Studies in Educational Evaluation, Dec 2020 https://doi.org/10.1016/j.stueduc.2020.100965

Testing Measurement Invariance of PISA 2015 Mathematics, Science, and ICT Scales Using the Alignment Method

Bryce Odell, Mark Gierl, Maria Cutumisu*

Centre for Research in Applied Measurement and Evaluation, Department of Educational Psychology, University of Alberta, Canada

{btodell, mgierl, cutumisu}@ualberta.ca

*Corresponding Author

Abstract

This study investigates measurement invariance of the mathematics, science, and ICT scales across the 47 countries that participated in the PISA 2015 *ICT Familiarity Questionnaire*. Knowing whether the same constructs and measurements can be reliably compared across countries constitutes an important goal. The Alignment method is employed to test the measurement invariance of the three scales. The results show that mathematics and science scores are highly invariant and can be used to compare countries, whereas the ICT scale is mostly non-invariant and cannot be used to reliably compare ICT means across all participating countries. Implications and limitations are discussed.

Keywords: Alignment; Measurement Invariance; PISA; ICT; Mathematics; Science

1. Introduction

It is undeniable that technology plays a large role in education across the world (Organization for Economic Co-operation and Development; OECD, 2015; Johnson et al., 2014). Thus, large international surveys like the Programme for International Student Assessment (PISA) have been measuring information and communication technology (ICT) across the participating countries. PISA ranks countries based on their academic performance, which is determined by the scores students get on measurements including mathematics, science, and reading (OECD, 2016). Researchers have also compared technology use between countries using PISA's *ICT Familiarity Questionnaire* (Zhang & Liu, 2016; Meng, Qiu, & Boyd-Wilson, 2018; Aryadoust, 2020). Before making cross-cultural comparisons based on these scales, researchers need to be confident that variables are measured the same way between groups. This introduces the question of measurement invariance. The current exploratory study aims to probe measurement invariance in the ICT scales from PISA 2015 to ascertain whether they can be used for cross-country comparisons using the novel Alignment method provided by Asparouhov and Muthén (2014) as well as validating the Alignment method against the established invariant mathematics and science scores from PISA 2015. This robust, yet flexible method of invariance testing overcomes the main limitations of the traditional confirmatory factor analysis (CFA), such as comparing complex models across many groups. In comparison, CFA is a method for testing factor structure and relationships between observed variables and their latent constructs (Suhr, 2006). As suggested by Raudenská (2020), this alignment method should be widely considered by researchers for its advantages that will be discussed in this study. Establishing measurement invariance is important when researchers aim to include multiple countries in studies based on their scale scores. If this is not done, researchers may be unknowingly comparing unrelated constructs and measures. Specifically, testing the measurement invariance of the ICT scale will establish whether or not students from all participating countries ca be compared based on their ICT interactions. The mathematics and science scales will be used to establish if the alignment method for determining measurement invariance is effective.

We start by outlining the conceptual background of measurement invariance and the PISA scales, and we review the literature on the topic of ICT, mathematics, and science comparisons across countries. Following that, we outline our research question, the methods used to answer the question, and the results. Finally, we conclude by discussing the results, limitations, future directions, and implications.

2. Conceptual Background

2.1. Information and Communication Technology (ICT)

ICT represents digital technology (e.g., computers, Internet connections, digital media, etc.) that is used for sharing and storing digital information and communicating with others (OECD, 2005). ICT includes the important feature of technology used for the purpose of communication and sharing (Murray, 2011; Daintith, 2009). ICT is defined by PISA as "the use of any equipment or software for processing or transmitting digital information that performs diverse general functions, whose options can be specified or programmed by its user" (OECD, 2005).

Current ICT access is not consistent across countries (OECD, 2017a). For instance, only one in five European students attends a school with access to high-speed Internet (European Union, 2019a). Concomitantly, countries such as Russia have the potential to achieve more integrated ICT use (Dneprovskaya, Bayaskalanova, Ruposov, & Shevtsova, 2018). There is a great demand for ICT to be integrated into higher education. However, to date, there is a lack of governmental support and organization for this to happen. Moreover, Nordic countries such as

Finland continue to have better access to technology and be Digital Frontrunners, whereas others like Bulgaria remain Digital Challengers (European Union, 2019a, 2019b, 2019c; Ridao-Cano & Bodewig, 2018; Novak et al., 2018). Also, there seems to be a link between ICT and academic achievement, as Digital Frontrunner countries often have higher classroom achievement than Digital Challengers (Novak et al., 2018).

2.2. Mathematics Literacy

Mathematics literacy was the main focus in PISA 2012, but was also assessed in PISA 2015, as it has been in all previous iterations of PISA. Many facets of modern life rely on one's understanding of and ability to use mathematics (OECD, 2013). Mathematics content knowledge, mathematical reasoning, and use of mathematical tools are essential to succeed in today's modern global society. For example, being able to estimate distances, managing finances, time management, and basic calculations are all tasks that rely on mathematics literacy and are performed daily by citizens. Mathematics literacy is united with scientific literacy as many of the skills are dependent on each other (National Mathematics Advisory Panel, 2008).

2.3. Science Literacy

Science literacy was the focus of PISA 2015, being extremely important for facing some of humanity's greatest challenges, such as reducing waste, global preventable diseases, and curbing climate change (United Nations Environment Programme; UNEP, 2019; CDC, 2019). It is important for citizens of our global society to have the scientific literacy skills to understand the legitimacy of these issues and be able to form educated opinions and partake in meaningful discussions (OECD, 2017a). A solid grasp on the scientific process and the ability to think scientifically is critically important in today's age of 'Fake News' to give students a fighting chance to sift through the enormous amounts of claims and stories that will get their attention (van der Linden, Maibach, Cook, Leiserowitz, & Lewandowsky, 2017).

2.4. Measurement Invariance

Comparing multiple groups has been a staple in social science research. However, this comparison becomes challenging when the scope expands to large-scale studies that use populous samples from separate countries as groups. In these cases, measurement invariance must be tested before comparisons of measures and scores can occur and before any cross-group or within country comparisons can be examined (Vandenberg & Lance, 2000). Measurement non-invariance occurs when the scores used to compare groups are not truly based on the intended construct, but rather on other variation due to factors such as time, methods, or culture (van de Vijver & Leung, 1997; Meade & Lautenschlager, 2004).

The goal of measurement invariance studies is to establish whether or not the same construct is being measured across different groups. Thus, measurement invariance is necessary to establish before cross-group comparisons can occur (Vandenberg & Lance, 2000). Specifically, measurement invariance aims to test whether constructs represent the same underlying attributes and measured scores have the same meaning in different conditions or groups (Meade & Lautenschlager, 2004). Measurement invariance must be established before means from different groups can be compared or else the comparison would not be meaningful (Millsap, 2012). Traditionally, three levels of measurement invariance are used in studies: configural, metric, and scalar (Kline, 2015). Each level assumes the previous level of invariance. Configural invariance is the most basic level that indicates similar latent constructs for different groups. Metric, or weak, invariance requires factor loadings to be equal across groups as well as having better fit statistics over the previous model. If metric invariance is established, it allows for comparison of group-estimated factor variances using significance tests. Scalar, or strong, invariance requires factor loadings and intercepts to be equal across groups to be certain that individuals from different groups respond the same way to a measure. If scalar invariance is established, then means can be compared validly across countries.

A review of measurement invariance testing by Byrne and van de Vijver (2017) found that researchers acknowledge the need for measurement invariance tests for multiple group comparisons. However, the majority of studies only used two groups for comparison. The amount of studies further drops as the number of focal groups increases. In the literature, exploratory measures are being used without confirmatory measures and this is often paired with a lack of measurement invariance testing and weak validity arguments (Kane, 2013). Measurement invariance often goes unexamined for measures that intend to be used on populations from different cultures (Dong & Dumas, 2020). Identifying a consistency and generalizability in scores is an important step in comparing groups internationally. Schuler et al. (2014) examined measurement invariance in disease research and found that a significant amount of measurement non-invariance is present when using multiple group confirmatory factor analysis (MGCFA). However, the researchers recommend that more testing for measurement invariance is conducted, as scalar invariance is rarely established. MGCFA is used for cross-group comparisons of a latent variable, while accounting for measurement invariance (Meredith, 1993; Steenkamp & Baumgartner, 1998). However, if some of the examined variables are non-invariant, then the user must exert considerable effort to make model modifications that allow for any comparisons (Muthén & Asparouhov, 2014).

