q04d ('I like to learn new things') have been most often released when testing MI across age groups and educational levels. Moreover, with respect to level of education, the item I_Q04j ('I like to get to the bottom of difficult things') has frequently been affected by parameter releases.

MI across gender

The models tested with respect to gender ranged from configural MI with 4 degrees of freedom (*df*) over weak MI with 7 *df* to strong MI with 22 *df*. As expected, the χ^2 tests were significant ($p < .01$) for all models. However, all models met the criteria for good model fit as indicated by a CFI > .97. Most RMSEA coefficients were acceptable (<.10), whereas the RMSEA for Ireland, Italy, Japan, Slovak Republic, and Spain slightly exceeded the cutoff value. Hence, the configural MI models generally showed acceptable model fit.

With respect to tests of weak and strong MI, we did not find significantly worse model fit for any of the countries. Changes in CFI mostly ranged between ΔCFI = .001 and ΔCFI = .005 with the exception of Czech Republic (ΔCFI = .021 for strong MI), which nevertheless showed a good overall model fit. Although the χ^2 increased when imposing MI restrictions, the RMSEA improved in all countries, probably due to the simultaneous increase in *df*. Thus, the assumption of strong MI across gender holds for all countries except for the Czech Republic, which showed partial strong MI. All strong MI models showed good model fit.

MI across age groups

The models tested with respect to age groups ranged from configural MI with 6 *df* over weak MI with 12 *df* to strong MI with 42 *df*. Paralleling the results described above, the χ^2 tests were significant ($p < .01$) for all models, but all models showed a CFI > .97 and a RMSEA < .10 with the exception of Ireland, Japan, and Spain, for which the RMSEA slightly exceeded this value. Thus, most configural MI models fitted the data reasonably well.

Inspecting potential worsening of model fit due to weak MI restrictions revealed that changes in CFI were less than .003 and the RMSEA improved in all countries except in Poland, where it did not change. Similarly, model fit did not worsen in most countries when imposing strong MI restrictions (ΔCFI < .015; RMSEA reduced, unchanged, or increased by less than .015) However, Austria, Czech Republic, Denmark, Germany, Sweden, Korea, and the UK failed to meet the cutoff criteria indicating substantial model change.

For Sweden, the CFI declined by .28 and the RMSEA increased by .045. Here, results support partial strong MI. For the remaining countries that exceeded the cutoff criteria, at least one of the fit indices indicated substantial changes in model fit (Austria: ΔCFI = .018; Czech Republic: ΔCFI = .022; Denmark: ΔCFI = .011; Germany: ΔCFI = .015; Korea: ΔRMSEA = .019; UK: ΔCFI = .012). Hence, we tested partial MI for these countries. The partial strong MI models showed markedly better model fit for all countries listed above except the Czech Republic. Thus, we decided to assume partial strong MI in regard to age-groups for Austria, Denmark, Germany, Korea, and the UK, and weak MI for Czech Republic.

Overall, we concluded that, despite of some countries failing to meet the cutoff criteria for strong MI, the assumption of strong MI across age groups holds for most countries included in the analyses. In countries with at least partial strong MI, these models showed good model fit.

MI across level of education

The models tested with respect to level of education ranged from configural MI with 6 *df* over weak MI with 28 *df* to strong MI with 42 *df*. Again, most models showed significant χ^2 tests ($p < .01$) with the exception of the configural ($p < .05$) and weak MI ($p = .16$) model for the Czech Republic. The CFI for all models was > .97 and the RMSEA < .10

(except Germany, Ireland, Japan, and Spain, for which the RMSEA slightly exceeded .10). Thus, model fit of configural MI models was acceptable for most countries.

With respect to MI restrictions, level of education turned out to perform similar to age groups. More specifically, the CFI change (.001 < ΔCFI < .023) was within the range deemed acceptable and the RMSEA, again, improved when imposing restrictions of weak and strong MI in two-thirds of the countries; the other seven countries are described in more detail.

When adding strong MI restrictions, the CFI dropped (.011 < ΔCFI < .035) and the RMSEA increased (.005 < ΔRMSEA < .025) substantially for Australia, Denmark, Ireland, Korea, the Netherlands, Norway, and the UK, so that we should not—strictly speaking—assume strong MI for these countries. However, as the model fit under partial strong MI conditions was markedly better—not significantly different from weak MI restrictions—these countries may still be treated as meeting strong MI assumptions. For the Czech Republic and Finland, model fit for both strong and partial strong MI was significantly worse than for the weak MI model (ΔCFI > .012; ΔRMSEA > .017). Therefore, we assumed at least partial strong MI across level of education within all countries except for the Czech Republic and Finland, which only met the conditions for weak MI. For all countries with (partial) strong MI, these models showed good model fit.

MI across migration background

The models tested with respect to language as an indicator of migration background ranged from configural MI with 4 *df* over weak MI with 7 *df* to strong MI with 22 *df*. All models showed significant χ^2 tests ($p < .01$) but their CFI was >.97 and the RMSEA < .10 in most countries (except Ireland, Slovak Republic, Spain, and the United States, for which the RMSEA slightly exceeded .10). Again, most configural MI models fitted the data reasonably well.

With respect to model comparisons, language turned out to perform similar to age groups and level of education. More specifically, the CFI change (.001 < ΔCFI < .008) was within the acceptable range and the RMSEA, again, improved when imposing restriction of weak and strong MI in most countries except Denmark, Norway and Sweden. For Denmark, the CFI dropped (.001 < ΔCFI < .025) and the RMSEA increased (.023 < ΔRMSEA < .031) substantially when imposing strong MI restrictions, so that we should not assume strong MI. However, as the model fit under partial strong MI conditions was markedly better, and not significantly different from weak MI restrictions, Denmark can still be treated as meeting strong MI assumptions. For Norway and Sweden, model fit for both strong and partial strong MI were significantly worse than the weak MI model (ΔCFI > .013; ΔRMSEA > .017). Therefore, we assume at least partial strong MI across migration background for all countries tested except for Norway and Sweden, which only met the conditions of weak MI. For countries showing (partial) strong MI, the respective models fitted well.

Discussion

This paper investigated measurement invariance of the recently proposed motivation-to-learn scale (Gorges et al. 2016) from the PIAAC background questionnaire across key socio-demographic variables—gender, age groups, level of education, and migration background—within 21 countries. In case of weak invariance (i.e., invariant factor loadings), this scale could be used to compare relations between motivation to learn and other variables, for instance basic skills or participation in further education across groups. In case of strong invariance (i.e., invariant intercepts or thresholds), the scale could be used to compare latent means across groups. In addition, as our analyses built on a multiple-group graded response model, residuals were fixed, thereby allowing comparisons of manifest scale scores under the condition of strong MI. Results supported the assumption of weak and at least partial strong MI across all grouping variables and countries included in the analyses except for the Czech Republic for age groups and level of education, Finland for level of education, and Norway and Sweden for migration background. Hence, taking these results together with the findings from Gorges et al. (2016), the proposed motivation-to-learn scale is remarkably robust and can be used for a broad range of comparative research.

Measurement invariance across socio-demographic groups in PIAAC

Based on our results, we conclude that the motivation-to-learn scale generally shows equivalent psychometric properties across the groups of interest. Because partial strong MI results may be treated as supporting strong MI assumptions, our discussion will focus on the countries and groups that did not show at least partial strong MI.

With respect to gender, we found (partial) strong MI in all of the countries for which we could test these group differences. Hence, researchers may investigate whether men are more motivated to learn or whether motivation-to-learn is more strongly related to participation in education for men compared to women, for example.

