

Annex A4. Quality assurance

Quality-assurance procedures were implemented in all parts of PISA 2022, as was done for all previous PISA surveys. The PISA 2022 Technical Standards (available at <https://www.oecd.org/pisa/>) specify the way in which PISA must be implemented in each country, economy and adjudicated region. The PISA Consortium monitors the implementation in each of these and adjudicates on their adherence to the standards.

The consistent quality and linguistic equivalence of the PISA 2022 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, i.e. two independent translations from the source language(s), with 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 2022 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 training 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 mathematics, reading and science assessments 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 2022 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 were collected by the PISA Consortium from each adjudicated entity (89 adjudication entities including countries, economies and regions) throughout the PISA 2022 assessment. These data were entered and collated in a central data-adjudication database on the quality of field operations, printing, translation, school and student sampling, and coding. This process identifies data issues that are in need of adjudication.

Comprehensive reports were then generated for the PISA Adjudication Group. This group is composed of 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 2022 Technical Report* (OECD, forthcoming^[1]).

Overall, the Adjudication Group's review suggests good adherence of national implementations of PISA to the technical standards in spite of the challenging circumstances that affected not only PISA operations but schooling more generally during the COVID-19 pandemic. Thanks to the reactivity and flexibility of participating countries and international contractors, to carefully constructed instruments, to a test design that is aligned to the main reporting goals and is supported by adequate sample design, and to the use of appropriate statistical methods for scaling, population estimates are highly reliable and comparable across countries and time, and particularly with 2018 results.

Nevertheless, a number of deviations from standards were noted and their consequences for data quality were reviewed in depth. The following overall patterns of deviations from standards were identified:

- About one in five of all adjudicated entities had exclusion rates exceeding the limits set by the technical standards (Standard 1.7).
- Seven entities failed to meet the required school response rates, with three of them failing to meet the stricter level of 65% before replacement (Standard 1.11). This is in line with earlier cycles of PISA.
- There was a significant increase in the number of entities that failed to meet the required student response rates (Standard 1.12): 10 entities did not meet this standard.
- There were delays in data submission in a significant number of entities (Standard 19.1): 14 entities did not meet this standard, and 13 only partially met it. The Adjudication Group noted that delayed submissions may affect the quality of the international contractors' work; and if shorter reporting timelines are expected, it may no longer be possible to accommodate such delays.
- A large number of entities did not conduct the field trial as intended (Standard 3.1) or did not attend all meetings (Standard 23.1). While this may also be a consequence of the pandemic, the Adjudication Group noted that these violations may be particularly consequential for new participants and for less-experienced teams. The Group underlined the importance of attendance at coder training sessions for ensuring comparability of the data.

At the international level, these frequent deviations should guide future efforts of the PISA Governing Board, the OECD Secretariat and Contractors to review the corresponding standards, prevent future deviations from standards, or mitigate the consequences of such violations.

At the level of individual adjudicated countries, economies and regions, in most cases, these issues did not result in major threats to the validity of reports, and the data could be declared fit for use. Where school or student participation rates fell short of the standard and created a potential threat for non-response/non-participation bias, countries/economies were requested to submit non-response-bias analyses. The evidence produced by countries/economies (and in some cases, by the sampling contractor) was reviewed by the Adjudication Group.

The Adjudication Group reviewed and discussed major adjudication issues in June 2023. The major adjudication issues reviewed by the group are listed below:

- The 13 adjudicated entities listed below did not meet one or more PISA sampling standards. See the Reader's Guide at the beginning of this volume for a detailed account of the sampling issues for each of the 13 entities. The results of these countries/economies are reported with annotations. Two groups can be distinguished among the 13 entities:
 - Entities that submitted technically strong analyses, which indicated that more than minimal bias was most likely introduced in the estimates due to low response rates (falling below PISA standards): Canada, Ireland, New Zealand, the United Kingdom and Scotland.
 - Entities that did not meet one or more PISA sampling standards and it is not possible to exclude the possibility of more than minimal bias based on the information available at the time of data adjudication: Australia, Denmark, Hong Kong (China), Jamaica, Latvia, the Netherlands, Panama and the United States.
- In **Ukraine**, the overall exclusion rate was 36.1%, when computed with respect to the original sampling frame, covering the entire country (See Annex A2). However, most exclusions resulted from the fact that survey operations could not be completed successfully in the regions most affected by war. Results for the remaining regions (18 out of 27) were deemed fit for reporting, but comparisons with previous results should be made only with great caution, and with due consideration of the differences in target populations.
- For **Viet Nam's** reading scores, a strong linkage to the international PISA scale could not be established as 40% of the items in reading (35 of 87) were assigned unique parameters. Viet Nam's reading results are reported in this volume with an annotation.
- In **Jordan**, and in the context of the country's transition from a paper- to a computer-based assessment, strong comparability of 2022 results in reading and science to the international scale could only be established by assigning new item parameters to most link items, and thus at the expense of trend comparability. For this reason, the Adjudication Group recommended limiting trends comparisons for Jordan to mathematics results.

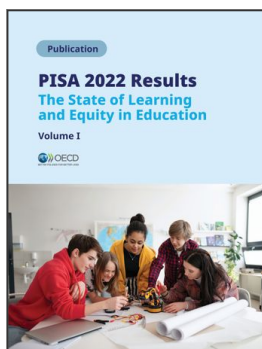
Nine other countries/economies, listed below, also did not meet one of the sampling standards, but the Adjudication Group did not judge these deviations to be consequential: Sweden (overall exclusion rate: 7.4%); Norway (overall exclusion rate: 7.3%); Lithuania (overall exclusion rate: 6.7%); Estonia (overall exclusion rate: 5.9%); Switzerland (overall exclusion rate: 5.8%); Türkiye (overall exclusion rate: 5.6%); Croatia (overall exclusion rate: 5.4%); Malta (student response rate: 79%); and Chinese Taipei (school response rates: 83% before replacement, 84% after replacement). No annotation is included when reporting data for these countries/economies in the international report.

While this could not be attributed to violations of the technical standards, the Adjudication Group also reviewed additional analyses conducted for Iceland and Norway, which reported that some students who were taking the test on Chromebooks experienced difficulties moving through the cognitive assessment due to overload on the PISA Consortium's server. While the PISA Consortium solved this problem during the testing period, 579 students in Iceland (17.2% of the final student sample, unweighted) and 584 students in Norway (8.8%) were assessed on Chromebooks before the problem was solved. According to Iceland, test administrators reported the issue having affected *at most* 13% of the unweighted final sample (438 students). The Adjudication Group reviewed the results of the additional analyses conducted by the PISA Consortium and confirmed that, overall, the data, including those of students who sat the test in these circumstances, were considered to be fit for reporting as their responses did show good fit with the model, and were not remarkably different from the performance of students in other schools. However the group noted that it is not possible to exclude the possibility that the issue affected students' engagement and

motivation to give their best effort when they sat the test. See *PISA 2022 Technical Report* (OECD, forthcoming^[1]) for details.

References

OECD (forthcoming), *PISA 2022 Technical Report*, OECD Publishing, Paris. [1]



From:
PISA 2022 Results (Volume I)
The State of Learning and Equity in Education

Access the complete publication at:
<https://doi.org/10.1787/53f23881-en>

Please cite this chapter as:

OECD (2023), "Quality assurance", in *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/3571108a-en>

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>.