The Alignment method was proposed as a solution to estimate the means and intercepts of many groups, while allowing for some flexibility in measurement invariance. It outperforms multiple-group techniques, such as multiple-group Structural Equation Modelling (SEM), as it facilitates invariance testing with many groups (Asparouhov & Muthén, 2014). SEM can be considered a combination of multiple regression and factor analysis that allows for the examination of multiple predictor and outcome variables simultaneously (Ullman & Bentler, 2003). The Alignment option is a better choice than a CFA, because a CFA may fail due to many modification indices and poor scalar model fit. Full invariance is rarely achieved in large datasets due to troublesome modification indices and complicated models from releasing constraints. The Alignment method simplifies and automates invariance testing among groups with expected non-invariance (e.g., culture in different countries; Muthén & Asparouhov, 2014; Byrne & van de Vijver, 2017). For example, Coromina and Bartolomé Peral (2020) used both CFA and Alignment methods to examine cross-cultural trust in government and concluded that the Alignment method was more flexible, realistic, and less restrictive than the traditional CFA method. Typically, using SEM to complete a CFA for multiple groups means making many one-to-one comparisons, each requiring a baseline model, which would result in a lot of tedious work (Byrne & van de Vijver, 2017).

2.5. The PISA ICT Scale

Given that the PISA data are self-reported for the ICT indicator variables, they are subjective and may not accurately represent the true scores of the students. As such, ICT is an interesting subjective variable that requires investigation into its comparability across different countries. The ICT variables from the *ICT Familiarity Questionnaire* include nine self-reported Likert-subscales on a logit scale, where 0 represents the OECD average (OECD, 2014a). Each question measures a slightly different aspect of ICT use, availability, and comfort: ENTUSE (IC008) for ICT use outside of school leisure; HOMESCH (IC010) for ICT use outside of school for schoolwork; and USESCH (IC011) for use of ICT at school in general (OECD, 2017b; 2014a). The COMPICT (IC009) scale is for students' perceived ICT competence. AUTICT

(IC013) measures students' perceived autonomy for ICT use. INTICT (IC015) captures students' ICT interest. SOIAICT (IC016) measures students' inclusion of ICT as a topic in social interaction. ICTSCH (IC014) measures ICT availability at school and ICTHOME (IC001) measures availability at home. Note here the possible issue of reference bias caused by translational differences in the ICT items (Heckman & Kautz, 2013).

2.6. The PISA Mathematics and Science Scales

PISA conceptualizes its construct of mathematical literacy as describing "the capacities of individuals to reason mathematically and use mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena." (OECD, 2017a). PISA assessments highlight the importance of understanding pure and abstract mathematical concepts as well as a student's ability to apply these concepts in practice. Using mathematics to form judgements and decisions in the assessment is reflective of the role of mathematics in a citizen's every-day life. PISA distills the mathematics domain into three aspects: identifying the mathematics component of a problem and solving it, the mathematics content knowledge necessary to test these skills, and the context of the mathematics problems.

PISA explains that the focus of their scientific literacy measure is both science and science-based technology (OECD, 2017a). Science is the process of asking informed questions and methodically seeking the answers and comparing them against other answers, while science technology provides solutions to the problems raised in the process of scientific inquiry. They are both critical to scientific literacy. In PISA 2015, scientific literacy represents a student's ability to "explain phenomena scientifically", "evaluate and design scientific enquiry", and "interpret data and evidence scientifically" in addition to content knowledge (OECD, 2017a).

The PISA assessment aims to understand students' ability to use a broad view of content knowledge to answer scientific questions that are important in our global society.

3. Literature Review

3.1 The Alignment Method

Current reviews report that scalar or strict measurement invariance is most often established in age and gender groups for measurement tools, however this is rarely established in cross-cultural groups (Dong & Dumas, 2020). This produces measures that are being used to score groups other than the ones used to establish their standardized scores, which could cause biased and invalid results based on cultural groups (Guenole & Brown, 2014). Byrne and van de Vijver (2017) examined family functioning from the *Family Value Scale* from 27 countries. These researchers found that the Alignment method overcame the limitations from the more traditional CFA methods and found more trustworthy results for comparability between countries. Lamm, Do, Rodriguez, Scales, and Roehlkepartain (2019) tested the measurement invariance of the *Developmental Asset Profile* across 30 countries and found a mix of invariant and non-invariant parameters, which led them to the conclusion that some comparisons can reliably be made. Fischer, Praetorius, and Klieme (2019) used the Alignment method to find that the cross-country comparisons based on teaching quality are limited and the comparable factors vary depending on the country. Amérigo et al. (2020) successfully used the alignment method to establish scalar measurement invariance of the Multidimensional Environmental Concern Scale across different cultures. The authors note the flexibility and realistic assumptions of the method as commendable attributes for assessing measurement invariance. Throughout the literature, the PISA ICT scale has never been tested for measurement invariance using the Alignment method from Muthén and Asparouhov (2017).

3.2. ICT

Previous researchers have examined the measurement invariance of PISA's ICT scales before testing and comparing various models. Meng, Qiu, and Boyd-Wilson (2018) established scalar measurement invariance for German and Chinese scores from PISA's 2015 ICT scales. However, they only examined two countries and highlighted the need for validation studies of the instrument across countries and cultures. Zhang and Liu (2016) examined PISA ICT scales through multiple iterations and countries. In order to discuss comparisons, the researchers chose similar measures from year to year despite acknowledging that there were differences between iterations. Thus, measurement invariance needs to be established between different iterations of measurements delivered in PISA's three-year increments. Aryadoust (2020) investigated measurement invariance of the information technology development index of the Pearson Test of English Academic reading assessment. Like many other researchers, they used MGCFA to test configural, metric, scalar, and structural invariance. Each country in the sample had a maximum of 470 participants and these were grouped into quartiles where the measurement invariance was then tested. The researchers established structural measurement invariance across the four groups. However, interpretations of these specific results are limited because they must be related back to the four groupings of countries, which are not identified.

3.3. Mathematics and Science

As mentioned before, it is critically important to test measurement invariance across scales before making comparisons between groups, however this also stands for other scales such as the mathematics and science scales (He, Barrera-Pedemonte, & Buchholz, 2019). These researchers remark on the comparison of countries by scales in PISA and Trends in International Mathematics and Science Study (TIMMS) data as well as on a lack of checking for measurement invariance prior to making comparisons. Again, MGCFA was used to establish metric invariance between cultures. He et al. (2019) call for more attention to testing measurement invariance of academic subjects such as the mathematics and science scales before using them for comparisons. Kalaycioglu (2015) examined and compared the relationships between SES, math self-efficacy, and math anxiety with math achievement in five countries in PISA 2012. Kalaycioglu (2015) determined that latent mean comparison was possible between these variables as they established partial scalar invariance with multiple group SEM. However, the invariance of the models as a whole is tested rather than the invariance of the separate PISA measures. The measurement invariance of the math achievement is not discussed, yet comparisons are made. In the 2015 PISA Technical report (OECD, 2017b), slope and differential item functioning parameters were constrained to establish metric invariance while testing their models for mathematics, science, reading, and financial literacy using IRT. With more constraints, the models become closer to scalar invariance. PISA is confident in their established measurement invariance for their mathematics and science literacy scales, as they use them as a major point of comparison and rank countries based on their mean scores. This provides us with an opportunity to evaluate other methods of measurement invariance testing using the same data.

In sum, the literature shows that measurement invariance is not always rigorously examined and, when it is, the most common method used is an MGCFA. As concluded by Byrne and van de Vijver (2017), this is most often done properly in groups of two, while comparisons of multiple groups are much fewer. Before researchers make comparisons between groups, they need to be confident that either they or previous researchers have established the proper level of measurement invariance. In the current study, we aim to do this for the ICT scale for all countries who participated, as well as for the mathematics and science measures for the same groups of participants who completed the *ICT Familiarity Questionnaire*. More information on the relationships among mathematics, science, and ICT is included in our prior research (Odell, Cutumisu, & Gierl, 2020).

4. Research Questions

Is the *ICT Familiarity Questionnaire* from PISA 2015 measurement invariant across participating countries? Is the Alignment method an effective tool for establishing measurement invariance in the mathematics and science scales from PISA 2015, which have already been validated as being scalar measurement invariant by OECD (2017b) using the IRT item-fit method (OECD, 2016)? It is critically important to establish measurement invariance in research that compares multiple countries, because the scales being examined need to measure the same constructs to be able to make confident comparisons. Specifically, the present research examines whether the ICT, mathematics, and science scales from PISA 2015 are measurement invariant across the participating countries. We advance several hypotheses.

Hypothesis 1: The Alignment method will confirm measurement invariance of the mathematics and science scales that has been established by PISA 2015, which confirms that it can be used as a useful test of invariance in similar situations. The mathematics and science scales have benefitted from more attention, development, and improvement over the many PISA iterations than the ICT scales in comparing different cultures (OECD, 1999, 2003, 2006, 2009, 2013, 2017a).

Hypothesis 2: The PISA 2015 ICT scale would not be completely measurement invariant across participating countries, because the *ICT Familiarity Questionnaire* is a less developed scale than mathematics and science. PISA does not provide ranked comparisons of countries based on their ICT scores in the way it provides a ranked list of participating countries based on their mathematics and science performance, which can only be done once scalar invariance is established.

Similar to our goals, McLarnon and Romero (2020) used the Alignment Method (Asparouhov & Muthén, 2014) to examine the measurement invariance of an existing questionnaire that has been used across many countries. These researchers found limitations in comparability of scores across countries, which implies other measures may not have been validated for cross-cultural comparisons.