With respect to our three age groups, only the Czech Republic failed to fulfill the requirement for at least partial strong MI. Hence, for all other countries the motivation-to-learn scale may also be used to investigate whether individuals from the different age groups are more or less motivated to learn. As Santiago, Gilmore, Nusche and Sammons (2012) pointed out with respect to the evaluation of the Czech education system, we know little about students' motivation to learn in this country. Our review of the literature for the adult population yields no additional details to explaining these results. Hence, further investigations of a potentially age-dependent interpretation of the motivation-to-learn scale in the Czech Republic are needed.

The Czech Republic and Finland are the only countries that fail to show at least partial strong MI across levels of education. Hence, the motivation-to-learn scale may be fully used for comparative research in all but these two countries, where it may be used for analyzing the relationship of motivation to learn to other variables. For example, researchers may test whether motivation to learn is differentially related to participation in (further) education, as has been the case in Gorges and Hollmann's (2015) study for Germany. With respect to the Czech Republic, the lack of strong MI can be related to the measurement of the education variable itself. As Schneider (2009) and Strakova (2008) show in their analyses, the aggregation of national categories to harmonized ones in ISCED-97 led to large losses of explanatory power in the Czech Republic (particularly the aggregation of ISCED 3A and 3C). Moreover, since 2006 the country invested in different projects related to gender sensitive education (e.g. Babanová and Miškolci 2007; EACEA P9 Eurydice 2010); this may also have impacted the perception of the motivation-to-learn scale and is a possible explanation for the lack of (partial) strong MI across levels of education. With respect to Finland, the weak MI could also be due to the aggregation of the national categories in ISCED-97 or to the national education reform in the 1970s (Kilpi 2008). However, these findings call for further investigation by country experts.

Finally, our results preclude comparing scale means for only two of the countries included in the analyses with respect to migration background. More specifically, migration background measured by test language shows weak MI in Norway and Sweden. Little is known from the PIAAC documentation about who (in terms of country of origin) took the test in a language other than native language in these countries. Respondents using a language other than the test language are probably heterogeneous and further research would need to look into this matter in more detail to explain the lack of MI in these two countries, especially because other Nordic countries showed no such results. Overall, comparing scale means may lead to invalid results in these countries, whereas in all other countries the scale is fit to be used in comparative research including comparisons of scale means across migration background.

We would like to emphasize that these results need to be interpreted in light of the motivation-to-learn scale implemented in the PIAAC background questionnaire. Due to the three items referring to deep approaches to learning (I_Q04j: 'I like to get to the bottom of difficult things', I_Q04l: 'I like to figure out how different ideas fit together', and I_Q04m: 'If I don't understand something, I look for additional information to make it clearer'), the scale may convey a specific interpretation of the term 'learning'. In particular, these items refer to active engagement in learning as opposed to potentially passively receiving knowledge, for example, by listening to a lecture. We believe that narrowing learning down to specific instances of knowledge or skill acquisition promotes equivalent interpretations of items across groups. With respect to established measures of motivation to learn, reference to a specific learning context (e.g., at school, in mathematics class; e.g., Gaspard et al. 2015) reflects a similar practice. Leaving more leeway for respondents to read their personal associations with learning and education into these respective terms may lead to less consistent item interpretation and, thus, may threaten measurement invariance. Nevertheless, investigations of adult motivation to learn would benefit from different and possibly broader conceptions of learning to account for the broad variety of ways in which adults learn (Merriam et al. 2012).

Methodological challenges of testing measurement invariance

From a methodological viewpoint, MI testing based on categorical response data brings some challenges. In our analyses, some models' initial RMSEA values slightly exceeded the conventional cutoff criterion of .10, but improved after further restriction. Improvements in the RMSEA values—as well as the initial exceeding of the cutoff value—may be partly explained by the fact that the RMSEA is based on the fit function, the degrees of freedom and the sample size, whereas the CFI merely compares the fit of the specified model with regard to a baseline (or independence) model. If the relation of misfit to degrees of freedom improves with additional parameter restrictions, the RMSEA value may drop. Conversely, in a model with few degrees of freedom even little misfit may lead to an increased RMSEA. Hence, the documented improvements of the RMSEA values indicate that the restrictions imposed when testing weak and strong MI only led to a marginal decrease in model fit when accounting for the changes in number of parameters to be estimated. Accordingly, the improvement of RMSEA is yet another indicator that the assumption of weak and (partial) strong MI holds for most models tested here. In this study, we considered all models with (partial) strong MI restrictions for countries in which this MI assumption holds to show a good model fit as indicated by the RMSEA and CFI.

Furthermore, we used multiple-group graded response models to evaluate the degree of measurement invariance across gender, age groups, level of education, and migration background using PIAAC data. In contrast to multiple-group confirmatory factor analyses models for continuous variables, graded response models are in line with item response theory (Samejima 1969; Takane and de Leeuw 1987). That means, these models employed here are more flexible and more appropriate in case of categorical (ordinal) response variables, as they do not assume a linear relationship between the observed and latent variables and are not based on the assumption of multivariate normal data. However, the estimation of complex item response models is more cumbersome than in confirmatory factor analyses models with continuous variables and often requires larger sample sizes. Simulation studies have repeatedly shown that estimation methods for categorical variables (e.g., WLSMV) outperform methods for continuous variables (e.g., ML) if there are less than five response categories and/or if the data is not normally distributed (Beauducel and Herzberg 2006; Rhemtulla et al. 2012).

In order to fit multiple-group graded response models in *Mplus* and test the degree of MI, the same number of categories must be present in all groups. If this requirement is violated for one of the groups, the model cannot be estimated and *Mplus* produces a warning message. To fit a model nonetheless, users could collapse adjacent response categories and pool the response frequencies. However, as the graded response model is not invariant across differential indicators per latent variable (i.e., items used to reflect a latent construct), this would result in fitting different models in some countries.

Limitations

Parts of the analyses were limited by the composition of some country samples. In particular, MI could not be tested in all 21 countries across all groups due to lack of responses in some combinations of country and (socio-economic) groups. Furthermore, this paper could not include the very recent release of data from the eight PIAAC countries participating in the second round surveyed in 2014 to 2015 (Chile, Greece, Indonesia, Israel, Lithuania, New Zealand, Singapore, Slovenia, and Turkey).

In the present paper, we used gender, three age groups, three levels of education, and migration background to indicate key socio-demographic characteristics and tested MI regarding these grouping variables within countries. Hence, our study did not aim at comparisons across different grouping variables and/or across countries and may not generalize to such purposes.

For most countries in PIAAC, age is not available as a continuous variable. Our approach of assigning participants to just three age groups further reduced information—and thus variance. Different approaches such as moderated factor analysis (Bauer and Hussong 2009; Curran et al. 2014) could have been used to test for MI across continuous grouping (or moderator) variables if a continuous age variable would have been available. Future studies should test MI using different age groups and bands, and—if possible—age as a continuous variable to further scrutinize the results presented here. With respect to future publications of PIAAC data, including age as a continuous variable would be most valuable for addressing age-related research questions, for example, whether motivation to learn is associated with retirement status rather than with age.

Furthermore, participants were identified as migrants if they took their skills assessment in their non-native language (i.e., the grouping variable was NATIVELANG). The skills assessment had been available in the respective countries' official language(s). As previously mentioned, some few countries provided the background questionnaire in additional languages, e.g. Turkish in Austria or Spanish in the US (Maehler et al. 2014). Participants in these countries may have been classified as migrants although they have responded to the motivation-to-learn scale in their native language. Hence, replication studies using different indicators of migration background are desirable.

