

ANNEX A4

Quality assurance

Quality assurance procedures were implemented in all parts of PISA 2018, as was done for all previous PISA surveys. The PISA 2018 Technical Standards (available on line at www.oecd.org/pisa) specify the way in which PISA must be implemented in each country, economy and adjudicated region. International contractors monitor the implementation in each of these and adjudicate on their adherence to the standards.

The consistent quality and linguistic equivalence of the PISA 2018 assessment instruments were facilitated by assessing the ease with which the original English version could be translated. Two source versions of the assessment instruments, in English and French, were prepared (except for the financial literacy assessment and the operational manuals, which were provided only in English) in order for countries to conduct a double translation design, i.e. two independent translations from the source language(s), and reconciliation by a third person. Detailed instructions for the localisation (adaptation, translation and validation) of the instruments for the field trial and for their review for the main survey, and translation/adaptation guidelines were supplied. An independent team of expert verifiers, appointed and trained by the PISA Consortium, verified each national version against the English and/or French source versions. These translators' mother tongue was the language of instruction in the country concerned, and the translators were knowledgeable about education systems. For further information on PISA translation procedures, see the *PISA 2018 Technical Report* (OECD, forthcoming^[1]).

The survey was implemented through standardised procedures. The PISA Consortium provided comprehensive manuals that explained the implementation of the survey, including precise instructions for the work of school co-ordinators and scripts for test administrators to use during the assessment sessions. Proposed adaptations to survey procedures, or proposed modifications to the assessment session script, were submitted to the PISA Consortium for approval prior to verification. The PISA Consortium then verified the national translation and adaptation of these manuals.

To establish the credibility of PISA as valid and unbiased and to encourage uniformity in conducting the assessment sessions, test administrators in participating countries were selected using the following criteria: it was required that the test administrator not be the reading, mathematics or science instructor of any student in the sessions he or she would conduct for PISA; and it was considered preferable that the test administrator not be a member of the staff of any school in the PISA sample. Participating countries organised an in-person training session for test administrators.

Participating countries and economies were required to ensure that test administrators worked with the school co-ordinator to prepare the assessment session, including reviewing and updating the Student Tracking Form; completing the Session Attendance Form, which is designed to record students' attendance and instruments allocation; completing the Session Report Form, which is designed to summarise session times, any disturbance to the session, etc.; ensuring that the number of test booklets and questionnaires collected from students tallied with the number sent to the school (for countries using the paper-based assessment) or ensuring that the number of USB sticks or external laptops used for the assessment were accounted for (for countries using the computer-based assessment); and sending or uploading the school questionnaire, student questionnaires, parent and teacher questionnaires (if applicable), and all test materials (both completed and not completed) to the national centre after the assessment.

The PISA Consortium responsible for overseeing survey operations implemented all phases of the PISA Quality Monitor (PQM) process: interviewing and hiring PQM candidates in each of the countries, organising their training, selecting the schools to visit, and collecting information from the PQM visits. PQMs are independent contractors located in participating countries who are hired by the international survey operations contractor. They visit a sample of schools to observe test administration and to record the implementation of the documented field-operations procedures in the main survey.

Typically, two or four PQMs were hired for each country, and they visited an average of 15 schools in each country. If there were adjudicated regions in a country, it was usually necessary to hire additional PQMs, as a minimum of five schools were observed in adjudicated regions.

Approximately one-third of test items are open-ended items in PISA. Reliable human coding is critical for ensuring the validity of assessment results within a country, as well as the comparability of assessment results across countries. Coder reliability in PISA 2018 was evaluated and reported at both within- and across-country levels. The evaluation of coder reliability was made possible by the design of multiple coding: a portion or all of the responses from each human-coded constructed-response item were coded by at least two human coders.

All quality-assurance data collected throughout the PISA 2018 assessment were entered and collated in a central data-adjudication database on the quality of field operations, printing, translation, school and student sampling, and coding. Comprehensive reports were then generated for the PISA Adjudication Group. This group was formed by the Technical Advisory Group and the Sampling Referee. Its role is to review the adjudication database and reports in order to recommend adequate treatment to preserve the quality of PISA data. For further information, see the *PISA 2018 Technical Report* (OECD, forthcoming^[1]). Overall, the review suggested good adherence of national implementations of PISA to the technical standards. Despite the overall high quality of data, a few countries' data failed to meet critical standards or presented inexplicable anomalies, such that the Adjudication Group recommends a special treatment of these data in databases and/or reporting.

The major issues for adjudication discussed at the adjudication meeting are listed below:

- In Viet Nam, while no major standard violation was identified, there were several minor violations and the adjudication group has identified 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. In particular, 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 are not consistent across field trial and main survey administrations, ruling out possible explanations of misfit in terms of familiarity, curriculum or cultural differences. For this reason, the OECD cannot currently assure full international comparability of the results.
- The Netherlands missed the standard for overall exclusions by a small margin. At the same time, in the Netherlands UH booklets, intended for students with special education needs, were assigned to about 17% of the non-excluded students. Because UH booklets do not cover the domain of financial literacy, the effective exclusion rate for the financial literacy additional sample is above 20%. The fact that students that receive support for learning in school were systematically excluded from the financial literacy sample results in a strong upward bias for the country mean and other population statistics. Therefore, the Netherlands' results in financial literacy may not be comparable to those of other countries or to results for the Netherlands from previous years. The Netherlands also missed the school response rate (before replacement) by a large margin, and could only reach close to an acceptable response rate through the use of replacement schools. Based on evidence provided in a non-response bias analysis, the Netherlands' results in reading, mathematics and science were accepted as largely comparable, but, in consideration of the low response rate amongst originally sampled schools, are reported with an annotation.
- Portugal did not meet the student-response rate standard. In Portugal, response rates dropped between 2015 and 2018. A student-non-response-bias analysis was submitted, investigating bias amongst students in grades 9 and above. Students in grades 7 and 8 represented 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 also 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. Therefore, Portugal's results are reported with an annotation.

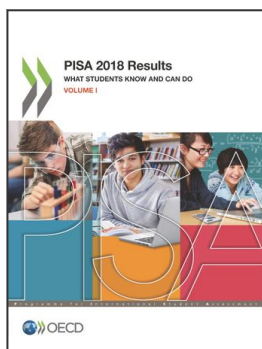
While the adjudication group did not consider the violation of response-rate standards by Hong Kong (China) and the United States (see Annex A2) as major adjudication issues, they noted several limitations in the data used in non-response-bias analyses submitted by Hong Kong (China) and the United States. In consideration of the lower response rates, compared to other countries, the data for Hong Kong (China) and the United States are reported with an annotation.

In Spain, while no major standard violation was identified, subsequent data analyses identified sub-optimal response behaviours of some students. This was especially evident in the reading-fluency items. The reporting of Spain's reading performance will be deferred as this issue will be further investigated. For more details, see Annex A9.

Reference

OECD (forthcoming), *PISA 2018 Technical Report*, OECD Publishing, Paris.

[1]



From:

PISA 2018 Results (Volume I)

What Students Know and Can Do

Access the complete publication at:

<https://doi.org/10.1787/5f07c754-en>

Please cite this chapter as:

OECD (2019), “Quality assurance”, in *PISA 2018 Results (Volume I): What Students Know and Can Do*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/7e39b257-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area. Extracts from publications may be subject to additional disclaimers, which are set out in the complete version of the publication, available at the link provided.

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at <http://www.oecd.org/termsandconditions>.