

Chapter 14

Data Adjudication

INTRODUCTION

Data adjudication is the process through which each national dataset is reviewed and a judgement about the appropriateness of the data for the main reporting goals is formed. The PISA Technical Standards (see Annex I) specify the way in which PISA must be implemented in each participating jurisdiction and adjudicated region. International contractors monitor the implementation in each of these and adjudicate on their adherence to the standards. This chapter describes the process used to adjudicate the PISA 2018 data for each of the adjudicated entities (i.e. the participating countries and economies – hereafter, “jurisdictions” – and the adjudicated regions) and gives the outcomes of data adjudication that are mainly based on the following aspects:

- the extent to which each adjudicated entity met PISA sampling standards
- the outcomes of the adaptation, translation, and verification process
- the outcomes of the PISA Quality Monitoring visits
- the quality and completeness of the submitted data, including concerns about the quality of the data that were identified during scaling and in preparation for reporting
- the outcomes of the international coding review.

Not all regions (i.e., subnational jurisdictions that report their results separately) opt to undergo the full adjudication that would allow their results to be compared statistically to all other participating economies and adjudicated regions. For example, the states of Australia are not adjudicated regions, whereas the Flemish Community of Belgium is an adjudicated region.

PISA 2018 Technical Standards

The areas covered in the PISA 2018 Technical Standards include the following:

Data Standards 1-15

- target population and sampling
- language of testing
- Field Trial participation
- adaptation of tests, questionnaires, and school-level materials
- translation of tests, questionnaires, and school-level materials
- testing of national software version
- technical support
- test administration
- training support
- implementation of national options
- security of the material
- quality monitoring
- assembling and printing paper-based material
- response coding
- data submission

Management standards 16-20

- communication with the international contractors
- notification of international and national options
- schedule for submission of materials
- management of data
- archiving of materials

National involvement standards 21-22

- national feedback
- meeting attendance

Implementing the standards – quality assurance

National Project Managers of participating jurisdictions are responsible for implementing the standards based on the international contractors' advice as contained in the various operational manuals and guidelines. Throughout the cycle of activities for each PISA survey, the international contractors carried out quality-assurance activities in two steps. The first step was to set up quality- assurance procedures using the operational manuals, as well as the agreement processes for national submissions on various aspects of the project. These processes gave the international contractor staff the opportunity to ensure that PISA implementation was planned in accordance with the PISA 2018 Technical Standards and to provide advice on taking rectifying action when required and before critical errors occurred. The second step was quality monitoring, which involved the systematic collection of data that monitored the implementation of the assessment in relation to the standards. For the data adjudication, information collected during both the quality-assurance and quality-monitoring activities was used to determine the level of compliance with the standards.

Information available for adjudication

The international contractors' quality monitoring of a participating jurisdiction's data collection is carried out from a range of perspectives during many stages of the PISA cycle. These perspectives include monitoring a participating jurisdiction's adherence to the deadlines, communication from the sampling contractor about each participating jurisdiction's sampling plan, information from the linguistic verification team, data from the PISA Quality Monitors, and information gathered from direct interviews at National Project Manager and Coder Training meetings. The information was combined together in the database so that:

- indications of non-compliance with the standards could be identified early on in order to enable rectifying measures
- the point at which the problem occurred could be easily identified
- information relating to the same PISA standard could be cross-checked between different areas or sources.

Many of these data collection procedures refer to specific key documents, specified in the National Project Manager's Manual and the Sampling Manual in particular. These are procedures that the international contractors require for Field Trial and Main Survey preparation from each National Centre. The data adjudication process provides a motivation for collating and summarising the specific information relating to PISA standards collected in

these documents, combined with information collected from specific quality monitoring procedures such as the PISA Quality Monitor visits and from information in the submitted data.