With respect to model fit, it should be noted that some of the configural MI levels slightly exceeded the cutoff value of .10 indicating acceptable model fit (Schermelleh-Engel et al. 2003). However, models with higher degrees of measurement invariance consistently show RMSEA values below .10. These findings may be partly explained by the fact that the RMSEA considers both the fit of model (i.e., fit function) and the degrees of freedom, and is known to favor more parsimonious models (Schermelleh-Engel et al. 2003). Because models with configural MI have a small number of degrees of freedom relative to their χ^2 value, they have a higher chance to be rejected by the RMSEA. In this study, the RMSEA steadily decreases when comparing models with configural and weak MI (up to −.048, for Japan). This indicates that the misfit of the less restrictive models (i.e., configural MI) is most likely due to few degrees of freedom (see the simulation study by Kenny et al. 2014, which also points to the RMSEA being problematic in small degrees of freedom models). We also note that the models with strong or partial strong MI restrictions—which are of substantive interest in this study—fit the data considerably well. Future simulation studies on the behavior of the RMSEA in small degree of freedom models using multiple-groups in particular may shed further light onto the interpretation in such contexts.

Finally, tests of partial MI were based on modification indices above 100. This has two shortcomings; first, the use of modification indices is essentially a data-driven approach, which calls for a cross-validation study. Second, the chosen cutoff value of 100 is—although based on previous analysis—somewhat arbitrary, which led to stricter decisions on parameters' equalities in larger samples. However, given that only very few parameters had to be freed to achieve partial MI, our results show that the motivation-to-learn scale is highly invariant across most groups within most countries. From a methodological point of view, it would be interesting to see whether recently suggested techniques for testing MI would yield similar results. For example, the alignment method by Asparouhov and Muthén (2014) is less restrictive than the traditional confirmatory approaches applied in the present study, and allows researchers to test the degree of approximate invariance. Here, we used a rather conservative approach for testing measurement invariance that is more likely to refute the assumption of MI.

Outlook and suggestions for future research

Our approach to test MI has been particularly conservative. Using different approaches may have led to more countries meeting criteria for (partial) strong MI. Therefore, we encourage researchers who would like to use the motivation-to-learn scale in their research but are unsure about its comparability or need information on MI for different groups to replicate and extend our MI analyses, using potentially more liberal approaches (e.g., approximate MI; Asparouhov and Muthén 2014). In addition, future research should attend to items involved in testing partial MI and use qualitative approaches such as cognitive interviews to reveal in what way item interpretations differ across groups (Collins 2003).

As mentioned before, testing MI allows for comparative research in two regards. First, when the assumption of weak MI has been met, motivation to learn may be included in regression or path analyses. Of particular interest could be whether motivation to learn differentially predicts participation in further education (Gorges and Hollmann 2015). Beyond that, motivation to learn may be conceptualized as a predictor of how much time individuals spend on potentially skill developing tasks at work and at home, and these analyses may be conducted with gender, levels of education, or age groups as moderators. With respect to research on skill mismatch, individuals' motivation to learn may be able to explain why a subgroup is particularly prone to be overqualified (e.g., Levels et al. 2013). Second, drawing on (partial) strong MI assumptions, researchers may be interested in whether contextual factors (e.g., a type of educational system) or personal factors (e.g., gender) are associated with higher or lower levels of motivation to learn. In addition, researchers might be interested in comparing motivation to learn across levels of education to identify potentials for participation in further education for less-educated individuals.

The implications of motivation in learning processes have been well documented (for overviews see Schunk et al. 2014; Wentzel and Wigfield 2009; Wigfield et al. 2006); however, research so far has been less focused on the adult population and hardly addressed cross-national comparisons (Gorges et al. 2016). Using the motivation-to-learn scale to gain insight into the role of motivation for adult learning and skill development thus enables pioneering research with PIAAC data. Yet, the items in this scale were developed in the literature on approaches to learning, therefore they do not represent a coherent theoretical concept of motivation as in motivational psychology (for a detailed discussion of the motivation-to-learn scale see Gorges et al. 2016). Future research should continue developing measures to assess different qualities of adult motivation to learn that are in line with established motivational theories in educational psychology.

Implementing a measure of adult motivation to learn in large-scale, cross-national and -cultural assessments remains a major challenge in future research. In order to work towards this goal, theoretical conceptualizations of everyday learning opportunities need to be taken into account. Noticing these differences and in response providing better-suited items is an ongoing task for developers of measurement instruments and surveys alike. Overall, however, the motivation-to-learn scale is among the few measurement instruments that has been systematically and rigorously tested with respect to MI across various socio-demographic groups. Given the potential of the PIAAC data for analyses from multiple disciplines, recommendations regarding the use of the motivation-to-learn scale in group comparisons will facilitate work on key questions of psychological, educational, and sociological researchers. Hence, this paper provides a promising starting point to ground further research.

Authors' contributions

JG had the lead for this manuscript and is expert in motivational research. JG wrote the theoretical background, the results, and the theoretical aspects of the discussion. DM is an expert for PIAAC data and wrote the method section except the statistical analyses written by TK. DM and TK conducted the analyses. JO added the sociological perspective on the group variables and country-specific parts of the discussion. All authors read and approved the final manuscript.

Author details

[1] Department of Psychology, Faculty of Psychology and Sports Science, Bielefeld University, PO Box 10 01 31, 33501 Bielefeld, Germany. [2] Leuphana University of Lüneburg, Scharnhorststr.1, 21335 Lüneburg, Germany. [3] Department of Survey Design and Methodology, GESIS Leibniz Institute for the Social Sciences, PO Box 12 21 55, 68072 Mannheim, Germany. [4] Institute of Sociology and Social Psychology, University of Cologne, Greinstr. 2, 50939 Cologne, Germany.

Acknowledgements

Work on this paper was supported by the College for Interdisciplinary Educational Research (CIDER), a Joint Initiative of the BMBF, the Jacobs Foundation and the Leibniz Association. The data used were provided by GESIS-Leibniz Institute for the Social Sciences.

Competing interests

The authors declare that they have no competing interests.

Appendix

See Tables 3, 4, 5, 6, 7.

Table 3 Detailed model fit with gender as grouping variable

Model	χ^2	*df*	*p*	CFI	ΔCFI	RMSEA [CI]	ΔRMSEA
Australia							
Configural	81.374	4	<.001	.996		.077 [.063–.092]	
Weak	84.874	7	<.001	.996	.000	.059 [.048–.070]	−.018
Strong	124.831	22	<.001	.995	−.001	.038 [.032–.045]	−.021
Austria							
Configural	84.728	4	<.001	.994		.099 [.081–.118]	
Weak	63.823	7	<.001	.996	+.002	.063 [.049–.077]	−.036
Strong	97.422	22	<.001	.994	−.002	.041 [.033–.049]	−.022
Canada							
Configural	153.708	4	<.001	.995		.058 [.051–.066]	
Weak	128.028	7	<.001	.996	+.001	.040 [.034–.046]	−.018
Strong	139.3	22	<.001	.996	.000	.022 [.019–.026]	−.018
Czech Republic							
Configural	13.218	4	<.001	.998		.032 [.014–.051]	
Weak	39.506	7	<.001	.993	−.005	.045 [.032–.059]	+.013
Strong	149.141	22	<.001	.972	*−.021*	.050 [.043–.058]	+.005
Partial strong	89.442	20	<.001	.985	−.008	.039 [.031–.047]	−.006
Denmark							
Configural	33.317	4	<.001	.998		.049 [.034–.064]	
Weak	39.003	7	<.001	.998	.000	.038 [.027–.051]	−.011
Strong	68.101	22	<.001	.997	−.001	.026 [.019–.033]	−.012
Estonia							
Configural	121.68	4	<.001	.996		.097 [.083–.112]	
Weak	130.72	7	<.001	.996	.000	.075 [.064–.087]	−.022
Strong	276.459	22	<.001	.991	−.005	.061 [.055–.067]	−.014
Partial strong	173.830	20	<.001	.995	−.001	.051 [.044–.058]	−.024
Finland							
Configural	102.075	4	<.001	.988		.104 [.087–.122]	
Weak	77.632	7	<.001	.991	+.003	.067 [.054–.080]	−.037
Strong	92.742	22	<.001	.991	.000	.038 [.030 to .046]	−.029
France							
Configural	98.941	4	<.001	.994		.090 [.076–.106]	
Weak	74.73	7	<.001	.996	+.002	.058 [.046–.070]	−.032
Strong	87.035	22	<.001	.996	.000	.032 [.025–.039]	−.026
Germany							
Configural	99.725	4	<.001	.992		.105 [.088–.124]	
Weak	99.61	7	<.001	.992	.000	.078 [.065–.092]	−.027
Strong	100.641	22	<.001	.993	+.001	.041 [.033–.049]	−.037
Ireland							
Configural	152.789	4	<.001	.992		.120 [.104–.136]	
Weak	120.73	7	<.001	.994	+.002	.079 [.067–.092]	−.041
Strong	158.997	22	<.001	.993	−.001	.049 [.042–.056]	−.030
Italy							
Configural	99.324	4	<.001	.994		.108 [.090–.127]	
Weak	72.843	7	<.001	.996	+.002	.068 [.054–.083]	−.040
Strong	94.207	22	<.001	.996	.000	.040 [.032–.049]	−.028
Japan							
Configural	153.559	4	<.001	.989		.130 [.113–.148]	
Weak	145.401	7	<.001	.990	+.001	.095 [.082–.108]	−.035