5. Methods

5.1. Data Source

PISA is a large collection of surveys that aim to test the content knowledge and applicability of knowledge of young students in many countries (OECD, 2016). A main goal of PISA is to assess the learning quality and equity across different groups to inform educators and policy. This research employs publicly-available data from the 2015 PISA database that contains approximately 540,000 students from 72 countries. Content specialists and measurement experts help design PISA to test the applicability of students' learning (OECD, 2017b). Additionally, the surveys are used to measure other variables related to education such as involvement with ICT, which constitutes the focus of this research. It is worth noting that, in PISA 2015, there was a move to include computer-based assessments where possible. Jerrim et al. (2018) explored the possibility that this change in mode would reduce the comparability between countries. However, the authors concluded that PISA demonstrates good comparability of the computer-based and paper-based assessments. The test items are delivered as paper or computer-based questionnaires and quizzes every three years. The data used for this research include the *ICT*

Familiarity Questionnaire (OECD, 2014a) and the *Student Questionnaire* (OECD, 2014b) from 47 participating countries. These scales were all collected using computer-based questionnaires.

5.2. Sample Description

The sample for the final SEM consisted of n = 11,810 students who attended school part-time or full-time in their countries (OECD, 2016; 2017b). The sample for the mathematics and science Alignment models were n = 369,450 and the sample for the ICT Alignment model was n = 332,522. The age of the students ranges from 15 years and 3 months to 16 years and 2 months. Participants were randomly chosen within their cluster sample of schools. Weighting variables were used to compare students from different schools, with the aim to equally represent participants. Countries chose whether to participate in the *ICT Familiarity Questionnaire*. In PISA 2015, only 47 countries had their students complete all sections of this additional questionnaire. The different participating countries have a wide array of economic and technological strength. The *Growing United: Upgrading Europe's Convergence Machine* report by the World Bank on the European Union was used to identify where on the scale some of the countries landed (Ridao-Cano & Bodewig, 2018). This report uses PISA and economic data to determine the divides that exist among European countries.

5.3. Measures

Information was collected from the students on their interactions with ICT as well as on their competence in mathematics and science. All measures used in this research were obtained from the PISA 2015 database mathematics and science plausible values and the nine ICT subscales from the *ICT Familiarity Questionnaire*. The structure of these three measures is discussed in the following paragraphs.

5.3.1. ICT

For the nine ICT subscales, students answered a set of items like, "How often do you use digital devices for the following activities at school." They were then presented with a list including websites, computers, simulations, and more. Responses were given on a 5-point ordinal scale with options from 1 = *Never or hardly ever* to 5 = *Every day*. Images of these scales are available in Appendix A. The availability subscales were indices calculated as the sum across all their comprising availability items, while the others were scaled indices computed based on IRT. Cronbach's alpha (α) is calculated by OECD and recorded for each subscale to compare internal consistencies between countries (OECD, 2017b). A value of 1 signifies perfect internal consistency, while a value of at least 0.7 indicates acceptable internal consistency. These variables are briefly described below and Appendix A shows examples of each subscale. A full list of scale reliabilities for all participating countries is included in Table 16.64 of the *2015 PISA Technical Report* (OECD, 2017b).

5.3.2. Mathematics and Science

To measure mathematics and science, different groups of students were given cognitive performance tests including similar sets of parallel, overlapping items from a total pool of 82 mathematics or science items (OECD, 2017b). PISA compared students' performance scores using Item Response Theory (IRT) to account for several challenges (Mislevy, Beaton, Kaplan, & Sheehan, 1992). Specifically, PISA uses complex sample designs (i.e., including unequal probabilities and stratifications) that must be taken into account when approximating scores. Distilling many test statistics into fewer plausible values enables researchers to add scores to their analyses rather than performing more advanced statistics to make the same comparisons. The IRT provides an improved estimation, even if the marginal analysis is not ideal. PISA calculated ten plausible values for each student in both subjects “using information from the student context questionnaire in a population model” (OECD, 2017b, p. 128). These values were based on students’ mathematics and science performance as well as other student background information. For all OECD participants the variable scales were on a scale of 0 to 1000 with a mean of 500 and a standard deviation of 100.

5.4. The Alignment Method

This method can be based on either maximum-likelihood estimation (MLE) or Bayes estimation and can be performed with either free or fixed estimation. Free Alignment is suggested by Asparouhov and Muthén (2014) as a better option than fixed Alignment, but it is possible that the model will not be identified. If that is the case, then fixed Alignment should be used. The difference between the two is that either all factor loadings and intercepts are freely estimated or one group is selected to have a factor mean set to 0 and an intercept set to 1. Free estimation allows more bias overall across all the parameters and works best when there is about 10% to 20% of non-invariant parameters. Muthén and Asparouhov (2014) provide an acceptable cutoff of 25% as the maximum amount of non-invariance allowed.

Alignment minimizes the amount of measurement non-invariance by estimating the factor means and variances and it identifies parameters without imposing scalar invariance. Different restrictions are imposed that optimize a simplicity function (Muthén & Asparouhov, 2014). The simplicity function is optimized at a few large non-invariant parameters and many approximately invariant parameters rather than many medium sized non-invariant parameters. According to Muthén and Asparouhov (2014, p. 2),

> "Adding a simplicity function gives the necessary restrictions to identify the model. The simplicity function minimizes with respect to [factor means] and [factor variance] the total loss/simplicity function *F* which accumulates the total measurement non-invariance over the items."

There are two steps that occur automatically during the Alignment analysis. The first is an estimation of a configural model, where loadings and intercepts estimated freely, factor means are fixed at 0, factor variances are fixed at one, while factor loadings and intercepts are freely estimated. In the end, the final Alignment model will have the same fit as this configural model. The second step is Alignment optimization, where factor means and variances are assigned values based on a pattern of parameter estimates using a simplicity function to minimize the total amount of non-invariance for every pair of groups and every intercept and loading using a simplicity function similar to rotations in an exploratory factor analysis (EFA). The estimation stops when the least amount of non-invariant parameters is achieved.

The Alignment output provides a table of which groups are invariant for a given parameter. The latent means for each parameter are also given for every group. This output also shows which groups have a significantly lower mean than the mean of the group being examined. The R-squared value is provided for each parameter and demonstrates the variance across groups that can be explained by the variation in factor means and variances (Byrne & van de Vijver, 2017). A value of 1 indicates complete invariance because the variability in item parameters is completely explained by group mean differences. A value near 0 indicates that the group mean differences explain none of the variability in item parameters (Asparouhov & Muthén, 2014).

5.5. Analytic Plan

5.5.1. ICT Alignment

The following steps explain the process of examining measurement invariance using the Alignment method. Initially, three structural equation Alignment models were devised to explore whether mathematics scores, science scores, and ICT scales are measurement invariant across the 47 participating countries, and to compare means. This method was used to answer the research question: *Are the ICT, science, and mathematics variables measurement invariant across the countries participating in the PISA ICT questionnaire?* In these models, the maximum likelihood estimation with robust standard errors (MLR) estimator was used to incorporate the weighting variable (W_FSTUWT) to allow for cross country comparisons. The country ID variable (CNTRYID) was used to define the 47 countries as latent classes. Table 1 shows all the country IDs. The default number of iterations was preserved, the number of random sets of starting values was set to 60 (double the default), and convergence was restricted to .001, as suggested by the *MPlus 8* user manual (Muthén & Muthén, 1998-2017). At first, free estimation was used to find a reference group for fixed estimation because the models were not identified with free Alignment. The reference groups used were the countries with the factor means closest to 0, either positive or negative. For the ICT model, the country with the three factor means closest to 0 was chosen as the reference group. The requested outputs were TECH1 and SVALUES for the starting values of the parameters, TECH8 for the optimization history, and ALIGN for factor loadings and intercept comparisons as well as measurement invariance of each country for the factors.

Several error messages of missing variables and variances of predictor variables equaling zero prompted an investigation to reveal that the residual variances of ICTHOME and ICTSCH were fixed for Germany (CNTRYID 276). Indeed, Germany was missing all values for ICTHOME and ICTSCH and was subsequently removed from all three of the Alignment models. The new sample size without Germany was n = 362,946. With Germany removed from the dataset, a free Alignment model was conducted and The Netherlands (CNTRYID 528) was found to have the sum of the absolute values of their factor means closest to 0, at .67. Therefore, the model was reconfigured to exclude Germany from the classes, resulting in 47 classes. *MPlus* did not include 30,424 cases, because they contained missing values. This updated model only produced one warning: 30,424 cases were not included due to missing variables of interest. This warning is not a concern, as there are 362,946 observations excluding Germany. In the end, 332,522 cases were included in the dataset of 47 countries for the ICT model and remained as 362,946 for the mathematics and science models, as there are no missing plausible values. See Appendix B for the *MPlus 8* code we used for the ICT Alignment analysis.