The quality monitoring information was collected from various quality monitoring instruments and procedures and covered the following main areas:

- international contractors' administration and management: information relating to administration processes, agreement of adaptation spreadsheets, submission of information
- item developers: issues identified in the coder query service and training of coders;
- translation: information from linguistic verification of test items, questionnaire items, and the test administration script
- sampling: information from the submitted data such as school and student response rates, exclusion rates and eligibility problems
- school-level materials: information from the agreement of adaptations to test administration procedures and field operations
- Final Optical Check team: information from the pre- and post-Main Survey Final Optical Checks of Main Survey booklets
- questionnaire Final Optical Check: issues arising from the Final Optical Check of the questionnaires
- National Centre Test Administrator or School Associate trainings
- National Centre quality monitoring: information gathered through interviews conducted during meetings of National Project Managers or at other times
- co-ordination of PISA Quality Monitor activities including recruitment
- PISA Quality Monitor reports: information gathered via the Data Collection Forms from PISA Quality Monitors and through their interactions with School Co-ordinators and Test Administrators
- data cleaners: issues identified during the data cleaning checks and from data cleaners' reports
- data analysis: information from item level reports, from the Field Trial data, and from data cleaning steps, including consistency checks
- data processing: issues relating to the eligibility of students tested
- questionnaire data: issues relating to the questionnaire data in the national questionnaire reports provided by the international contractor
- Main Survey and Field Trial Reviews: information provided by the National Project Managers in the Field Trial and Main Survey Review Questionnaires.

Quality monitoring reports

There were two types of PISA quality monitoring reports: The Session Report Form containing data for each session in each school, and the Data Collection Form detailing the general observations across all schools visited by PQMs. The Session Report Form was completed by the Test Administrator after each test session and also contained data related to test administration. The data from this report were recorded by the National Centre and submitted as part of the national dataset to ETS. The PISA Quality Monitor reports contained data related to test administration in selected schools, and the PISA quality monitoring data were collected independently of the National Project Manager.

Data adjudication process

Data adjudication is the process through which each national dataset is reviewed and a judgement about the appropriateness of the data for the main reporting goals is formed. The different steps in the data adjudication process ensure that the final judgement is transparent, based on evidence, and defensible. The data adjudication process achieved this through the following steps:

Step 1: The Consortium collected quality-assurance and quality monitoring data throughout the survey administration period. The international project manager compiled this information into an adjudication database that was updated or amended as new information arose and provided an overview of the national implementation of PISA throughout the cycle. .

Step 2: The international project manager compiled individual reports for each jurisdiction that contained quality-assurance data for key areas of project implementation.

Step 3: The international project director, together with the consortium leads, identified data issues that were in need of adjudication. Where necessary, the relevant National Project Manager was contacted to provide additional information. After this stage, for each dataset, a summary report detailing how the PISA Technical Standards had been met was constructed..

Step 4: The PISA Adjudication Group, formed by representatives of the OECD and of international contractors, the Technical Advisory Group and the Sampling Referee, reviewed the summary reports to recommend adequate treatment of the data from each adjudicated entity in international PISA products (database and reports).

Step 5: The recommendations of the Data Adjudication Group were presented to the PISA Governing Board representatives and to the countries concerned.

Monitoring compliance to any single standard occurred through responses to one or more quality-assurance questions regarding test implementation and national procedures which may come from more than one area. For example, the session report data were used in conjunction with the PISA Quality Monitor reports, computer system tracking of timings, and information from the adaptation of national manuals to assess compliance with the PISA session timing standard (Standard 6.1, Annex F).

Information was collected in relation to these standards through a variety of mechanisms:

- through PISA Quality Monitor reports
- through the Field Trial and Main Survey reviews
- through information negotiated and stored on the PISA Portal website (the portal which was used in PISA 2018)
- through a system database specific to the implementation of PISA tasks
- through the formal and informal exchanges between the international contractors and National Centres over matters such as sampling, translation and verification, specially requested analyses (such as non-response bias analysis)
- through a detailed post-hoc inspection of all Main Survey assessment materials (test booklets)
- through the data cleaning and data submission process.

For PISA 2018, an adjudication database was developed to capture, summarise, and store the most important information derived from these various information sources. The staff members of the international contractor who led each area of work were responsible for identifying relevant information and entering it into the database. This means that at the time of data adjudication, relevant information was easily accessible for making recommendations about the fitness of use of data from each PISA adjudicated entity.

The adjudication database captured information related to the major phases of the data operation: field operations, sampling, questionnaires, and tests. Within each of these phases, the specific activities are identified, and linked directly to the corresponding standards.

Within each section of the database, specific comments are entered that describe the situation of concern, the source of the evidence about that situation, and the recommended action. Each entry is classified as serious, minor, or of no importance for adjudication. Typically, events classified as serious would warrant close expert scrutiny and possibly action affecting adjudication outcomes. For example, cognitive data for Viet Nam were found to be inconsistent with item-response patterns observed in prior administrations of PISA (including both Field Trial and Main Study administrations), resulting in the exclusion of Viet Nam's performance scores from international comparisons and comparisons with results for Viet Nam from previous years. Events classified as minor would typically not directly affect adjudication outcomes but will be reported back to National Centres to assist them in reviewing their national procedures.