Table 3 continued

Model	χ^2	*df*	*p*	CFI	ΔCFI	RMSEA [CI]	ΔRMSEA
Strong	207.013	22	<.001	.987	−.003	.062 [.054–.070]	−.033
Partial strong	161.373	18	<.001	.990	.000	.060 [.052–.069]	−.035
Korea							
Configural	65.833	4	<.001	.998		.074 [.059–.091]	
Weak	55.934	7	<.001	.998	.000	.050 [.038–.063]	−.024
Strong	137.537	22	<.001	.996	−.002	.043 [.037–.050]	−.007
The Netherlands							
Configural	60.788	4	<.001	.996		.082 [.065–.101]	
Weak	48.642	7	<.001	.997	+.001	.053 [.040–.068]	−.029
Strong	71.33	22	<.001	.997	.000	.033 [.024–.041]	−.020
Norway							
Configural	34.036	4	<.001	.996		.061 [.043–.081]	
Weak	53.634	7	<.001	.995	−.001	.058 [.044–.073]	−.003
Strong	56.746	22	<.001	.996	+.001	.028 [.019–.037]	−.030
Poland							
Configural	22.623	4	<.001	.999		.044 [.027–.062]	
Weak	30.195	7	<.001	.999	.000	.037 [.024–.051]	−.007
Strong	68.606	22	<.001	.998	−.001	.029 [.022–.037]	−.008
Slovak Republic							
Configural	114.657	4	<.001	.998		.110 [.093–.128]	
Weak	96.867	7	<.001	.998	.000	.075 [.062–.089]	−.035
Strong	100.765	22	<.001	.998	.000	.040 [.032–.048]	−.035
Spain							
Configural	152.478	4	<.001	.987		.122 [.106–.139]	
Weak	137.571	7	<.001	.988	+.001	.087 [.074–.099]	−.035
Strong	155.157	22	<.001	.988	.000	.049 [.042–.057]	−.038
Sweden							
Configural	39.923	4	<.001	.996		.070 [.052–.091]	
Weak	28.547	7	<.001	.998	+.002	.041 [.026–.058]	−.029
Strong	61.956	22	<.001	.996	−.002	.032 [.023–.041]	−.009
United Kingdom							
Configural	117.83	4	<.001	.995		.087 [.074–.100]	
Weak	102.074	7	<.001	.996	+.001	.060 [.050–.070]	−.027
Strong	116.039	22	<.001	.996	.000	.034 [.028–.040]	−.026
United States							
Configural	85.259	4	<.001	.994		.100 [.082–.119]	
Weak	68.033	7	<.001	.996	+.002	.065 [.052–.080]	−.035
Strong	97.729	22	<.001	.994	−.002	.041 [.033–.050]	−.024

df, degrees of freedom; CFI, comparative fit index; RMSEA, root mean square error of approximation; CI, confidence interval; following Cheung and Rensvold (2002) and Chen (2007), model fit of the more restrictive model should be considered to be significantly worse if the CFI drops by more than .01 and the RMSEA increases by more than .015; changes in CFI and RMSEA that exceed these cutoff values are printed in italics; partial MI has only been tested if assumptions of full MI did not hold

Table 4 Detailed model fit with age groups as grouping variable

Model	χ^2	*df*	*p*	CFI	ΔCFI	RMSEA [CI]	ΔRMSEA
Australia							
Configural	70.339	6	<.001	.997		.071 [.056–.086]	
Weak	53.053	12	<.001	.998	+.001	.040 [.029–.051]	−.031
Strong	229.439	42	<.001	.991	−.007	.046 [.040–.051]	+.006
Austria							
Configural	69.963	6	<.001	.995		.088 [.070–.107]	
Weak	63.326	12	<.001	.996	+.001	.056 [.043–.070]	−.032
Strong	333.863	42	<.001	.978	−*.018*	.071 [.064–.078]	+.015
Partial strong	165.872	38	<.001	.990	−.006	.049 [.042–.057]	−.007
Canada							
Configural	138.55	6	<.001	.996		.055 [.047–.063]	
Weak	143.366	12	<.001	.996	.000	.039 [.033–.044]	−.016
Strong	463.435	42	<.001	.987	−.009	.037 [.034–.040]	−.002
Czech Republic							
Configural	10.956	6	<.001	.999		.023 [.000–.045]	
Weak	29.365	12	<.001	.996	−.003	.031 [.017–.045]	+.008
Strong	155.095	42	<.001	.974	−*.022*	.042 [.035–.049]	+.011
Partial strong	108.095	38	<.001	.984	−*.012*	.035 [.027–.042]	+.004
Denmark							
Configural	35.701	6	<.001	.998		.049 [.034–.065]	
Weak	50.137	12	<.001	.998	.000	.039 [.028–.051]	−.010
Strong	250.675	42	<.001	.987	−*.011*	.049 [.043–.055]	+.010
Partial strong	124.509	38	<.001	.995	−.003	.033 [.027–.040]	−.006
Estonia							
Configural	116.955	6	<.001	.996		.094 [.800–.110]	
Weak	118.751	12	<.001	.996	.000	.065 [.055–.076]	−.029
Strong	311.634	42	<.001	.990	−.006	.056 [.050–.061]	−.009
France							
Configural	105.361	6	<.001	.994		.093 [.078–.108]	
Weak	92.977	12	<.001	.995	+.001	.059 [.048–.071]	−.034
Strong	266.046	42	<.001	.987	−.008	.053 [.047–.059]	−.006
Germany							
Configural	66.486	6	<.001	.995		.084 [.066–.102]	
Weak	49.271	12	<.001	.997	+.002	.046 [.033–.060]	−.038
Strong	250.396	42	<.001	.982	−*.015*	.059 [.052–.066]	+.013
Partial strong	136.768	38	<.001	.992	−.005	.043 [.035–.050]	−.003
Ireland							
Configural	136.305	6	<.001	.994		.112 [.096–.129]	
Weak	117.613	12	<.001	.995	+.001	.071 [.060–.083]	−.041
Strong	302.272	42	<.001	.987	−.008	.060 [.054–.066]	−.011
Italy							
Configural	90.465	6	<.001	.995		.102 [.084–.121]	
Weak	99.735	12	<.001	.995	.000	.073 [.061–.087]	−.029
Strong	150.705	42	<.001	.993	−.002	.044 [.036–.051]	−.029
Japan							
Configural	134.913	6	<.001	.991		.121 [.104–.139]	
Weak	102.297	12	<.001	.994	+.003	.072 [.059–.085]	−.049
Strong	239.764	42	<.001	.986	−.008	.057 [.050–.064]	−.015

