[PISA 2018 Results Volume VI]

[Are Students Ready to Thrive in an Interconnected World?]

OECD

ANNEX A5. Measurement invariance and comparability of scaled indices across countries

Measurement invariance

This annex explores the comparability of PISA 2018 global competence scaled indices presented in Annex A1 across the participating countries and economies. The statistical property used to assess comparability across countries is called measurement invariance. It indicates that the same construct/index is being measured and that it can be interpreted in a similar manner across countries/economies.

Lack of comparability of an index could arise from:

- underlying differences (cultural and otherwise) between countries/economies (in this case, the same question might have different meanings for the different groups)
- variations in translation of the questionnaire items across different languages
- adaptations of questionnaire items to a country/economy context in a way that could make them less comparable across countries/economies
- use of previously validated scales on populations that are different from the ones for which the scales were validated.

Results of the measurement invariance analyses for each index/construct and for each country/economy and questionnaire item are provided in the corresponding Excel file (Annex A5). In addition to PISA's standard approach to examine the comparability of scaled indices which is described at the end of this annex, this report applied the alignment method (Asparouhov and Muthén, 2014[1]). The alignment method can be used to estimate group-specific factor means and variances under configural invariance. A strength of the method is the ability to estimate models for many groups. The method is a valuable alternative to multiple-group confirmatory factor analysis for detecting non-equivalent parameters.

Three measures were computed to assess the comparability of indices:

1. The **invariance index** (factor score) indicates whether an index/construct is comparable overall across countries/economies (Tables VI.A5.1, VI.A5.3, VI.A5.5, and VI.A5.7, available online). As a rule of thumb, when the invariance index exceeds 0.75, the construct is considered as equivalent and comparable across countries/economics regardless of the values of the invariance index for items. However, some caution might still be needed when interpreting results for countries/economies with a low proportion of invariant parameters. This is further explored below.

2. The **invariance index for items** indicates if an item is comparable across countries/economies (Tables VI.A5.1, VI.A5.3, VI.A5.5, and VI.A5.7, available online). For the purpose of this report, an item that exceeds the threshold of 0.6 is considered comparable. Items that do not exceed this threshold could be excluded to improve comparability of the index/construct. However, the exclusion of items weakens the conceptual validity of the index, as it narrows its meaning. Thus, one should balance the need for comparability with that of conceptual validity of an index.

3. The **proportion of invariant parameters** are country-level analyses to check comparability for each country/economy (Tables VI.A5.2, VI.A5.4, VI.A5.6, and VI.A5.8, available online). The threshold for the proportion of invariant parameters is set at 0.25.

The invariance index represents levels of comparability across countries. This index indicates the proportion of variation across countries in the configural model that can be explained by variation in the factor means and variances, which is similar to an R-squared for regression analyses: it shows model fit but does not indicate which observation diverges from the regression line. As such, the invariance index does not show whether each country/economy is comparable to all others. Thus, country-level analyses are needed to check comparability for each country/economy. This is done by computing the **proportion of invariant parameters**. For instance, if five questionnaire items are used to construct an index, there

will be 10 parameters (i.e. factor loadings and intercepts for five items). The alignment method provides information on which parameter is invariant for a country/economy. Thus, the proportion of invariant parameters for each country/economy can be computed. The threshold for the proportion of invariant parameters is set at 0.25. It is not necessary for the threshold to be as high as that for the invariance index (0.75) because each country does not have to follow the same strict criteria. The results of the measurement invariance analyses for each index are presented below.

Student-level scaled indices

Of the ten student-level scaled indices (i.e. students' attitudes and dispositions) used in this report, six had an invariance index exceeding the threshold of 0.75 and can thus be considered comparable. These are the indices of students' awareness of global issues, students' attitudes towards immigrants, students' interest in learning about other cultures, students' respect for people from other cultures, students' cognitive adaptability and students' awareness of intercultural communication.

Three indices came close to attaining the threshold: students' self-efficacy regarding global issues, students' perspective taking and students' perception of discrimination at school. The only index that fell short, with an invariance index of 0.69, was the index of students' agency regarding global issues. For this index, caution is advised when interpreting the results.