5.5.2. Mathematics and Science Alignment

Measurement invariance of the mathematics and science scales was tested to evaluate the validity of the Alignment method. Two other similar models were created to test the measurement invariance of mathematics and science plausible values. These models are the same as the ICT model, except that the latent variables “MATH” and “SCIENCE” were created from their respective ten plausible values. The fixed country was based on each free Alignment model. For the mathematics model, the fixed country was Croatia (CNTRYID 191), whereas for the science model, the fixed country was Ireland (CNTRYID 372). The plausible values that make up the latent variables were divided by 100 to bring them closer to the ICT scale and ease interpretation of results. See Appendices C and D for the *MPlus 8* code we used for the mathematics and science Alignment analysis.

6. Results

6.1. ICT Alignment

The Alignment method was used to answer the following research question: *Is ICT measurement invariant across the countries participating in the PISA ICT Familiarity Questionnaire?* The Alignment model shows which item intercepts and factor loadings are invariant in all the groups. Table 2 presents the non-invariance percentages of the factor loadings and intercepts of the nine ICT scales. If Muthén and Asparouhov's (2014) cutoff of 25% as the maximum amount of non-invariance is surpassed, then the latent mean estimation may be untrustworthy. All the 9 ICT scales have high amounts of non-invariance, with an average of 68.60% for intercepts and 34.8% for factor loadings. The finding of fewer non-invariant factor loadings than intercepts follows the trend of previous researchers (Crane, Belle, & Larson, 2004; Meiring, van de Vijver, Rothmann, & Barrick, 2005; Byrne & van de Vijver, 2017). The factor intercepts of all 9 ICT scales were above the 25% cutoff. The least non-invariant intercept parameters were ICTHOME and ICTSCH with 42.60% non-invariance and the highest parameter was SOCIAICT with 87.20% non-invariance. The factor loadings were more mixed for non-invariant parameters, with five of nine being over the cutoff. The invariant factors were the loadings of ICTHOME (6.40%), HOMESCH (12.80%), ICTSCH (14.90%), and COMPICT (23.4%). The latent ICT subscales with the lowest amount of non-invariance would be the most useful when comparing countries. All the subscales have significant non-invariance in all 47 countries. Table 2 shows the R-squared values for each of the nine ICT scales. The R-squared value of ICTHOME was congruous with the non-invariance output as it was the highest at .94, which means that it is the most invariant parameter. According to the R-squared, COMPICT was the next-most invariant, which is similar to the non-invariance output. With this high amount of non-invariance, the latent mean estimates or their comparisons in other models cannot be trusted using this data. The factor loadings of only four scales (ICTHOME, HOMESCH, ICTSCH, and COMPICT) can be considered measurement invariant and none of the intercepts are measurement invariant. Therefore, a trustworthy comparison cannot be established between all included countries.

6.2. Mathematics and Science Alignment

The invariance output of the mathematics and science model had extremely low non-invariance in both factor loadings and intercepts. All plausible values included in the separate mathematics and science Alignment models were invariant. Mathematics intercepts and factor loadings had an average of 9.40% and 6.0%, respectively. Science was even lower with intercepts and factor loading percentages of .83% and 1.0%, respectively. Table 5 shows the percent non-invariance for intercepts and factor loadings for the ten plausible values for both the mathematics and science scales. These levels of non-invariance are well below the 25% cutoff, which suggests that the plausible values provided by PISA are invariant and can be used when comparing mathematics and science knowledge across countries. Consequently, the factor means are comparable across the 47 countries included in this study. According to both the present results and those from PISA 2015 results (OECD, 2016), Singapore is by far the leading country in mathematics and science proficiency. The comparisons of the country averages for mathematics and science scores match those from PISA identically. Table 3 and Table 4 show the ranked order of countries by factor means.

7. Discussion

7.1. ICT

Are the ICT, science, and mathematics variables measurement invariant across the countries participating in the PISA ICT questionnaire? As outlined in *Results*, ICT displayed a large amount of non-invariance according to the Alignment output. The factor loadings of ICTHOME, HOMESCH, ICTSCH, and COMPICT were the only ones that fell below the acceptable cutoff, which means that they would be the most trustworthy and useful in comparing means across all the participating countries. However, only the factor loadings of these four items are measurement invariant, meaning scalar invariance cannot be established and factor means cannot be compared (Millsap, 2012). The other five indicators are not invariant across the participating countries, which means that they cannot be compared as well. As measurement invariance is a prerequisite for valid comparisons of group differences in ICT scores, the current results indicate that, overall, ICT cannot be compared across the PISA countries that participated in the ICT questionnaire. This also shows that any cross-cultural educational research regarding ICT, especially revolving around policy, must be mindful of the disparities in ICT resources (e.g., infrastructure) and approaches across countries. The lack of evidence of measurement invariance makes it difficult to determine whether group differences reflect actual differences in the constructs measured rather than measurement error. Future research is needed to identify variables that may be possible sources of the measurement non-invariance results.

It is possible that ICT does not measure the same constructs in these vastly different countries. However, this does not mean that these countries cannot be compared by their ICT. As a group, the ICT profiles of these 47 countries are not similar enough to be compared or to trust that their scales are measuring the same constructs. Alternatively, if the countries were organized into subgroups based on their similar cultural involvement with technology, then the PISA representation of ICT may become measurement invariant. This could result in organizing countries based on their place on the Frontrunner-Challenger scale. If smaller groups of more similar countries are compared, then measurement invariance could be established.

Alternatively, this lack of ICT invariance could be a positive aspect, as these scales allow for national adaptions, introducing changes in the scales for countries. Therefore, the invariance may be due to the changes in the scales from different nationalities.

7.2. Mathematics and Science

The mathematics and science plausible values had very low levels of non-invariance across 47 countries. As a result, both mathematics plausible values and science plausible values can be reliably compared across countries. Therefore, when comparing the countries on a scale, based on the means of their plausible values, these results can provide the confidence that they represent the same knowledge constructs despite being measured in culturally different countries.

It is important to note the contrast in measurement invariance between the mathematics and science variables and the ICT scales. The mathematics and science questions and scales have received more attention since the inception of PISA as compared to the *ICT Familiarity Questionnaire* that was added in 2003 (OECD 2003, 2005, 2014a, 2017). The main difference, however, is that the ICT scale tests a construct that is more directly linked to culture than the more homogenous mathematics and science knowledge. As mentioned earlier, the mathematics and science scales have already been established as invariant by PISA. Therefore, this section of the analysis can be seen as testing the Alignment method rather than the invariance of the two scales. The Alignment method correctly showed measurement invariance in the already established measurement invariant mathematics and science PISA scales. Taken together, the present findings distinguish the ICT from the more established STEM subjects (mathematics and science) in terms of measurement invariance and require additional research to interpret the noninvariance of ICT across the countries participating in the ICT questionnaire.

7.3. Educational Implications

Overall, ICT was non-invariant. Thus, it appears that technology is treated differently depending on the country. This could be due to different cultural relationships with technology or nuanced differences in measuring these interactions with technology. Either way, ICT interactions cannot be broadly compared across large groups of countries. Researchers may need to test for measurement invariance in the countries of interest and for the ICT factors of interest before they are able to make cross country comparisons using PISA's ICT scale. One of the implications of this study is a more active role that formal education may consider taking to develop students' skills related to living in the 21st-century digital age to better prepare them to be contributing members to the information society. Conversely, mathematics and science scores are measurement invariant and can be compared across countries when measured by PISA. This implies that mathematics and science learning is either more universal than ICT, has been taught longer than ICT in formal settings, or more effort was put into the design of those measures.

7.4. Limitations

As the ICT scales are capped, some very highly-rated responses may be restricted to the limit of the scale, thereby causing a ceiling effect. The ICT scales could benefit by extending to higher levels of use and exposure for technologically-advanced countries. Certain aspects of the

ICT scales such as self-report format and translations into many languages may introduce reference bias between different cultural groups, which could in turn introduce measurement inconsistencies (Heckman & Kautz, 2013). A limitation of the Alignment method that was not encountered in this research is that it works well unless there are small group sizes or a high proportion of significant non-invariant parameters. The Alignment method has several disadvantages: cross loadings cannot be accommodated, models with covariates cannot be estimated, relationships between variables cannot be estimated, and there are no parameters to assess model fit (Asparouhov & Muthén, 2014; Marsh et al., 2018). These limitations restrict the Alignment method to an exploratory tool. The non-invariance cutoff point is 25% and the minimum number of groups is 30 (Muthén & Asparouhov, 2014). In addition, this method is relatively new and still requires further validation, such as simulation studies to test cutoff values. However, this method is still extremely useful for automating and simplifying the process of establishing measurement invariance. As described by Asparouhov and Muthén (2013), the Alignment method can estimate group-specific factor means and variances without exact measurement invariance. It is more appropriate for estimating models for many groups than the more common CFA, which struggles with scalar model fit and large modification indices. Using the Alignment method, researchers are able to minimize the amount of measurement non-invariance more than if they used a CFA (Muthén & Asparouhov 2013). Furthermore, in the mathematics and science models, only the countries who participated in the *ICT Familiarity Questionnaire* were included. This was done to keep similar samples for the models, however it would be beneficial to see if measurement invariance holds when the remaining countries are added. These limitations should not be overlooked when drawing conclusions, and future researchers should make their best efforts to address them to improve the research on this topic (Rutkowski & Rutkowski, 2016).