Table 4 continued

Model	χ^2	*df*	*p*	CFI	ΔCFI	RMSEA [CI]	ΔRMSEA
Korea							
Configural	59.904	6	<.001	.998		.069 [.054–.086]	
Weak	58.323	12	<.001	.999	.001	.046 [.034–.058]	−.023
Strong	374.008	42	<.001	.989	−.010	.065 [.059–.071]	+*.019*
Partial strong	242.069	36	<.001	.993	−.006	.055 [.049–.062]	+.009
The Netherlands							
Configural	60.941	6	<.001	.997		.081 [.063–.100]	
Weak	74.174	12	<.001	.996	−.001	.061 [.048–.075]	−.020
Strong	224.406	42	<.001	.989	−.007	.056 [.049–.063]	−.005
Poland							
Configural	29.587	6	<.001	.999		.049 [.032–.067]	
Weak	59.505	12	<.001	.998	−.001	.049 [.037–.062]	.000
Strong	208.863	42	<.001	.991	−.007	.049 [.043–.056]	.000
Slovak Republic							
Configural	102.937	6	<.001	.998		.103 [.086–.121]	
Weak	106.906	12	<.001	.998	.000	.072 [.060–.085]	−.031
Strong	213.481	42	<.001	.996	−.002	.052 [.045–.059]	−.020
Spain							
Configural	160.601	6	<.001	.986		.125 [.108–.142]	
Weak	142.012	12	<.001	.988	+.002	.081 [.069–.093]	−.044
Strong	272.084	42	<.001	.979	−.009	.057 [.051–.064]	−.024
Sweden							
Configural	19.070	6	<.001	.999		.042 [.022–.065]	
Weak	23.160	12	<.001	.999	.000	.028 [.009–.045]	−.014
Strong	309.409	42	<.001	.971	−*.028*	.073 [.065–.080]	+*.045*
Partial strong	110.016	38	<.001	.992	−.007	.040 [.031–.048]	+.012
United Kingdom							
Configural	106.198	6	<.001	.995		.081 [.068–.095]	
Weak	86.582	12	<.001	.997	+.002	.050 [.040–.060]	−.031
Strong	363.668	42	<.001	.985	−*.012*	.055 [.050–.060]	+.005
Partial strong	169.435	38	<.001	.995	−.002	.037 [.031–.043]	−.013
United States							
Configural	87.444	6	<.001	.994		.100 [.082–.119]	
Weak	73.386	12	<.001	.996	+.002	.061 [.048–.075]	−.039
Strong	168.389	42	<.001	.991	−.005	.047 [.040–.055]	−.014

df, degrees of freedom; CFI, comparative fit index; RMSEA, root mean square error of approximation; CI, confidence interval; following Cheung and Rensvold (2002) and Chen (2007), model fit of the more restrictive model should be considered to be significantly worse if the CFI drops by more than .01 and the RMSEA increases by more than .015; changes in CFI and RMSEA that exceed these cutoff values are printed in italic; partial MI has only been tested if assumptions of full MI did not hold

Table 5 Detailed model fit with level of education as grouping variable

Model	χ^2	*df*	*p*	CFI	ΔCFI	RMSEA [CI]	ΔRMSEA
Australia							
Configural	60.600	6	<.001	.997		.065 [.051–.080]	
Weak	55.650	12	<.001	.998	+.001	.041 [.031–.052]	−.024
Strong	352.464	42	<.001	.985	*−.013*	.059 [.053–.064]	*+.018*
partial strong	168.260	36	<.001	.993	−.005	.041 [.035–.048]	.000
Austria							
Configural	90.464	6	<.001	.993		.101 [.083–.120]	
Weak	78.036	12	<.001	.995	+.002	.063 [.050–.077]	−.038
Strong	173.244	42	<.001	.990	−.005	.048 [.040–.055]	−.015
Canada							
Configural	136.690	6	<.001	.996		.054 [.047–.063]	
Weak	141.781	12	<.001	.996	.000	.038 [.033–.044]	−.016
Strong	433.715	42	<.001	.987	−.009	.036 [.033–.039]	−.002
Czech Republic							
Configural	14.133	6	.030	.998		.030 [.009–.050]	
Weak	16.602	12	.100	.999	+.001	.016 [.000–.033]	−.014
Strong	113.970	42	<.001	.982	*−.017*	.033 [.026–.041]	*+.017*
Partial strong	105.489	38	<.001	.983	*−.016*	.034 [.026–.042]	*+.018*
Denmark							
Configural	37.178	6	<.001	.998		.050 [.035–.066]	
Weak	38.582	12	<.001	.998	.000	.033 [.021–.045]	−.017
Strong	259.534	42	<.001	.984	*−.014*	.050 [.044–.056]	*+.017*
Partial strong	160.743	38	<.001	.991	−.007	.040 [.033–.046]	+.007
Estonia							
Configural	142.473	6	<.001	.994		.105 [.090–.120]	
Weak	133.942	12	<.001	.995	+.001	.070 [.060–.081]	−.035
Strong	295.322	42	<.001	.989	−.006	.054 [.048–.060]	−.016
Finland							
Configural	99.526	6	<.001	.988		.101 [.084–.119]	
Weak	102.328	12	<.001	.989	+.001	.070 [.058–.083]	−.031
Strong	409.204	42	<.001	.954	*−.035*	.076 [.069–.083]	+.006
Partial strong	219.040	34	<.001	.977	*−.012*	.060 [.052–.068]	−.010
France							
Configural	113.081	6	<.001	.993		.069 [.081–.112]	
Weak	140.387	12	<.001	.991	−.002	.074 [.064–.086]	+.005
Strong	278.606	42	<.001	.984	−.007	.054 [.048–.060]	−.020
Germany							
Configural	110.476	6	<.001	.990		.110 [.093–.129]	
Weak	103.551	12	<.001	.991	+.001	.073 [.060–.086]	−.037
Strong	217.898	42	<.001	.983	−.008	.054 [.047–.061]	−.019
Ireland							
Configural	129.028	6	<.001	.994		.109 [.093–.125]	
Weak	115.225	12	<.001	.995	+.001	.070 [.059–.082]	−.039
Strong	404.758	42	<.001	.981	*−.014*	.071 [.064–.077]	+.001
Partial strong	216.412	34	<.001	.991	−.004	.056 [.049–.063]	−.015
Italy							
Configural	77.084	6	<.001	.995		.094 [.076–.113]	
Weak	87.66	12	<.001	.995	.000	.068 [.055–.082]	−.026
Strong	156.213	42	<.001	.993	−.002	.045 [.037–.052]	−.023