Students' awareness of global issues (GCAWARE)

In general, this construct had an invariance index above 0.75 and can be considered as comparable across countries and economies. However, among the seven questionnaire items used in the scaling of this index, only one had an item invariance index that fell slightly short of the threshold of 0.6. This item asked students how informed they are about climate change and global warming. Moreover, five countries/economies had a proportion of invariant parameters below 0.25: Baku (Azerbaijan), Brunei Darussalam, Korea, Singapore and Spain. Caution is advised when interpreting findings based on this index for those countries/economies.

Students' self-efficacy regarding global issues (GCSELFEFF)

This construct fell slightly short of the invariance index threshold of 0.75. Among the six questionnaire items used in the scaling of this index, two had an item invariance index that fell slightly below the threshold of 0.6. Those items focused on students' ability to perform two tasks: explain how carbon-dioxide emissions affect global climate change and discuss the different reasons why people become refugees. Moreover, four countries had a proportion of invariant parameters below 0.25: Australia, Canada, New Zealand and Singapore.

Students' ability to understand the perspectives of others (PERSPECT)

This construct fell slightly short of the invariance index threshold of 0.75. Of the five items used in scaling this index, one had an item invariance index below 0.6. This item asked students how well the following statement describes them: "When I'm upset at someone, I try to take the perspective of that person for a while". Only two countries, Portugal and Chinese Taipei, had a proportion of invariant parameters below 0.25. Those countries/economies may be less comparable with others, and caution is advised when interpreting results based on this index for them.

Students' interest in learning about other cultures (INTCULT)

In general, this construct had an invariance index above 0.75 and can be considered as comparable across countries and economies. Of the four questionnaire items used in scaling this index, one had an

item invariance index below 0.6. This item asked students how well the following statement describes them: "I want to learn more about the religions of the world". This finding shows that the notion of religion might be understood differently by students in different countries/economies. No country or economy had a proportion of invariant parameters that fell below the threshold of 0.25. Therefore, results based on this index can be seen as comparable across all countries/economies.

Students' respect for people from other cultures (RESPECT)

In general, this construct had an invariance index above 0.75 and can be considered as comparable across countries and economies. All questionnaire items used for the scaling of the index of students' respect for people from other cultures had an item invariance index exceeding the cut-off of 0.6. Only three countries/economies had a proportion of invariant parameters below the threshold of 0.25: Hong Kong (China), Ireland and New Zealand.

Students' cognitive adaptability (COGFLEX)

In general, this construct had an invariance index above 0.75 and can be considered as comparable across countries and economies. All questionnaire items used for the scaling of this index had an item invariance index exceeding the cut-off point of 0.6 and, in all countries and economies, the proportion of invariant parameters exceeded 0.25. Thus, this index can be considered as fully comparable across all countries/economies.

Students' attitudes towards immigrants (ATTIMM)

In general, this construct had an invariance index above 0.75 and can be considered as comparable across countries and economies. All questionnaire items used for the scaling of students' attitudes towards immigrants had an item invariance index exceeding the cut-off point of 0.6. Only in Korea did the proportion of invariant parameters fall short of the threshold of 0.25. For Korea, caution is advised when comparing results on this index to other countries/economies.

Students' awareness of intercultural communication (AWACOM)

In general, this construct had an invariance index above 0.75 and can be considered as comparable across countries and economies. All questionnaire items used for the scaling of this index had an item invariance index exceeding the cut-off point of 0.6 and, in all countries and economies, the proportion of invariant parameters exceeded 0.25. Thus this index can be considered as fully comparable across all countries/economies.

Students' agency regarding global issues (GLOBMIND)

This construct had the lowest invariance index of 0.69. Of the six statements used to scale this index, two had an item invariance index below 0.6: "I think of myself as a citizen of the world" and "It is right to boycott companies that are known to provide poor workplace conditions for their employees". Three countries (Hungary, Spain and Romania) had a proportion of invariant parameters below 0.25. Results based on this index for those countries should be interpreted with caution.