7.5. Future Directions

The next step in this research is to organize groups of countries with similar ICT profiles that will display measurement invariance for the PISA ICT subscales. Once this is achieved, the Alignment within CFA (AwC) method can be used to obtain estimated means and model estimates because comparing multiple countries simultaneously to assess the relation between ICT and academic achievement may present a methodological challenge. To address this challenge, more complex methods such as AwC are required to optimally perform this complex analysis (Marsh et al., 2018). In the future, the AwC method can be used to address this limitation, as it offers a confirmatory aspect to measurement invariance testing. The AwC approach can be used to extend the Alignment method used in SEM analyses into a confirmatory tool to address a multitude of issues such as covariates and latent variable relationship estimates, as it constitutes a combination between the Alignment and CFA methods (Marsh et al., 2018). As the Alignment method only deals with variance and means of factors, the AwC method allows for regression estimation of multiple groups. This method can test for measurement invariance across populations using the relaxed-fit style of the Alignment method and it can also provide model estimations. The Alignment method rather than AwC was used in this research because of the exploratory rather than confirmatory stance taken to investigate the measurement invariance of the scales rather than the factor structure. Furthermore, PISA 2018 data is now public and this analysis can be replicated to monitor any change in ICT variables, academic scores, and their associations. The current study used PISA 2015, as the analysis, results, and discussion had been completed before the release of the PISA 2018 data. Future studies will focus on research involving adolescents, exploring the relationship between ICT and academic achievement in more ecologically-valid settings. As addressed by Dong and Dumas (2020), measurement invariance is often established prior to testing a measure. However, due to the importance of measurement invariance in validly comparing groups, more focus should be placed on measurement invariance during the creation of measures.

8. Conclusion

The ICT means of all the countries that participated in the PISA *ICT Familiarity Questionnaire* cannot be reliably compared because it is possible that the ICT measure is not homogeneous for all countries. Given that measurement invariance of ICT was not established for the group of 47 countries, it would be valuable to know if comparisons of smaller, more similar groups are possible. This investigation could allow researchers to compare students' ICT in multiple different countries on a smaller scale using the available data. On the other hand, the mathematics and science plausible values were confirmed as measurement invariant across all countries. For instance, the quizzes used to collect the mathematics and science knowledge as well as the IRT methods to determine the plausible values were successful in creating the same scales across differing cultures. Therefore, the present work shows that the Alignment method is capable of accurately measuring and confirming the measurement invariance of the mathematics and science scales across countries.

Acknowledgements

We would like to thank the editor and the reviewers for their valuable feedback. We would also like to thank the Centre for Mathematics, Science, and Technology Education

(CMASTE) at the University of Alberta, the Social Sciences and Humanities Research Council of Canada - Insight Development Grant (SSHRC IDG) RES0034954 and Insight Grant (SSHRC IG) RES0048110, the Natural Sciences and Engineering Research Council (NSERC DG) RES0043209, and the Killam Cornerstone Operating Grant RES0043207 for supporting this research.

References

Amérigo, M., Garcia, J. A., Perez-Lopez, R., Cassullo, G., Ramos, A., Venumbaka, S. K., & Aragones, J. I. (2020). Analysis of the structure and factorial invariance of the Multidimensional Environmental Concern Scale (MECS). *Psicothema, 32*(2), 275-283.

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(4), 495-508.

Aryadoust, V. (2020). Measurement invariance across information and communication technology development index and gender: The case of the Pearson Test of English Academic reading. *Studies in Educational Evaluation, 64*, 100813.

Byrne, B. M., & van de Vijver, F. J. (2017). The maximum likelihood Alignment approach to testing for approximate measurement invariance: A paradigmatic cross-cultural application. *Psicothema, 29*(4), 539-551.

CDC (May, 2019). About CDC 24-7: Mission, role, and pledge. Centers for Disease Control and Prevention. Retrieved from https://www.cdc.gov/about/organization/mission.htm

Coromina, L., & Peral, E. B. (2020). Comparing Alignment and Multiple Group CFA for Analysing Political Trust in Europe During the Crisis. *Methodology, 16*(1), 21-40.

Crane, P. K., Belle, G. V., & Larson, E. B. (2004). Test bias in a cognitive test: differential item functioning in the CASI. *Statistics in Medicine, 23*(2), 241-256.

Daintith, J. (2009). "IT" *A Dictionary of Physics*. Oxford University Press.

Dneprovskaya, N. V., Bayaskalanova, T. A., Ruposov, V. L., & Shevtsova, I. V. (2018, September). Study of digitization of russian higher education as basis for smart education. In *2018 IEEE International Conference "Quality Management, Transport and Information Security, Information Technologies*"(IT&QM&IS) (pp. 607-611). IEEE.

Dong, Y., & Dumas, D. (2020). Are personality measures valid for different populations? A systematic review of measurement invariance across cultures, gender, and age. *Personality and Individual Differences, 160,* 109956.

European Union (March, 2019a). 2nd survey of schools: ICT in Education. Digital Single Market. Retrieved from https://ec.europa.eu/digital-single-market/en/news/2nd-survey-schools-ict-education

European Union (March, 2019b). 2nd survey of schools: ICT in Education. Bulgaria Country Report. Digital Single Market. DOI: 10.2759/83567.

European Union (March, 2019c). 2nd survey of schools: ICT in Education. Finland Country Report. Digital Single Market. DOI: 10.2759/364303.

Fischer, J., Praetorius, A. K., & Klieme, E. (2019). The impact of linguistic similarity on cross-cultural comparability of students' perceptions of teaching quality. *Educational Assessment, Evaluation and Accountability, 31*(2), 201-220.

Guenole, N., & Brown, A. (2014). The consequences of ignoring measurement invariance for path coefficients in structural equation models. *Frontiers in Psychology, 5,* 980.

He, J., Barrera-Pedemonte, F., & Buchholz, J. (2019). Cross-cultural comparability of noncognitive constructs in TIMSS and PISA. *Assessment in Education: Principles, Policy & Practice, 26*(4), 369-385.

Heckman, J. J., & Kautz, T. (2013). *Fostering and measuring skills: Interventions that improve character and cognition* (No. w19656). National Bureau of Economic Research.

Jerrim, J., Micklewright, J., Heine, J. H., Salzer, C., & McKeown, C. (2018). PISA 2015: how big is the 'mode effect' and what has been done about it? *Oxford Review of Education, 44*(4), 476-493.

Johnson, L., Becker, S., Estrada, V., Freeman, A., Kampylis, P., Vuorikari, R., & Punie, Y. (2014). *NMC Horizon Report Europe: 2014 Schools Edition*. The New Media Consortium.

Kalaycioglu, D. B. (2015). The Influence of Socioeconomic Status, Self-Efficacy, and Anxiety on Mathematics Achievement in England, Greece, Hong Kong, the Netherlands, Turkey, and the USA. *Educational Sciences: Theory and Practice, 15*(5), 1391-1401.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. New York, NY, USA: Guilford.

Lamm, R., Do, T., Rodriguez, M.C., Scales, P., & Roehlkepartain, E. (2019, April). Measurement invariance of an international developmental assets measure: Alignment of 29 countries. Paper presented at the annual meeting of the *National Council on Measurement in Education*, Toronto, Canada.

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B., & Dicke, T. (2018). What to do when scalar invariance fails: The extended Alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychological Methods*, *23*(3), 524.

McLarnon, M. J., & Romero, E. F. (2020). Cross-cultural equivalence of shortened versions of the Eysenck Personality Questionnaire: An application of the alignment method. *Personality and Individual Differences, 163,* 110074.

Meade, A. W., & Lautenschlager, G. J. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods, 7*, 361– 388.

Meiring, D., Van de Vijver, A. J. R., Rothmann, S., & Barrick, M. R. (2005). Construct, item and method bias of cognitive and personality tests in South Africa. *SA Journal of Industrial Psychology, 31*(1), 1-8.

Meng, L., Qiu, C., & Boyd-Wilson, B. (2018). Measurement invariance of the ICT engagement construct and its association with students' performance in China and Germany: Evidence from PISA 2015 data. *British Journal of Educational Technology.* 1-19. doi:10.1111/bjet.12729.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525-543.

Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. Routledge.

Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement, 29*(2), 133-161.

Murray, J. (2011, December). Modern Network Architecture: Cloud network architecture and ICT. *IT Knowledge Exchange*. Retrieved from https://itknowledgeexchange.techtarget.com/modern-network-architecture/cloud-network-architecture-and-ict/

Muthén, B., & Asparouhov, T. (2014). IRT studies of many groups: The Alignment method. *Frontiers in Psychology, 5*. doi:10.3389/ fpsyg.2014.00978

Muthén, B., & Asparouhov, T. (2017). Recent methods for the study of measurement invariance with many groups: Alignment and random effects. *Sociological Methods & Research. Advance Online Publication.* doi:10.1177/0049124117701488

Muthén, L. K., & Muthén, B. O. (1998-2017). MPlus User's Guide Eighth Edition. Los Angeles, CA: Muthén & Muthén.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. US Department of Education.