Data adjudication outcomes

It was expected that the data adjudication would result in a range of possible recommendations to the PISA Governing Board. Some possible, foreseen recommendations included:

- that the data be declared fit for use
- that some data be removed for a particular participating jurisdiction or adjudicated region, such as the removal of data for some open-ended items or the removal of data for some schools
- that rectifying action be performed by the National Project Manager, such as providing additional evidence to demonstrate that there was no non-response bias, or rescoring open-ended items
- that the data not be endorsed for use in certain types of analyses
- that the data not be endorsed for inclusion in the PISA 2018 database.

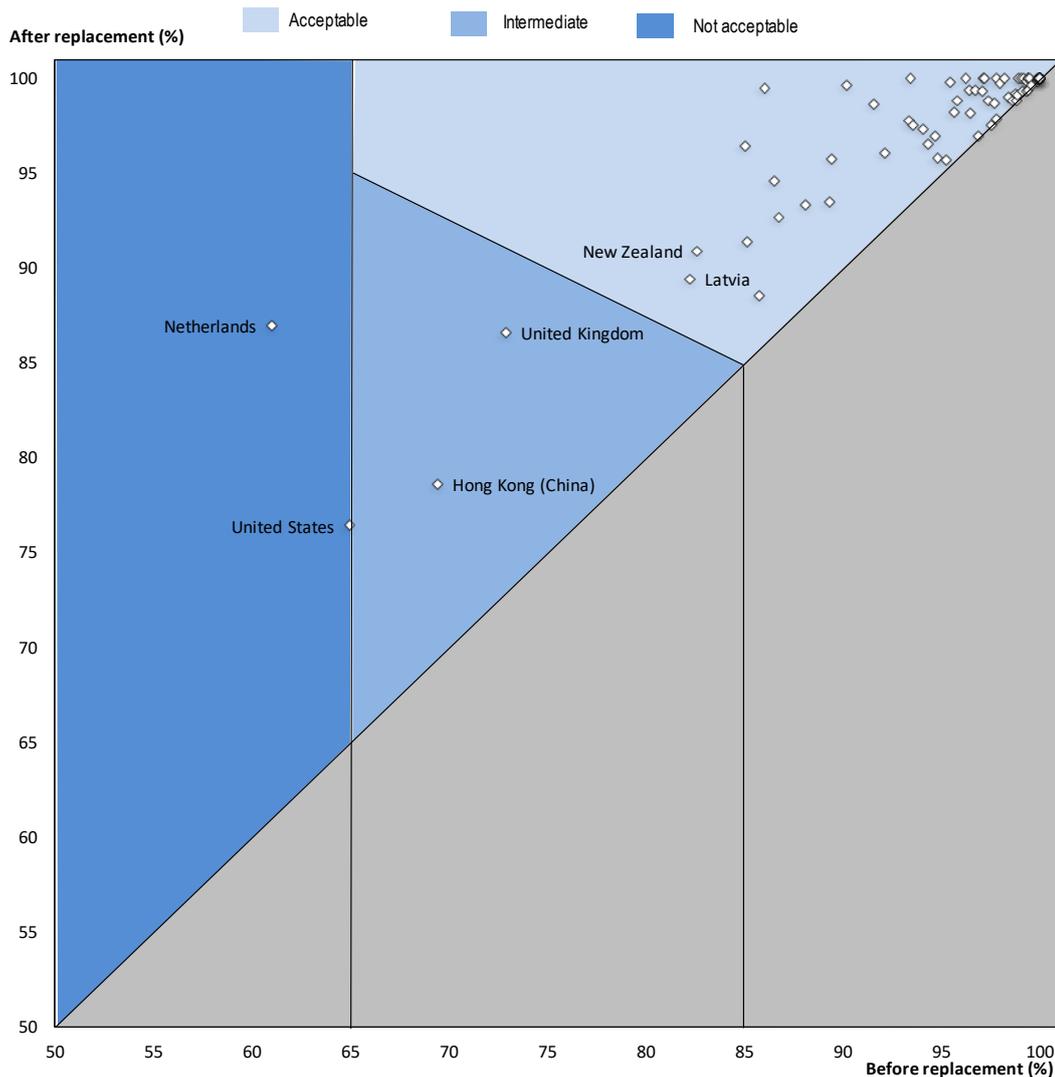
Throughout PISA 2018, the international contractors concentrated their quality control activities to ensure that the highest scientific standards were met. However, during data adjudication a wider definition of quality was used, especially when considering data that were at risk. In particular, the underlying criterion used in adjudication was fitness for use; that is, data were endorsed for use if they were deemed to be fit for meeting the major intended purposes of PISA.