Table 5 continued

Model	χ^2	*df*	*p*	CFI	ΔCFI	RMSEA [CI]	ΔRMSEA
Japan							
Configural	163.744	6	<.001	.988		.134 [.117–.152]	
Weak	143.148	12	<.001	.990	+.002	.086 [.074–.099]	−.048
Strong	284.701	42	<.001	.981	−.009	.063 [.056–.070]	−.023
Korea							
Configural	64.351	6	<.001	.998		.072 [.057–.089]	
Weak	58.634	12	<.001	.998	.000	.046 [.034–.058]	−.026
Strong	440.681	42	<.001	.984	*−.014*	.071 [.065–.078]	*+.025*
Partial strong	232.12	31	<.001	.992	−.006	.059 [.052–.066]	+.013
The Netherlands							
Configural	62.104	6	<.001	.996		.082 [.064–.101]	
Weak	49.041	12	<.001	.997	+.001	.047 [.034–.061]	−.035
Strong	226.269	42	<.001	.986	*−.011*	.056 [.049–.063]	+.009
Partial strong	152.187	34	<.001	.991	−.006	.050 [.042–.058]	+.003
Norway							
Configural	66.560	6	<.001	.993		.087 [.069–.107]	
Weak	49.041	12	<.001	.997	+.004	.047 [.034–.061]	−.040
Strong	194.135	42	<.001	.982	*−.015*	.052 [.045–.060]	+.005
Partial strong	88.911	34	<.001	.994	−.003	.035 [.026–.044]	−.012
Poland							
Configural	23.953	6	<.001	.999		.043 [.026–.061]	
Weak	54.087	12	<.001	.997	−.002	.046 [.034–.059]	+.003
Strong	165.163	42	<.001	.992	−.005	.042 [.036–.049]	−.004
Spain							
Configural	144.966	6	<.001	.986		.118 [.102–.135]	
Weak	131.807	12	<.001	.988	+.002	.078 [.066–.090]	−.040
Strong	236.637	42	<.001	.981	−.007	.053 [.046–.060]	−.025
Sweden							
Configural	47.064	6	<.001	.995		.075 [.056–.096]	
Weak	31.601	12	<.001	.998	+.003	.037 [.021–.053]	−.038
Strong	114.695	42	<.001	.992	−.006	.038 [.030–.046]	+.001
United Kingdom							
Configural	100.643	6	<.001	.995		.079 [.066–.093]	
Weak	78.061	12	<.001	.996	+.001	.047 [.037–.057]	−.032
Strong	323.446	42	<.001	.985	*−.011*	.052 [.046–.057]	+.005
Partial strong	140.926	34	<.001	.994	−.002	.035 [.029–.041]	−.012
United States							
Configural	79.823	6	<.001	.994		.095 [.077–.114]	
Weak	78.191	12	<.001	.995	+.001	.064 [.051–.077]	−.031
Strong	237.685	42	<.001	.985	−.010	.059 [.051–.066]	−.005

df, degrees of freedom; CFI, comparative fit index; RMSEA, root mean square error of approximation; CI, confidence interval; following Cheung and Rensvold (2002) and Chen (2007), model fit of the more restrictive model should be considered to be significantly worse if the CFI drops by more than .01 and the RMSEA increases by more than .015; changes in CFI and RMSEA that exceed these cutoff values are printed in italic; partial MI has only been tested if assumptions of full MI did not hold

Table 6 Detailed model fit with migration background (i.e., native language same as test language) as grouping variable

Model	χ^2	*df*	*p*	CFI	ΔCFI	RMSEA [CI]	ΔRMSEA
Australia							
Configural	85.551	4	<.001	.996		.079 [.065–.094]	
Weak	65.431	7	<.001	.997	+.001	.051 [.040–.062]	−.028
Strong	146.139	22	<.001	.994	−.003	.042 [.035–.048]	−.019
Austria							
Configural	72.68	4	<.001	.995		.091 [.073–.110]	
Weak	59.994	7	<.001	.996	+.001	.061 [.047–.075]	−.030
Strong	118.108	22	<.001	.993	−.003	.046 [.038–.054]	−.015
Canada							
Configural	164.491	4	<.001	.995		.060 [.053–.068]	
Weak	136.178	7	<.001	.996	+.001	.041 [.035–.047]	−.019
Strong	264.565	22	<.001	.992	−.004	.032 [.028–.035]	−.009
Denmark							
Configural	37.121	4	<.001	.998		.052 [.037–.067]	
Weak	35.175	7	<.001	.998	.000	.036 [.025–.048]	−.016
Strong	259.287	22	<.001	.984	−*.014*	.059 [.053–.065]	+*.023*
Partial strong	119.111	19	<.001	.993	−.005	.041 [.034–.048]	+.005
France							
Configural	106.976	4	<.001	.994		.094 [.079–.110]	
Weak	75.066	7	<.001	.996	+.002	.058 [.047–.070]	−.036
Strong	96.052	22	<.001	.995	−.001	.034 [.027–.041]	−.024
Germany							
Configural	82.137	4	<.001	.993		.095 [.078–.114]	
Weak	92.633	7	<.001	.993	.000	.075 [.062–.089]	−.020
Strong	121.364	22	<.001	.991	−.002	.046 [.038–.054]	−.029
Ireland							
Configural	132.024	4	<.001	.993		.111 [.095–.127]	
Weak	91.795	7	<.001	.996	+.003	.068 [.056–.081]	−.043
Strong	246.612	22	<.001	.988	−.008	.063 [.056–.070]	−.005
Italy							
Configural	70.996	4	<.001	.996		.091 [.073–.110]	
Weak	68.216	7	<.001	.996	.000	.066 [.052–.080]	−.025
Strong	85.641	22	<.001	.996	.000	.038 [.030–.046]	−.028
The Netherlands							
Configural	57.754	4	<.001	.997		.080 [.063–.099]	
Weak	46.094	7	<.001	.998	+.001	.052 [.038–.066]	−.028
Strong	110.544	22	<.001	.995	−.003	.044 [.036–.052]	−.008
Norway							
Configural	37.756	4	<.001	.996		.065 [.047–.085]	
Weak	31.434	7	<.001	.997	+.001	.042 [.028–.057]	−.023
Strong	254.662	22	<.001	.972	−*.025*	.073 [.065–.081]	+*.031*
Partial strong	159.394	20	<.001	.983	−*.014*	.059 [.051–.068]	+*.017*
Slovak Republic							
Configural	110.189	4	<.001	.998		.108 [.091–.126]	
Weak	86.835	7	<.001	.998	.000	.071 [.058–.084]	−.037
Strong	116.951	22	<.001	.998	.000	.044 [.036–.051]	−.027
Spain							
Configural	144.822	4	<.001	.987		.119 [.103–.136]	

Table 6 continued

Model	χ^2	*df*	*p*	CFI	ΔCFI	RMSEA [CI]	ΔRMSEA
Weak	142.655	7	<.001	.987	.000	.088 [.076–.101]	−.031
Strong	120.345	22	<.001	.991	+.004	.042 [.035–.050]	−.046
Sweden							
Configural	37.098	4	<.001	.996		.068 [.049–.088]	
Weak	27.512	7	<.001	.998	+.002	.040 [.025–.057]	−.028
Strong	160.727	22	<.001	.985	−*.013*	.059 [.051–.068]	+*.019*
Partial strong	156.084	20	<.001	.985	−*.013*	.061 [.053–.070]	+*.021*
United Kingdom							
Configural	119.614	4	<.001	.995		.087 [.074–.101]	
Weak	100.439	7	<.001	.996	+.001	.059 [.049–.070]	−.028
Strong	212.626	22	<.001	.991	−.004	.048 [.042–.054]	−.011
United States							
Configural	100.785	4	<.001	.993		.109 [.091–.128]	
Weak	74.171	7	<.001	.995	+.002	.069 [.055–.083]	−.040
Strong	109.647	22	<.001	.994	−.001	.044 [.036–.053]	−.025

df, degrees of freedom; CFI, comparative fit index; RMSEA, root mean square error of approximation; CI, confidence interval; following Cheung and Rensvold (2002) and Chen (2007), model fit of the more restrictive model should be considered to be significantly worse if the CFI drops by more than .01 and the RMSEA increases by more than .015; changes in CFI and RMSEA that exceed these cutoff values are printed in italic; partial MI has only been tested if assumptions of full MI did not hold

Table 7 Overview of freed parameters for testing partial measurement invariance (item: thresholds that have been freed)

Country	Gender	Age groups	Level of education	Migration background
Australia			I_Q04d: 3/4	
Austria		I_Q04d: 3/4		
Czech Republic	I_Q04j: 1/2			
Denmark		I_Q04d: 3/4	I_Q04d: 3/4	I_Q04j: 2/3 I_Q04d: 4
Germany		I_Q04d: 3/4		
Ireland			I_Q04j: 3/4 I_Q04d: 3/4	
Korea		I_Q04d: 1 I_Q04m: 1/2	I_Q04j: 2* I_Q04j: 3/4 I_Q04d: 3/4	
The Netherlands			I_Q04j: 3/4 I_Q04d: 3/4	
Norway			I_Q04j: 3/4 I_Q04d: 3/4	
Sweden		I_Q04d: 3/4		
United Kingdom		I_Q04d: 3/4	I_Q04j: 3/4 I_Q04d: 3/4	

Item I_Q04d, "I like learning new things."; Item I_Q04j, "I like to get to the bottom of difficult things."; Item I_Q04l, "I like to figure out how different ideas fit together."; Item I_Q04m, "If I don't understand something, I look for additional information to make it clearer."; for Korea, the threshold of item I_Q04j has been freed only for the low educational level, which had also free factor loadings for item I_Q04l and I_Q04m

Publisher's Note

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

Received: 27 August 2016 Accepted: 7 July 2017
Published online: 17 July 2017

References

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.