Novak, J., Purta, M., Marciniak, T., Ignatowicz, K., Rozenbaum, K., & Yearwood, K., (2018). *The rise of Digital Challengers: How digitization can become the next growth engine for Central and Eastern Europe*. McKinsey & Company.

Odell, B., Cutumisu, M., & Gierl, M. (2020). A scoping review of the relationship between students' ICT and performance in mathematics and science in the PISA data. *Social Psychology of Education, 23*. 1449-1481. DOI: 10.1007/s11218-020-09591-x

OECD (1999). *PISA 2000 Measuring Student Knowledge and Skills.* PISA, OECD Publishing.

OECD (2003). *The PISA 2003 Assessment Framework: Mathematics, Reading, Science, and Problem Solving Knowledge and Skills.* PISA, OECD Publishing.

OECD (2005). *PISA 2003 Technical Report*. PISA, OECD Publishing, Paris. Retrieved from http://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/35188570.pdf.

OECD (2006). *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006.* PISA, OECD Publishing, Paris, https://doi.org/10.1787/9789264026407-en.

OECD (2009). *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science.* PISA, OECD Publishing. ISBN: 9789264059603

OECD (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy.* PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264190511-en

OECD (2014a). *PISA 2015 ICT Familiarity Questionnaire*. OECD Publishing, Paris. Retrieved from: http://www.oecd.org/pisa/data/CY6_QST_MS_ICQ_Final.pdf.

OECD (2014b). *PISA 2015 Student Questionnaire*. PISA, OECD Publishing, Paris. Retrieved from http://www.oecd.org/pisa/data/CY6_QST_MS_STQ_CBA_Final.pdf

OECD (2015). *Students, Computers and Learning: Making the Connection*, PISA, OECD Publishing.

OECD (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*, PISA, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264266490-en

OECD (2017a). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving*, revised edition, *OECD Publishing, Paris.*

OECD (2017b). *PISA 2015 Technical Report*, PISA, OECD Publishing, Paris. Retrieved from https://www.oecd.org/pisa/sitedocument/PISA-2015-technical-report-final.pdf.

Raudenská, P. (2020). The cross-country and cross-time measurement invariance of positive and negative affect scales: Evidence from European Social Survey. *Social Science Research, 86*, 102369.

Ridao-Cano, C., & Bodewig, C. (2018). Growing united: Upgrading Europe's convergence machine. World Bank Report on the European Union.

Rutkowski, L., & Rutkowski, D. (2016). A call for a more measured approach to reporting and interpreting PISA results. *Educational Researcher, 45*(4), 252-257.

Schuler, M., Musekamp, G., Bengel, J., Nolte, S., Osborne, R. H., & Faller, H. (2014). Measurement invariance across chronic conditions: a systematic review and an empirical investigation of the Health Education Impact Questionnaire (heiQ™). *Health and Quality of Life Outcomes, 12*(1), 56.

Steenkamp, J. B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*(1), 78-90.

Suhr, D. D. (2006). Exploratory or confirmatory factor analysis? 31st Annual SAS Users Group International Conference: Statistics and Data Analysis.

Ullman, J. B., & Bentler, P. M. (2003). Structural equation modeling. *Handbook of psychology*, 607-634.

UNEP (2019). Programme report 2018. United Nations Environment Programme. Retrieved from www.unenvironment.org/annualreport/2018/index.php

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.

Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis of comparative research*. Allyn & Bacon.

Van Der Linden, S., Maibach, E., Cook, J., Leiserowitz, A., & Lewandowsky, S. (2017). Inoculating against misinformation. *Science, 358*(6367), 1141-1142.

Zhang, D., & Liu, L. (2016). How does ICT use influence students' achievements in math and science over time? Evidence from PISA 2000 to 2012. *Eurasia Journal of Mathematics, Science & Technology Education*, *12*(9).

Tables

Table 1

List of the 47 Countries in the Analysis, with Latent Class Label, Country ID, and Three-Letter Code. Note Germany (276) was removed from the models.

Latent Class	CNTRID	Country	Latent Class	CNTRID	Country
1	36	AUS	25	410	KOR
2	40	AUT	26	428	LVA
3	56	BEL	27	440	LTU
4	76	BRA	28	442	LUX
5	100	BGR	29	446	MAC
6	152	CHL	30	484	MEX
7	158	TAP	31	528	NLD
8	170	COL	32	554	NZL
9	188	CRI	33	604	PER
10	191	HRV	34	616	POL
11	203	CZE	35	620	PRT
12	208	DNK	36	643	RUS
13	214	DOM	37	702	SGP
14	233	EST	38	703	SVK
15	246	FIN	39	705	SVN
16	250	FRA	40	724	ESP
17	300	GRC	41	752	SWE
18	344	HKG	42	756	CHE
19	348	HUN	43	764	THA
20	352	ISL	44	826	GBR
21	372	IRL	45	858	URY
22	376	ISR	46	970	QCH
23	380	ITA	47	971	QES
24	392	JPN			

Table 2

R-squared and Non-Invariance Percentage Values for the 9 ICT Scales from the Alignment Method.

	R-Squared	% non-invariant factor loadings	% non-invariant intercept
ENTUSE	.10	25.5	66.0
SOIAICT	.16	51.1	87.2
AUTICT	.25	46.8	70.2
ICTSCH	.30	14.9	42.6
USESCH	.67	87.2	85.1
INTICT	.76	44.7	66.0
HOMESCH	.78	12.8	72.3
COMPICT	.83	23.4	85.1
ICTHOME	.94	6.4	42.6

Note: *R-Squared* values closer to 1 indicate more invariance, while a value close to 0 indicates less invariance for that parameter. A higher percentage indicates more non-invariance for that parameter. The ICT scales are arranged from smallest to largest *R-Squared* value.

Table 3

Mean Comparison of the MATH Latent Variable for 47 Included Countries from the Alignment Model.

Rank	Country	Factor Means	Rank	Country	Factor Means	Rank	Country	Factor Means
1	SGP	1.23	17	AUT	0.40	33	HUN	0.16
2	HKG	1.03	18	NZL	0.38	34	SVK	0.14
3	MAC	0.98	19	RUS	0.37	35	ISR	0.07
4	TAP	0.96	20	SWE	0.37	36	HRV	0.00
5	JPN	0.84	21	AUS	0.37	37	GRC	-0.13
6	QCH	0.82	22	FRA	0.35	38	BGR	-0.28
7	KOR	0.74	23	GBR	0.35	39	CHL	-0.51
8	CHE	0.70	24	CZE	0.35	40	URY	-0.56
9	EST	0.68	25	PRT	0.34	41	THA	-0.60
10	NLD	0.59	26	ITA	0.31	42	MEX	-0.69
11	DNK	0.58	27	ISL	0.29	43	CRI	-0.78
12	FIN	0.58	28	QES	0.27	44	COL	-0.91
13	SVN	0.56	29	ESP	0.27	45	PER	-0.95
14	BEL	0.53	30	LUX	0.27	46	BRA	-1.07
15	POL	0.50	31	LVA	0.22	47	DOM	-1.67
16	IRL	0.49	32	LTU	0.18			

Table 4

Mean Comparison of the SCIENCE Latent Variable for 47 Included Countries from the Alignment Model.

Rank	Country	Factor Means	Rank	Country	Factor Means	Rank	Country	Factor Means
1	SGP	0.63	17	BEL	-0.01	33	LTU	-0.32
2	JPN	0.42	18	DNK	-0.01	34	ISL	-0.35
3	EST	0.37	19	POL	-0.01	35	ISR	-0.42
4	TAP	0.35	20	PRT	-0.02	36	SVK	-0.49
5	FIN	0.33	21	AUT	-0.09	37	GRC	-0.56
6	MAC	0.31	22	FRA	-0.09	38	CHL	-0.65
7	HKG	0.25	23	QES	-0.09	39	BGR	-0.67
8	QCH	0.18	24	SWE	-0.11	40	URY	-0.79
9	KOR	0.16	25	CZE	-0.11	41	THA	-0.96
10	NZL	0.13	26	ESP	-0.11	42	CRI	-0.98
11	SVN	0.12	27	LVA	-0.15	43	COL	-1.02
12	AUS	0.09	28	RUS	-0.19	44	MEX	-1.02
13	GBR	0.08	29	LUX	-0.23	45	BRA	-1.20
14	NLD	0.07	30	ITA	-0.26	46	PER	-1.25
15	CHE	0.04	31	HUN	-0.30	47	DOM	-2.01
16	IRL	0.00	32	HRV	-0.32			

Table 5

Non-Invariance Percentage Values for the 10 Plausible Values from the Mathematics and Science Scales from the Alignment Method.