GENERAL OUTCOMES

Overview of response rate issues

The PISA school response rate requirements are discussed in Chapter 4. Figure 14.1 is a scatter plot of the attained PISA school response rates before and after replacements. Those jurisdictions that are plotted in the light blue shaded region were regarded as fully satisfying the PISA school response rate criterion.

Figure 14.1: Attained school response rates



Latvia and New Zealand had a response rate below the 85% level before the use of replacement schools but cleared the acceptable level after the replacement schools were included.

Four jurisdictions – Hong Kong (China), the Netherlands, the United Kingdom and the United States – failed to meet the school non-response rate (see Figure 14.1). One participating jurisdiction – Portugal – fell short of the student response-rate standard. After reviewing the sampling outcomes, the consortium asked these five jurisdictions to provide additional data

that would assist the adjudication group in making a balanced judgement about the threat of the non-response to the accuracy of inferences which could be made from the PISA data.

Detailed participating jurisdiction comments

It is important to recognise that PISA data adjudication is a late but not necessarily final step in the quality assurance process. By the time each participating jurisdiction was adjudicated at the Technical Advisory Group meeting in June 2019, the quality assurance and monitoring processes outlined earlier in this chapter and in Chapter 7 had been implemented. Data adjudication focused on residual issues that remained after these quality assurance processes had been carried out.

The remaining issues fall under two broad categories: (1) adaptations to the recommended international standard procedures in a participating jurisdiction's data collection plan, and (2) a failure to meet international standards at the implementation stage.

Departures from standard procedures in the national data collection plan

With such a broad and diverse range of participation, it is to be expected that the international best practice approaches to data collection articulated in the PISA Technical Standards document may not be achieved in all national and local contexts. This may be the case for a number of reasons. For example, it may be contrary to national protocols to have unannounced visits of quality monitors to schools to observe test administration. Typically these were discussed with international contractor experts in advance of the assessment, and alternative approaches were considered jointly between the National Project Manager and the international contractor. In isolated departures from best practice, a judgement might easily be made by international contractor experts that there was minimal risk to the quality of the data collection plan. Such isolated departures are not reported in the participating jurisdiction summaries below.

On the other hand, it may not have been straightforward to determine in advance of the assessment how more extensive or multiple departures from PISA Technical Standards may interact with each other and with other aspects of a participating jurisdiction's data collection plan. Cases such as these were considered as part of the data adjudication process and are included in the participating jurisdiction summaries below.

Departures from standards arising from implementation

Departures from the standards at the implementation stage range from errors within the National Centre (e.g., during the final stages of preparing materials, or in the administration of the coding operation following data collection), a failure to meet documented targets during data collection, for example a shortfall from the minimum school and student sample sizes.

A component of the data adjudication process was to consider the cases of multiple, or more complex departures from the PISA standard procedures, as well as to consider the impact of errors or shortfalls across all aspects of each participating jurisdiction's data collection plan and implementation, and make an evaluation with respect to the quality and international comparability of the PISA results. Notable departures from the standards are reported in the participating jurisdiction summaries below.

Several jurisdictions exceeded the limit on student- and school-level exclusions (5% at most; see Tables 11.1 and 11.2). In jurisdictions where other violations of sampling standards were observed or where the combined level of exclusions exceeded 7%, further information was requested to support the case that no bias would result from exclusions. The number of such cases shows a notable increase over the level observed in previous cycles; at the same time, the average exclusion rates have barely increased, and several countries were able to reduce their exclusion rates with respect to 2015.

A small number of jurisdictions failed to reach the required minimum sample sizes specified in the standards. Such cases were considered as part of the data adjudication process. Even a minor deviation in sample size might be considered a substantive enough issue to report, for example in jurisdictions where standard errors tend to be higher for a given sample size. In PISA 2018, most deviations observed were either minor deviations (i.e. shortfalls of fewer than 50 students or 5 schools) or in jurisdictions that nevertheless achieved comparable standard errors on the major survey estimates; they are therefore not reported below.