Students' perception of discrimination at school (DISCRIMA)

This construct fell slightly short of the invariance index threshold of 0.75. Of the four questions used in the scaling of the index of students' perception of discrimination at school, one asking students whether teachers have misconceptions about the history of some cultural groups had an item invariance

index below 0.6. Only two countries, Thailand and Greece, had a proportion of invariant parameters below 0.25.

In general, most students' scaled indices are comparable across countries/economies, with few exceptions. Caution is advised when interpreting the results for countries with a proportion of invariant parameters below the cut-off of 0.25.

School-level scaled indices

Principals' views on teachers' multicultural beliefs (SCMCEG)

The construct of principals' views on teachers' multicultural beliefs had an invariance index of 0.81, exceeding the cut-off of 0.75. It can thus be considered as comparable across countries and economies. The proportion of invariant parameters was high for all countries, exceeding the cut-off of 0.25. Thus this index can be deemed comparable across all countries/economies.

Teacher-level scaled indices

Of the four scaled indices based on teacher reports, three had an invariance index exceeding the cut-off of 0.75 and thus can be considered as comparable across countries/economies. The only index that did not reach this threshold was the index of teacher training on global competence (invariance index = 0.69). Caution should be considered when interpreting results based on this index.

Teacher training on global competence (GCTRAIN)

This construct had the lowest invariance index of 0.69. Of the five questionnaire items used to scale this index, the one asking teachers whether they received training on conflict resolution strategies had an item invariance index below the cut-off of 0.6. In all countries and economies, the proportion of invariant parameters exceeded 0.25. Thus, the relatively low comparability of this index between countries/economies is due to one questionnaire item, rather than to underlying differences between countries or economies.

Teachers' multicultural and egalitarian beliefs (TCMCEG)

In general, this construct had an invariance index above 0.75 and can be considered as comparable across countries and economies. All questionnaire items used for the scaling of this index had an item invariance index exceeding the cut-off point of 0.6 and, in all countries and economies, the proportion of invariant parameters exceeded 0.25. Thus this index can be considered as fully comparable across all countries/economies.

Teachers' self-efficacy in multicultural environments (GCSELF)

In general, this construct had an invariance index above 0.75 and can be considered as comparable across countries and economies. All questionnaire items used for the scaling of this index had an item invariance index exceeding the cut-off point of 0.6. However, one country (Spain) had a proportion of invariant parameters below 0.25, making it less comparable with other countries/economies on this index.

Teachers' attitudes towards immigrants (TCATTIMM)

In general, this construct had an invariance index above 0.75 and can be considered as comparable across countries and economies. One questionnaire item out of four had an item invariance index below 0.6. This item asked teachers whether immigrant children should have the same opportunities for

education that other children in the country have. Moreover, Spain was the only country with a proportion of invariant parameters below 0.25.

In general, three out of four scaled indices based on teachers' reported data are comparable between countries and economies. Caution is recommended when interpreting the results on teacher self-efficacy regarding global issues and teachers' attitudes towards immigrants in Spain.

Parent-level scaled indices

Of the three scaled indices based on parents' reports, only one (index of parents' awareness of global issues) had an invariance index exceeding the cut-off of 0.75 and thus can be considered as comparable across countries/economies. The other two indices of parents' interest in learning about other cultures (invariance index = 0.69) and parents' attitudes towards immigrants (invariance index = 0.72) had invariance indices below the cut-off. Caution is advised when interpreting the results for these two indices.

Parents' awareness of global issues (GCAWAREP)

In general, this construct had an invariance index above 0.75 and can be considered as comparable across countries and economies. All questionnaire items used for the scaling of this index had an item invariance index exceeding the cut-off point of 0.6 and, in all countries and economies, the proportion of invariant parameters exceeded 0.25. Thus this index can be considered as fully comparable across all countries/economies.

Parents' interest in learning about other cultures (INTCULTP)

This construct had the lowest invariance index of 0.69. Of the four questionnaire items used in scaling this index, the one on interest in learning about the religions of the world had an item invariance index below 0.6. As indicated above, religion might be understood differently by people in different countries. For all countries/economies taking the parent questionnaire, the proportion of invariant parameters exceeded 0.25. Thus the relative low comparability of this index is due to the low comparability of one item rather than underlying differences between countries/economies.