	Mathematics		Science	
	% non-invariant factor loadings	% non-invariant intercept	% non-invariant factor loadings	% non-invariant intercept
PV1	12.5	4.2	0.0	2.1
PV2	2.1	10.4	0.0	2.1
PV3	4.2	4.2	0.0	2.1
PV4	6.25	8.3	2.1	0.0
PV5	4.2	4.2	2.1	0.0
PV6	12.5	12.5	0.0	0.0
PV7	6.25	8.3	0.0	2.1
PV8	4.2	14.6	6.3	0.0
PV9	4.2	20.8	0.0	0.0
PV10	6.25	6.25	0.0	0.0

Appendix A

ICT questions taken from 2015 PISA *ICT Familiarity Questionnaire* which are used to determine the predictor variables (OECD, 2017b).

ICTHOME: ICT available at home

IC001 Are any of these devices available for you to use at home?

(Please select one response in each row.)

		Yes, and I use it	*Yes, but I don't use it*	*No*
IC001Q01TA	Desktop computer	□1	□2	□3
IC001Q02TA	Portable laptop, or notebook	□1	□2	□3
IC001Q03TA	<Tablet computer> (e.g. <iPad®>, <BlackBerry® PlayBook™>)	□1	□2	□3
IC001Q04TA	Internet connection	□1	□2	□3
IC001Q05TA	<Video games console>, e.g. <Sony® PlayStation®>	□1	□2	□3
IC001Q06TA	<Cell phone> (without Internet access)	□1	□2	□3
IC001Q07TA	<Cell phone> (with Internet access)	□1	□2	□3
IC001Q08TA	Portable music player (Mp3/Mp4 player, iPod® or similar)	□1	□2	□3
IC001Q09TA	Printer	□1	□2	□3
IC001Q10TA	USB (memory) stick	□1	□2	□3
IC001Q11TA	<ebook reader>, e.g. <Amazon® Kindle™>	□1	□2	□3

COMPICT: Students' perceived ICT competence

IC014 Thinking about your experience with digital media and digital devices: To what extent do you disagree or agree with the following statements?

(Please think of different kinds of digital devices such as for example desktop computers, portable laptops, notebooks, smartphones, tablet computers, cell phones without internet access, game consoles, or internet-connected television)

(Please select one response in each row.)

		Strongly disagree	*Disagree*	*Agree*	*Strongly agree*
IC014Q03NA	I feel comfortable using digital devices that I am less familiar with.	☐1	☐2	☐3	☐4
IC014Q04NA	If my friends and relatives want to buy new digital devices or applications, I can give them advice.	☐1	☐2	☐3	☐4
IC014Q06NA	I feel comfortable using my digital devices at home.	☐1	☐2	☐3	☐4
IC014Q08NA	When I come across problems with digital devices, I think I can solve them.	☐1	☐2	☐3	☐4
IC014Q09NA	If my friends and relatives have a problem with digital devices, I can help them.	☐1	☐2	☐3	☐4

ENTUSE: ICT use outside of school for leisure

IC008 How often do you use digital devices for the following activities outside of school?

(Please select one response in each row.)

		Never or hardly ever	*Once or twice a month*	*Once or twice a week*	*Almost every day*	*Every day*
IC008Q01TA	Playing one-player games.	☐1	☐2	☐3	☐4	☐5
IC008Q02TA	Playing collaborative online games.	☐1	☐2	☐3	☐4	☐5
IC008Q03TA	Using email.	☐1	☐2	☐3	☐4	☐5
IC008Q04TA	<Chatting online> (e.g. <MSN®>).	☐1	☐2	☐3	☐4	☐5
IC008Q05TA	Participating in social networks (e.g. <Facebook>, <MySpace>).	☐1	☐2	☐3	☐4	☐5
IC008Q07NA	Playing online games via social networks (e.g. <Farmville®>, <The Sims Social>).	☐1	☐2	☐3	☐4	☐5
IC008Q08TA	Browsing the Internet for fun (such as watching videos, e.g. <YouTube™>).	☐1	☐2	☐3	☐4	☐5
IC008Q09TA	Reading news on the Internet (e.g. current affairs).	☐1	☐2	☐3	☐4	☐5
IC008Q10TA	Obtaining practical information from the Internet (e.g. locations, dates of events).	☐1	☐2	☐3	☐4	☐5
IC008Q11TA	Downloading music, films, games or software from the internet.	☐1	☐2	☐3	☐4	☐5
IC008Q12TA	Uploading your own created contents for sharing (e.g. music, poetry, videos, computer programs).	☐1	☐2	☐3	☐4	☐5
IC008Q13NA	Downloading new apps on a mobile device.	☐1	☐2	☐3	☐4	☐5

ICTSCH: ICT available at school

IC009 Are any of these devices available for you to use at school?

(Please select one response in each row.)

		Yes, and I use it	*Yes, but I don't use it*	*No*
IC009Q01TA	Desktop computer	☐1	☐2	☐3
IC009Q02TA	Portable laptop or notebook	☐1	☐2	☐3
IC009Q03TA	<Tablet computer> (e.g. <iPad®>, <BlackBerry® PlayBook™>)	☐1	☐2	☐3
IC009Q05NA	Internet-connected school computers	☐1	☐2	☐3
IC009Q06NA	Internet connection via wireless network	☐1	☐2	☐3
IC009Q07NA	Storage space for school-related data, e.g. a folder for own documents	☐1	☐2	☐3
IC009Q08TA	USB (memory) stick	☐1	☐2	☐3
IC009Q09TA	<ebook reader>, e.g. <Amazon® Kindle™>	☐1	☐2	☐3
IC009Q10NA	Data projector, e.g. for slide presentations	☐1	☐2	☐3
IC009Q11NA	Interactive whiteboard, e.g. <SmartBoard®>	☐1	☐2	☐3

USESCH: Use of ICT at school in general

IC011 **How often do you use digital devices for the following activities <u>at school</u>?**

(Please select one response in each row.)

		Never or hardly ever	*Once or twice a month*	*Once or twice a week*	*Almost every day*	*Every day*
IC011Q01TA	<Chatting online> at school.	☐1	☐2	☐3	☐4	☐5
IC011Q02TA	Using email at school.	☐1	☐2	☐3	☐4	☐5
IC011Q03TA	Browsing the Internet for schoolwork.	☐1	☐2	☐3	☐4	☐5
IC011Q04TA	Downloading, uploading or browsing material from the school's website (e.g. <intranet>).	☐1	☐2	☐3	☐4	☐5
IC011Q05TA	Posting my work on the school's website.	☐1	☐2	☐3	☐4	☐5
IC011Q06TA	Playing simulations at school.	☐1	☐2	☐3	☐4	☐5
IC011Q07TA	Practicing and drilling, such as for foreign language learning or mathematics.	☐1	☐2	☐3	☐4	☐5
IC011Q08TA	Doing homework on a school computer.	☐1	☐2	☐3	☐4	☐5
IC011Q09TA	Using school computers for group work and communication with other	☐1	☐2	☐3	☐4	☐5

INTICT: Students' ICT interest

IC013 Thinking about your experience with digital media and digital devices: to what extent do you disagree or agree with the following statements?

(Please think of different kinds of digital devices such as for example desktop computers, portable laptops, notebooks, smartphones, tablet computers, cell phones without internet access, game consoles, or internet-connected television)

(Please select one response in each row.)

		Strongly disagree	*Disagree*	*Agree*	*Strongly agree*
IC013Q01NA	I forget about time when I'm using digital devices.	☐1	☐2	☐3	☐4
IC013Q04NA	The Internet is a great resource for obtaining information I am interested in (e.g. news, sports, dictionary).	☐1	☐2	☐3	☐4
IC013Q05NA	It is very useful to have social networks on the Internet.	☐1	☐2	☐3	☐4
IC013Q11NA	I am really excited discovering new digital devices or applications.	☐1	☐2	☐3	☐4
IC013Q12NA	I really feel bad if no internet connection is possible.	☐1	☐2	☐3	☐4
IC013Q13NA	I like using digital devices.	☐1	☐2	☐3	☐4

SOIAICT: Students' ICT as a topic in conversations

IC016 Thinking about your experience with digital media and digital devices: to what extent do you disagree or agree with the following statements?

(Please select one response in each row.)

		Strongly disagree	*Disagree*	*Agree*	*Strongly agree*
IC016Q01NA	To learn something new about digital devices, I like to talk about them with my friends.	☐1	☐2	☐3	☐4
IC016Q02NA	I like to exchange solutions to problems with digital devices with others on the internet.	☐1	☐2	☐3	☐4
IC016Q04NA	I like to meet friends and play computer and video games with them.	☐1	☐2	☐3	☐4
IC016Q05NA	I like to share information about digital devices with my friends.	☐1	☐2	☐3	☐4
IC016Q07NA	I learn a lot about digital media by discussing with my friends and relatives.	☐1	☐2	☐3	☐4

AUTICT: Students' perceived social interaction and autonomy related to ICT use

IC015 Thinking about your experience with digital media and digital devices: to what extent do you disagree or agree with the following statements?

(Please select one response in each row.)