Babanová, A., & Miškolci, J. (2007). *Genderově citlivá výchova: Kde začít [Geschlechtersensible Bildung: Wo soll ich anfangen]?* Prague: Žába na prameni. http://www.rppp.cz/index.php?s=file_download&id=85.

Baltes, P. B., Lindenberger, U., & Staudinger, U. M. (2006). Life span theory in developmental psychology. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology (vol 1). Theoretical models of human development* (6th ed., pp. 569–664). Hoboken, NJ: Wiley.

Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods, 14,* 101–125. doi:10.1037/a0015583.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13*(2), 186–203. doi:10.1207/s15328007sem1302_2.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy, 70*(5/2), 9–49. doi:10.1086/258724.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246. doi:10.1037/0033-2909.107.2.238.

Boeren, E., Nicaise, I., & Baert, H. (2010). Theoretical models of participation in adult education: The need for an integrated model. *International Journal of Lifelong Education, 29*(1), 45–61. doi:10.1080/02601370903471270.

Bose, C., & Kim, M. (Eds.). (2009). *Global gender research: Transnational perspectives (perspectives on gender)*. New York: Routledge.

Breen, R., & Jonsson, J. O. (2005). Inequality of opportunity in comparative perspective. Recent research on educational attainment and social mobility. *Annual Review of Sociology, 31,* 223–243. doi:10.1146/annurev.soc.31.041304.122232.

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*(3), 456–466. doi:10.1037/0033-2909.105.3.456.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464–504. doi:10.1080/10705510701301834.

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.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255. doi:10.1207/S15328007SEM0902_5.

Choy, D., Deng, F., Chai, C. S., Koh, H. J., & Tsai, P. (2016). Singapore primary and secondary students' motivated approaches for learning: A validation study. *Learning and Individual Differences, 45,* 282–290. doi:10.1016/j.lindif.2015.11.019.

Chrisler, J. C., & McCreary, D. R. (Eds.). (2010). *Handbook of gender research in psychology*. Heidelberg: Springer. doi:10.1007/978-1-4419-1465-1.

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research, 12*(3), 229–238. doi:10.1023/A:1023254226592.

Courtney, S. (1992). *Why adults learn: Towards a theory of participation in adult education*. London: Routledge.

Curran, P. J., McGinley, J. S., Bauer, D. J., Hussong, A. M., Burns, A., Chassin, L., et al. (2014). A moderated nonlinear factor model for the development of commensurate measures in integrative data analysis. *Multivariate Behavioral Research, 49,* 214–231. doi:10.1080/00273171.2014.889594.

Desjardins, R. (2003). Determinants of literacy proficiency: A lifelong-lifewide learning perspective. *International Journal of Educational Research, 39*(3), 205–245. doi:10.1016/j.ijer.2004.04.004.

Desjardins, R., Rubenson, K., & Milana, M. (2006). *Unequal chances to participate in adult learning: International perspectives*. Paris: UNESCO.

EACEA P9 Eurydice (Education, Audiovisual and Culture Executive Agency). (2010). *Gender differences in educational outcomes: Study on the measures taken and the current situation in Europe*. http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/120en.pdf.

EACEA P9 Eurydice. (2012). *Key data on education in Europe 2012*. doi:10.2797/77414.

Eccles, J. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist, 44*(2), 78–89. doi:10.1080/00461520902832368.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103–127. doi:10.1037/a0018053.

Freund, P. A., Kuhn, J., & Holling, H. (2011). Measuring current achievement motivation with the QCM: Short form development and investigation of measurement invariance. *Personality and Individual Differences, 51*(5), 629–634. doi:10.1016/j.paid.2011.05.033.

Gaspard, H., Dicke, A. L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., et al. (2015). More value through greater differentiation: Gender differences in value beliefs about math. *Journal of Educational Psychology, 107*(3), 663–677. doi:10.1037/edu0000003.

Gorges, J. (2015). Warum (nicht) an Weiterbildung teilnehmen? Ein erwartungs-wert-theoretischer Blick auf die Motivation erwachsener Lerner. [Why (not) participate in further education? An expectancy-value perspective on adult