Particular attention has been paid to the achievement of the specified response rates of 85% for schools, 80% for students within schools and no more than 5% of students excluded from the assessment. Five jurisdictions were required to provide additional data to support the case that no bias would result from failure to meet the response-rate standards.

Anomalies in submitted data, particularly inconsistencies and deviations from the expected patterns, were also investigated; most cases could be explained and solved through a resubmission of data. The one case that could not be solved is noted below.

If a participating jurisdiction is not listed below then it fully met the PISA standards. Further, in the case of minor deviations from the standards, unless otherwise noted, additional data were usually available to suggest the data were suitable for use.

Australia

There was a total of 5.72% exclusions in Australia; data were included in the final database.

Canada

There was a total of 6.87% exclusions in Canada; in consideration of the fact that the nature and amount of exclusions in 2015 was similar (7.49%), data were included in the final database.

Cyprus

There was a total of 5.99% exclusions in Cyprus; data were included in the final database.

Denmark

There was a total of 5.70% exclusions in Denmark; data were included in the final database.

Hong Kong (China)

The school participation rate, before replacement, was 69%; after replacement, it was 79%. A significant portion of non-participation is the result of some schools failing to reach a 50%

response rate among students, rather than of schools not participating at all. In line with established PISA procedures, data from responding students in these schools are included in the database, but the schools are considered as non-responding for the purpose of assessing compliance with response-rate standards.

A school-non-response bias analysis was submitted, showing that, within each sampling stratum, school size and school average achievement (based on respondents) are unrelated to student response rates. In addition, the national centre documented a moderate correlation between school size and PISA performance and explained that the non-response issue was related to low initiative in participation by schools, teacher and students, and that adequate remedial actions had been taken once the problem was identified. In consideration of the nature of the non-response problem, of the fact that a raw, but direct measure of school performance is used to assign schools to sampling strata (and therefore, differential non-response across strata is unlikely to cause bias), and of the available information, the adjudication group concluded that there is limited risk of bias due to non-response, and recommended the inclusion of data in the final database.

Iceland

There was a total of 5.99% exclusions in Iceland; data were included in the final database.

Israel

There was a total of 10.21% exclusions in Israel.

The high exclusion rate in Israel was the result of the lack of participation by a particular type of boys' school. These schools were considered to be non-responding schools in cycles up to 2015 but were treated as school-level exclusions in 2018.

Kazakhstan

There was a total of 5.87% exclusions in Kazakhstan; data were included in the final database.

Lebanon

Lebanon administered a field-trial version of the student questionnaire. As a result, the database does not contain information about several questions that were rotated in the field trial across questionnaire booklets. This violation of the PISA technical standard does not affect performance data, but results in incomplete questionnaire data for Lebanon.

Luxembourg

There was a total of 7.92% exclusions in Luxembourg; in consideration of the fact that the nature and amount of exclusions in 2015 was similar (8.16%), data were included in the final database.

The Netherlands

There was a total of 6.24% exclusions in the Netherlands.

The school participation rate, before replacement, was 61%; after replacement, it was 87%.

Some 17% of responding students (weighted) were assigned to UH booklets, intended for students with special education needs and containing only reading, mathematics and science questions, and a shorter version of the student questionnaire.

A school-non-response bias analysis was submitted, relying on a direct measure of school performance in a national assessment, and comparing the final PISA sample to the original PISA sample, including non-responding schools and excluding replacement schools. Based on the available information, the Adjudication Group concluded that there is limited risk of bias due to non-response, and recommended the inclusion of data in the final database for analyses of reading, mathematics and science performance.

However, for analyses of financial literacy performance and of most questionnaire scales, the effective rate of exclusion is increased by the high proportion of students assigned to UH booklets. The Adjudication Group concluded that results for financial literacy are not comparable to those of other jurisdictions. Data for the Netherlands (financial literacy sample) were therefore not included in the international dataset, and are available as a separate dataset.

New Zealand

There was a total of 6.78% exclusions in New Zealand. In consideration of the fact that the nature and amount of exclusions in 2015 was similar (6.54%), data were included in the final database.

Norway

There was a total of 7.88% exclusions in Norway. In consideration of the fact that the nature and amount of exclusions in 2015 was similar (6.75%), data were included in the final database.