		Strongly disagree	*Disagree*	*Agree*	*Strongly agree*
IC015Q02NA	If I need new software, I install it by myself.	☐1	☐2	☐3	☐4
IC015Q03NA	I read information about digital devices to be independent.	☐1	☐2	☐3	☐4
IC015Q05NA	I use digital devices as I want to use them.	☐1	☐2	☐3	☐4
IC015Q07NA	If I have a problem with digital devices I start to solve it on my own.	☐1	☐2	☐3	☐4
IC015Q09NA	If I need a new application, I choose it by myself.	☐1	☐2	☐3	☐4

HOMESCH: ICT use outside of school for schoolwork

IC010 How often do you use digital devices for the following activities outside of school?

(Please select one response in each row.)

		Never or hardly ever	*Once or twice a month*	*Once or twice a week*	*Almost every day*	*Every day*
IC010Q01TA	Browsing the Internet for schoolwork (e.g. for preparing an essay or presentation).	☐1	☐2	☐3	☐4	☐5
IC010Q02NA	Browsing the Internet to follow up lessons, e.g. for finding explanations.	☐1	☐2	☐3	☐4	☐5
IC010Q03TA	Using email for communication with other students about schoolwork.	☐1	☐2	☐3	☐4	☐5
IC010Q04TA	Using email for communication with teachers and submission of homework or other schoolwork.	☐1	☐2	☐3	☐4	☐5
IC010Q05NA	Using social networks for communication with other students about schoolwork (e.g. <Facebook>, <MySpace>).	☐1	☐2	☐3	☐4	☐5
IC010Q06NA	Using social networks for Communication with teachers (e.g. <Facebook>, <MySpace>).	☐1	☐2	☐3	☐4	☐5
IC010Q07TA	Downloading, uploading or browsing material from my school's website (e.g. timetable or course materials).	☐1	☐2	☐3	☐4	☐5
IC010Q08TA	Checking the school's website for announcements, e.g. absence of teachers.	☐1	☐2	☐3	☐4	☐5
IC010Q09NA	Doing homework on a computer.	☐1	☐2	☐3	☐4	☐5
IC010Q10NA	Doing homework on a mobile device.	☐1	☐2	☐3	☐4	☐5
IC010Q11NA	Downloading learning apps on a mobile device.	☐1	☐2	☐3	☐4	☐5
IC010Q12NA	Downloading science learning apps on a mobile device.	☐1	☐2	☐3	☐4	☐5

Appendix B

MPlus 8 analysis code for ICT Alignment.

```
TITLE: ICT model Alignment_Fixed_3fac;

DATA: FILE IS PISA data Current_July22.csv;

VARIABLE:
    NAMES ARE
    CNTRYID CNTSCHID CNTSTUID NatCen Region ESCS GENDER
    HOMESCH ENTUSE USESCH INTICT COMPICT
   AUTICT SOIAICT ICTHOME ICTSCH
   W_FSTUWT
   PV1MATH PV2MATH PV3MATH PV4MATH PV5MATH
   PV6MATH PV7MATH PV8MATH PV9MATH PV10MATH
   PV1SCIE PV2SCIE PV3SCIE PV4SCIE PV5SCIE
   PV6SCIE PV7SCIE PV8SCIE PV9SCIE PV10SCIE;

    USEVARIABLES ARE
    HOMESCH ENTUSE USESCH INTICT COMPICT
    AUTICT SOIAICT ICTHOME ICTSCH;

   MISSING ARE ALL(99 97);
    WEIGHT IS W_FSTUWT;

    Classes= c(47);
    knownclass= c(CNTRYID= 36      40 56  76     100    152    158    170    188
    191    203    208
    214    233    246    250    300    344    348    352    372    376    380    392
    410    428    440    442    446
    484    528    554    604    616    620    643    702    703    705    724    752
    756    764    826    858    970    971);

  ANALYSIS:
   TYPE= Mixture;
    ESTIMATOR = MLR;
     AITERATIONS = 5000;
     PROCESSORS=8;
     ASTARTS = 60
     ACONVERGENCE=.001;
     alignment=fixed (528); !Netherlands

    MODEL:

    %OVERALL%

      [ICTUSE];
    ICTUSE BY HOMESCH ENTUSE USESCH;
```

```
        [ICTCOMF];
   ICTCOMF BY INTICT COMPICT AUTICT SOIAICT;
        [ICTAVB];
   ICTAVB BY ICTHOME ICTSCH;


OUTPUT:
SVALUES;
TECH1;
TECH8;
align;
```

Appendix C

MPlus 8 analysis code for Mathematics Alignment.

```
TITLE: MATH Alignment;
DATA: FILE IS PISA_July22.csv;

  VARIABLE:
     NAMES ARE
     CNTRYID CNTSCHID CNTSTUID NatCen Region ESCS GENDER
     HOMESCH ENTUSE USESCH INTICT COMPICT AUTICT SOIAICT ICTHOME
      ICTSCH
     W_FSTUWT
    PV1MATH PV2MATH PV3MATH PV4MATH PV5MATH PV6MATH PV7MATH PV8MATH
    PV9MATH PV10MATH
    PV1SCIE PV2SCIE PV3SCIE PV4SCIE PV5SCIE PV6SCIE PV7SCIE PV8SCIE
    PV9SCIE PV10SCIE;

     USEVARIABLES ARE
    PV1MATH PV2MATH PV3MATH PV4MATH PV5MATH
    PV6MATH PV7MATH PV8MATH PV9MATH PV10MATH;

    MISSING ARE ALL(99 97);
     WEIGHT IS W_FSTUWT;

     Classes= c(47);
     knownclass= c(CNTRYID= 36      40 56  76    100   152   158   170   188
     191   203   208
     214   233   246   250   300   344   348   352   372   376   380   392
     410   428   440   442   446
     484   528   554   604   616   620   643   702   703   705   724   752
     756   764   826   858   970   971);

   DEFINE:
    PV1MATH = PV1MATH/100;
    PV2MATH = PV2MATH/100;
    PV3MATH = PV3MATH/100;
    PV4MATH = PV4MATH/100;
    PV5MATH = PV5MATH/100;
    PV6MATH = PV6MATH/100;
    PV7MATH = PV7MATH/100;
    PV8MATH = PV8MATH/100;
    PV9MATH = PV9MATH/100;
    PV10MATH = PV10MATH/100;

   ANALYSIS:
    TYPE= Mixture;
     ESTIMATOR = MLR;
      PROCESSORS=6;
      ACONVERGENCE=.001;
     alignment=fixed(191);
```

```
 MODEL:
 %OVERALL%
   [MATH];
 MATH BY PV1MATH
         PV2MATH
         PV3MATH
         PV4MATH
         PV5MATH
         PV6MATH
         PV7MATH
         PV8MATH
         PV9MATH
         PV10MATH

OUTPUT:
TECH1;
TECH8;
align;
```

Appendix D

MPlus 8 analysis code for Science Alignment.

```
TITLE: SCIENCE Alignment;

DATA: FILE IS PISA_July22.csv;

VARIABLE:
     NAMES ARE
      CNTRYID CNTSCHID CNTSTUID NatCen Region ESCS GENDER
      HOMESCH ENTUSE USESCH INTICT COMPICT AUTICT SOIAICT ICTHOME
      ICTSCH
      W_FSTUWT
      PV1MATH PV2MATH PV3MATH PV4MATH PV5MATH PV6MATH PV7MATH
      PV8MATH PV9MATH PV10MATH
      PV1SCIE PV2SCIE PV3SCIE PV4SCIE PV5SCIE PV6SCIE PV7SCIE PV8SCIE
      PV9SCIE PV10SCIE;

     USEVARIABLES ARE
     PV1SCIE PV2SCIE PV3SCIE PV4SCIE PV5SCIE
     PV6SCIE PV7SCIE PV8SCIE PV9SCIE PV10SCIE ;

  MISSING ARE ALL(99 97);
     WEIGHT IS W_FSTUWT;


     Classes= c(47);
     knownclass= c(CNTRYID= 36     40 56  76     100    152    158    170    188
     191    203    208
     214    233    246    250    300    344    348    352    372    376    380    392
     410    428    440    442    446
     484    528    554    604    616    620    643    702    703    705    724    752
     756    764    826    858    970    971);

 DEFINE:

     PV1SCIE = PV1SCIE /100;
     PV2SCIE = PV2SCIE /100;
     PV3SCIE = PV3SCIE /100;
     PV4SCIE = PV4SCIE /100;
     PV5SCIE = PV5SCIE /100;
     PV6SCIE = PV6SCIE /100;
     PV7SCIE = PV7SCIE /100;
     PV8SCIE = PV8SCIE /100;
     PV9SCIE = PV9SCIE /100;
     PV10SCIE = PV10SCIE /100;

 ANALYSIS:
  TYPE= Mixture;
     ESTIMATOR = MLR;
```

```
      PROCESSORS=6;
      ACONVERGENCE=.001;
      alignment=fixed(372);

   MODEL:

      %OVERALL%

     [SCIENCE];
   SCIENCE BY PV1SCIE
              PV2SCIE
              PV3SCIE
              PV4SCIE
              PV5SCIE
              PV6SCIE
              PV7SCIE
              PV8SCIE
              PV9SCIE
              PV10SCIE;

OUTPUT:
TECH1;
TECH8;
align;
```