Parents' attitudes towards immigrants (ATTIMMP)

This construct had the lowest invariance index of 0.72. Of the four questions asked to parents about their attitudes towards immigrants, the question on whether immigrant children should have the same opportunities for education that other children in the country have had an item invariance index below 0.6 making it less comparable across countries. For all countries/economies, the proportion of invariant parameters exceeded 0.25. Thus the relative low comparability of this index is due to the low comparability of one item rather than underlying differences between countries/economies.

PISA 2018 standard approach to examine the comparability of scaled indices

While the forthcoming PISA 2018 Technical Report (OECD, Forthcoming[2]) will explain in detail the scaling procedures and the construct validation of all context- questionnaire data, this section presents a summary of the standard analyses carried out to validate the cross-country comparability of the main scaled indices included in the PISA 2018 database. The internal consistency of scaled indices and the invariance of item parameters are the two approaches that PISA 2018 used to examine the comparability of scaled indices across school systems.

Internal consistency refers to the extent to which the items that make up an index are inter-related. Cronbach's Alpha was used to check the internal consistency of each scale within the

countries/economies and to compare it amongst countries/economies. The coefficient of Cronbach's Alpha ranges from 0 to 1, with higher values indicating higher internal consistency. Similar and high values across countries/economies are an indication of having measured reliably across countries/economies. Commonly accepted cut-off values are 0.9 for excellent, 0.8 for good, and 0.7 for acceptable internal consistency. Table VI.A5.9, available online, presents the Cronbach's Alpha for the scaled indices in this volume.

PISA 2018 examined the cross-country comparability of scaled indices also through the invariance of item parameters. The idea was to test whether the item parameters of an index could be assumed to be the same (invariant) across groups of participating countries and language groups. In a first step, groups were defined based on samples of at least 300 students responding to the same language-version questionnaire in a country. In a second step, international and student parameters were estimated based on students across all groups. In a third step, the root mean square deviance (RMSD) item-fit statistics was calculated for each group and item. Values close to zero signal a good item-fit, indicating that the international model describes student responses within individual groups accurately. Any group receiving a value above 0.3 was flagged and a group-specific item parameter was calculated. Steps 2 and 3 were then repeated until all items exhibited RMSD values below 0.3. The RMSD values will be reported in the forthcoming PISA 2018 Technical Report. Amongst the main indices examined in this volume, some needed just one round to ensure that all items exhibited acceptable levels of RMSD, whereas other indices needed several iterations:

- One round: Self-efficacy regarding global issues (gcselfeff), Parents' awareness of global issues (gcawarep), Perspective taking (perspect), Respect for people from other cultures (respect), Students' attitudes towards immigrants (attimm), Cognitive adaptability (cogflex), Parents' interest in learning about other cultures (intcultp), and Awareness of intercultural communication (awacom)
- Several rounds: Students' awareness of global issues (gcaware; 2 rounds), Enjoyment of reading (joyread, 2 rounds), Students interest in learning about other cultures (intcult, 2 rounds), and Global mindedness (globmind, 2 rounds)

Tables available on line

https://doi.org/10.1787/888934171134

- Table VI.A5.1   Student-level indices, item-level analyses
- Table VI.A5.2   Student-level indices, country-level analyses
- Table VI.A5.3   School-level indices, item-level analyses
- Table VI.A5.4   School-level indices, country-level analyses
- Table VI.A5.5   Teacher-level indices, item-level analyses
- Table VI.A5.6   Teacher-level indices, country-level analyses
- Table VI.A5.7   Parent-level indices, item-level analyses
- Table VI.A5.8   Parent-level indices, country-level analyses
- Table VI.A5.9   Internal consistency of the main scaled indices

References

Asparouhov, T. and B. Muthén (2014), "Multiple-Group Factor Analysis Alignment", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 21/4, pp. 495-508, http://dx.doi.org/10.1080/10705511.2014.919210.   [1]

OECD (Forthcoming), *PISA 2018 Technical Report*.   [2]