The Republic of North Macedonia

The Republic of North Macedonia administered a field-trial version of the student questionnaire. As a result, the database does not contain information about several questions that were rotated in the field trial across questionnaire booklets. This violation of the PISA technical standard does not affect performance data, but results in incomplete questionnaire data for the Republic of North Macedonia.

Portugal

The weighted student-response rate, after replacement, was 76%.

A student-non-response-bias analysis was submitted, investigating bias amongst students in grades 9 and above (students in grades 7 and 8 represented only about 11% of the total sample, but 20% of the non-respondents). A comparison of the linked responding and non-responding cases, using sampling weights, revealed that non-respondents tended to score about one-third of a standard deviation below respondents on the national mathematics examination (implying a “raw” upward bias of about 10% of a standard deviation on population statistics that are based on respondents only). At the same time, a significant proportion of the performance differences could be accounted for by variables considered in non-response adjustments (including grade

level). Nevertheless, a residual upward bias in population statistics remained, even when using non-response adjusted weights. The non-response bias analysis therefore implies a small upward bias for PISA 2018 performance results in Portugal. The Adjudication Group however considered that trend comparisons and performance comparisons with other countries may not be particularly affected, because an upward bias of that size cannot be excluded even in countries that met the response-rate standard or for previous cycles of PISA. Data were included in the final database, and are reported with an annotation.

Sweden

There was a total of 11.09% exclusions in Sweden, a marked increase over previous cycles.

Because of Swedish data-collection laws, the reasons for exclusion could not be explicitly stated in student-tracking forms. However, the adjudication group considered it plausible that this increase was the consequence of a large and temporary increase in immigrant flows and therefore of students with limited language experience. It is expected that the exclusion rate will decrease to previous levels in future cycles of PISA, as such inflows stabilise or shrink. Data were included in the final database.

Switzerland

There was a total of 6.68% exclusions in Switzerland; data were included in the final database.

Turkey

There was a total of 5.66% exclusions in Turkey; data were included in the final database.

The increase in exclusions over previous cycles could be attributed to a particular type of non-formal education institutions which were previously not listed in sampling frames, and which were listed, but excluded, in 2018.

The United Kingdom

The school participation rate, before replacement, was 73%; after replacement, it was 87%.

A school-non-response bias analysis was submitted, limited to England (the largest subnational entity within the United Kingdom) and relying on a direct measure of school performance in a national assessment. This analysis investigated differences between responding and non-responding schools and between originally sampled schools and replacement schools. This supported the case that no notable bias would result from non-response. Based on the available information, the Adjudication Group concluded that there is limited risk of bias due to non-response, and recommended the inclusion of data in the final database.

Data for the Scotland, an adjudicated subnational entity within the United Kingdom, fully met the standards.

The United States

The school participation rate, before replacement, was 65%; after replacement, it was 76%.

A school-non-response bias analysis was submitted, indicating that, after replacement schools and non-response adjustments are taken into account, a number of characteristics (not including direct measures of school performance) are balanced across respondents and non-respondents. Based on the available information, the Adjudication Group concluded that non-response adjustments could minimise the risk of bias due to non-response, and recommended the inclusion of data in the final database.

Viet Nam

In Viet Nam, while no major violation of implementation standards was identified, there were several minor violations. Furthermore, during scaling and reporting, the PISA consortium identified several technical issues affecting the comparability of their data, an essential dimension of data quality in PISA.

Viet Nam's cognitive data show poor fit to the item-response-theory model, with more significant misfit than any other country/ language group (see Chapter 12). In particular, Viet Nam's cognitive response data do not follow a predictable pattern in terms of item difficulty, as they do in other countries; and selected-response questions, as a group, appeared to be significantly easier for students in Viet Nam than expected, given the usual relationship between open-ended and selected-response questions reflected in the international model parameters. In addition, for several selected-response items, response patterns were not consistent across administrations (PISA 2015 and PISA 2018). Finally, the corresponding cognitive data for the field trial also differ from main-study data in unexpected way. This accumulation of inconsistencies over time implies that the large number of unique parameters for the main study is not a reflection of language specificities, item translation issues, or alignment with the curriculum, but rather of the specific conditions under which the main study is administered.

The Adjudication Group recommended not including Viet Nam's data in the international dataset, and making them available as a separate dataset. In particular, and pending further investigations on the nature and origin of the observed issues, comparisons of performance in reading, mathematics and science with all other countries should be avoided; Viet Nam's mean performance should not be ranked together with other countries' performance; and Viet Nam's results should not be interpreted in terms of the described proficiency levels.