learners' motivation]. *Zeitschrift für Erziehungswissenschaft [Journal of Educational Research], 18*(1, Special Issue 30), 9–28. doi:10.1007/s11618-014-0598-y.
Gorges, J., & Hollmann, J. (2015). Motivationale Faktoren der Weiterbildungsbeteiligung bei hohem, mittlerem und niedrigem Bildungsniveau [Motivational factors influencing participation in further education with a high, medium and low level of education]. *Zeitschrift für Erziehungswissenschaft [Journal of Educational Research], 18*(1), 51–69. doi:10.1007/s11618-014-0595-1.
Gorges, J., Maehler, D. B., Koch, T., & Offerhaus, J. (2016). Who likes to learn new things: Measuring adult motivation to learn with PIAAC data from 21 countries. *Large-scale Assessments in Education, 4*(9), 1–22. doi:10.1186/s40536-016-0024-4.
Grouzet, F. M., Otis, N., & Pelletier, L. G. (2006). Longitudinal cross-gender factorial invariance of the Academic Motivation Scale. *Structural Equation Modeling, 13*(1), 73–98. doi:10.1207/s15328007sem1301_4.
Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. F. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3–38). Mahwah, NJ: Lawrence Erlbaum.
Heinz, W. R. (2003). From work trajectories to negotiated careers. The contingent work life course. In J. T. Mortimer & M. J. Shanahan (Eds.), *Handbook of the life course* (pp. 185–204). New York: Kluwer Academic/Plenum Publishers. doi:10.1007/978-0-306-48247-2_9.
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), 1–55. doi:10.1080/10705519909540118.
Ishida, H., Müller, W., & Ridge, J. M. (1995). Class origin, class destination, and education. A cross-national study of ten industrial nations. *American Journal of Sociology, 101*(1), 145–193. doi:10.1086/230701.
Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research, 44*(3), 486–507.
Kilpi, E. (2008). Education in Finland and the ISCED-97. In S. Schneider (Ed.), *The International Standard Classification of Education (ISCED-97). An evaluation of content and criterion validity for 15 European countries*. Mannheimer Zentrum für Europäische Sozialforschung: Mannheim.
Kirsch, I. S., Jungeblut, A., Jenkins, L., & Kolstad, A. (2002). *Adult literacy in America: A first look at the findings of the National Adult Literacy Survey*. Washington, DC: National Center for Education Statistics.
Kosovich, J. J., Hulleman, C. S., Barron, K. E., & Getty, S. (2015). A practical measure of student motivation: Establishing validity evidence for the expectancy-value-cost scale in middle school. *Journal of Early Adolescence, 35*(5–6), 790–816. doi:10.1177/0272431614556890.
Levels, M., van der Velden, R. K. W., & Allen, J. P. (2013). *Skill mismatch and skill use in developed countries: Evidence from the PIAAC study. Research Centre for Education and the Labour Market (ROA) Research Memorandum, 017*. Maastricht: Maastricht University.
Litalien, D., Guay, F., & Morin, A. S. (2015). Motivation for PhD studies: Scale development and validation. *Learning and Individual Differences, 41,* 1–13. doi:10.1016/j.lindif.2015.05.006.
Maehler, D. B., Martin, S., & Rammstedt, B. (2017). Coverage of the migrant population in large-scale assessment surveys: Experiences from PIAAC in Germany. *Large-scale Assessments in Education, 5,* 9. doi:10.1186/s40536-017-0044-8.
Maehler, D. B., Massing, N., Helmschrott, S., Rammstedt, B., Staudinger, U. M., & Wolf, C. (2013). Grundlegende Kompetenzen in verschiedenen Bevölkerungsgruppen [Basic skills of different population groups]. In B. Rammstedt (Ed.), *Grundlegende Kompetenzen Erwachsener im internationalen Vergleich—Ergebnisse von PIAAC 2012 [Basic skills of adults in international comparison—Results from PIAAC 2012]* (pp. 77–124). Münster: Waxmann.
Maehler, D. B., Massing, N., & Rammstedt, B. (2014). *Grundlegende Kompetenzen Erwachsener mit Migrationshintergrund im internationalen Vergleich: PIAAC 2012. [Basic skills of adults with migration background in international comparison: PIAAC 2012]*. Münster: Waxmann.
Maehr, M. L., & Zusho, A. (2009). Achievement goal theory. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation in school* (pp. 77–104). New York: Routledge.
Manninen, J. (2005). Development of participation models. From single predicting elements to complex system models. In ERDI (Ed.), *Participation in adult education. Theory, research, practice* (pp. 11–22). Bonn: Editor.
Marsh, H. W. (1993). The multidimensional structure of academic self-concept: Invariance over gender and age. *American Educational Research Journal, 30*(4), 841–860. doi:10.3102/00028312030004841.
Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543. doi:10.1007/BF02294825.
Merriam, S. B., Caffarella, R. S., & Baumgartner, L. M. (2012). *Learning in adulthood: A comprehensive guide*. San Francisco: Wiley & Sons.
Muthén, B., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. *Mplus Web Notes, 4*(5), 1–22.
Muthén, L., & Muthén, B. (1998–2012). *Mplus user's guide. Version 7*. Los Angeles, CA: Muthén & Muthén.
O'Connell, P. J. (1999). *Adults in training: An international comparison of continuing education and training. Center for Educational Research and Innovation, Report No. CERI/WD(99)1*. Paris: OECD.
OECD. (2005). *Promoting adult learning*. Paris: OECD.
OECD. (2010). *PISA 2009 at a glance*. Paris: OECD.
OECD. (2013a). *OECD Skills Outlook 2013: First results from the Survey of Adult Skills*. Paris: OECD.
OECD. (2013b). *Technical report of the Survey of Adult Skills*. Paris: OECD.
OECD [Organization for Economic Co-Operation and Development]. (2004). *Learning for tomorrow's world. First results from PISA 2003*. Paris: OECD.
OECD, & Statistics Canada. (2000). *Literacy in the information age: Final report of the International Adult Literacy Survey*. Paris: OECD.
Reder, S. (1994). Practice-engagement theory: A sociocultural approach to literacy across languages and cultures. In B. M. Ferdman, R.-M. Weber, & A. G. Ramirez (Eds.), *literacy across languages and cultures* (pp. 33–74). Albany, NY: State University of New York Press.

Reder, S., & Bynner, J. (Eds.). (2009). *Tracking adult literacy and numeracy skills: Findings from longitudinal research*. New York, NY: Taylor & Francis.

Rhemtulla, M., Brosseau-Liard, P. E., & 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. doi:10.1177/0013164413498257.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67. doi:10.1006/ceps.1999.1020.

Samejima, F. (1969). Estimation of ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 17,* 1–100.

Santiago, P., Gilmore, A., Nusche, D., & Sammons, P. (2012). *OECD reviews of evaluation and assessment in education: Czech Republic 2012*. Paris: OECD Publishing. doi:10.1787/97892641167.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23–74.

Schneider, S. (2009). *Confusing credentials: The cross-nationally comparable measurement of educational attainment*. DPhil: University of Oxford.

Schunk, D. H., Meece, J. R., & Pintrich, P. R. (2014). *Motivation in education: Theory, research and applications*. London: Pearson Education.

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. doi:10.1177/0013164413481802.

Skelton, C., & Francis, B. (2006). *The SAGE handbook of gender and education*. London: Sage.

Stanat, P., Rauch, D., & Segeritz, M. (2010). Schülerinnen und Schüler mit Migrationshintergrund [Students with migration background]. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider, & P. Stanat (Eds.), *PISA 2009. Bilanz nach einem Jahrzehnt [Taking stock after one decade]* (pp. 200–230). Münster: Waxmann.

Statistics Canada. (2013). *Skills in Canada: First results from the Programme for International Assessment of Adult Competencies (PIAAC), 2012*. Statistics Canada Catalogue No. 89-555-X. Ottawa.

Statistics Canada, & OECD. (2005). *Learning a living: First results of the adult literacy and life skills survey*. Paris: OECD.

Staudinger, U. M., Marsiske, M., & Baltes, P. B. (1995). Resilience and reserve capacity in later adulthood: Potentials and limits of development across the life span. In D. Cicchetti & D. Cohen (Eds.), *Developmental psychopathology. Risk, disorder, and adaptation (vol 2)* (pp. 801–847). New York, NY: Wiley.

Steenkamp, J.-B., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national research. *Journal of Consumer Research, 25*(1), 78–90. doi:10.1086/209528.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioural Research, 25*(2), 173–180. doi:10.1207/s15327906mbr2502_4.

Strakova, J. (2008). The Czech educational system and evaluation of the ISCED-97 implementation. In S. Schneider (Ed.), *The International Standard Classification of Education (ISCED-97). An evaluation of content and criterion validity for 15 European countries*. Mannheimer Zentrum für Europäische Sozialforschung: Mannheim.

Su, X., McBride, R. E., & Xiang, P. (2015). College students' achievement goal orientation and motivational regulations in physical activity classes: A test of gender invariance. *Journal of Teaching in Physical Education, 34*(1), 2–17. doi:10.1123/jtpe.2013-0151.

Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika, 52,* 393–408. doi:10.1007/BF02294363.

UNESCO [United Nations Educational, Scientific, and Cultural Organization]. (2011). *International Standard Classification of Education. ISCED 2011*. Paris: UNESCO. http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf.

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. doi:10.1177/109442810031002.

Watt, H. M. G., & Eccles, J. S. (Eds.). (2010). *Gender and occupational outcomes: Longitudinal assessment of individual, social, and cultural influences*. Washington, DC: APA.

Wentzel, K. R., & Wigfield, A. (Eds.). (2009). *Handbook of motivation in school*. New York: Routledge.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81. doi:10.1006/ceps.1999.1015.

Wigfield, A., Eccles, J. S., Schiefele, U., Roeser, R. W., & Davis-Kean, P. (2006). Development of achievement motivation. In N. Eisenberg (Ed.), *Handbook of child psychology: Social, emotional, and personality development* (Vol. 3, pp. 933–1002). Hoboken: Wiley.

Wigfield, A., Tonks, S., & Klauda, S. L. (2009). Expectancy-value theory. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 55–75). New York: Routledge.

Woo, S. E., Gibbons, A. M., & Thornton, G. C. (2007). Latent mean differences in the facets of achievement motivation of undergraduate students and adult workers in the US. *Personality and Individual Differences, 43*(7), 1687–1697. doi:10.1016/j.paid.2007.05.006.

Zhu, X., Sun, H., Chen, A., & Ennis, C. (2012). Measurement invariance of expectancy-value questionnaire in physical education. *Measurement in Physical Education and Exercise Science, 16*(1), 41–54. doi:10.1080/1091367X.2012.639629.