

Annex A5. How comparable are the PISA 2022 computer- and paper-based tests?

Computer-based administration (CBA) was the primary mode of delivery for PISA 2022. In PISA 2022, 77 out of 81 countries and economies took the CBA version of the PISA test. Four countries (Cambodia, Guatemala, Paraguay, and Viet Nam) used a paper-based version of the assessment (PBA). This annex describes the differences between paper- and computer-based instruments, and what they imply for the interpretation of results.

Differences in test administration

Starting with the 2015 assessment cycle, the PISA test was delivered mainly on computers. Existing tasks were adapted for delivery on screen; new tasks items were developed that made use of the affordances of computer-based testing. The computer-based delivery mode allows PISA to measure new and expanded aspects of the domain constructs. In mathematics, new material for PISA 2022 included items developed to assess mathematical reasoning as a separate process classification, and items that leveraged the use of the digital environment (e.g. spreadsheets, simulators, data generators, drag-and-drop, etc.). A mixed-design that included computer-based multistage adaptive testing was also adopted for the mathematics literacy domain to further improve measurement accuracy and efficiency, especially at the extremes of the proficiency scale (on adaptive testing, see Annex A9 of this report and in *PISA 2022 Technical Report* (OECD, Forthcoming^[1])).

Paper-based assessment instruments in each domain comprise a subset of the test-items included in the computer-based version of the tests in prior cycles. In PISA 2022, a paper-based version of the assessment that included only trend units was developed for the four countries that chose not to implement the computer-delivered survey (i.e. “old” PBA). However, only one participant (Viet Nam) used the same paper-based materials as in the 2015 and 2018 cycles (based on items that were first used in PISA 2012 or earlier). The other paper-based participants administered a “new” PBA instrument that was first used in the PISA for Development (PISA-D) assessment. This “new” paper-based instrument contained a substantial amount of material that was first used in the PISA 2015 computer-based tests, or taken from other assessments, including the “Literacy Assessment and Monitoring Programme” (LAMP), the OECD Survey of Adult Skills (PIAAC), and PISA for Schools.

Table I.A5.1 presents differences in the computer- and paper-based assessments in PISA 2022, respectively. All new items for mathematics were developed as computer-based items. No new items were developed for science or reading in PISA 2022.

Table I.A5.1. Differences between computer- and paper-based administration in PISA 2022

Domain	Computer-Based Administration			Paper-based Administration	
	New	Trend	Total CBA items	PBA instrument used by Viet Nam	New instrument used by Cambodia, Guatemala and Paraguay
Mathematics	Adaptive design: 160 items	Adaptive design: 74 items	234	6 clusters: 71 items Same set of items that PBA participants used in 2018 and 2015 These items were all taken from the 2012 cycle	4 clusters: 63 items
Reading	No new item development for 2022	Adaptive design: 197 items	197	6 clusters: 87 items Same set of items that all PBA participants used in 2018 and 2015 Prior to 2015, these items were last used in 2012 and 2009	4 clusters: 66 items
Science	No new item development for 2022	6 clusters: 115 items (76 from the 2015 cycle; 39 used prior to 2015)	115	6 clusters: 85 items Same set of items that all PBA participants used in 2018 and 2015 Prior to 2015, these items were last used in 2012, 2006 and 2003	4 clusters: 66 items

Source: OECD, PISA 2022 Database; PISA 2022 Technical Report (OECD, Forthcoming⁽¹⁾).

Comparability of computer-based and paper-based tests

In order to ensure comparability of results between the computer-delivered tasks and the paper-based tasks that were used in previous PISA assessments (and are still in use in countries that use paper instruments), for the test items common to the two administration modes, the invariance of item characteristics was investigated using statistical procedures.

Most importantly, these included a randomised mode-effect study in the PISA 2015 field trial that compared students' responses to paper-based and computer-delivered versions of the same test items across equivalent international samples¹. The goal was to examine whether test items presented in one mode (e.g. paper-based assessment) function differently when presented in another mode (e.g. computer-based assessment). Results of the mode-effect study showed that for the majority of items, the results supported the comparability across the two modes of assessment (i.e. there were very few samples with any significant differences in difficulty and discrimination parameters between CBA and PBA). For some items, however, the computer-delivered version was found to have a different relationship with student proficiency from the corresponding, original paper version. Such tasks had different difficulty parameters (and sometimes different discrimination parameters) in countries that delivered the test on computer. In effect, this partial invariance approach both accounts for and corrects the potential effect of mode differences on test scores.

Table I.A5.2 shows the number of anchor items that support the reporting of results from the computer-based and paper-based assessments on a common scale. The large number of items with common difficulty and discrimination parameters (i.e. "scalar invariant") indicates a strong link between the scales. This strong link corroborates the validity of mean comparisons across countries that delivered the test in different modes.

At the same time, Table I.A5.2 also shows that a large number of items used in the PISA 2022 computer-based tests of reading and, to a lesser extent, science, were not delivered on paper. Caution is therefore required when drawing conclusions about the meaning of scale scores from paper-based tests, when the evidence that supports these conclusions is based on the full set of items. For example, the proficiency of students who sat the PISA 2022 paper-based test of mathematics should be described in terms of the PISA 2012 proficiency levels, not the PISA 2022 proficiency levels. This means, for example, that even though PISA 2022 developed a description of the skills of students who scored below Level 1b in mathematics, it remains unclear whether students who scored within the range of Level 1c on the paper-based tests have acquired these basic mathematics skills.

Table I.A5.2. Anchor items across paper- and computer-based scales

Scalar-invariant, metric-invariant and unique items in PISA 2022 paper and computer tests

	Mathematics	Reading	Science
Items with common difficulty and discrimination parameters across modes (scalar invariant)	38	25	29
Items with common discrimination parameters across modes, but distinct difficulty parameters (metric invariant)	23	20	10
Items with mode-specific parameters (distinct difficulty and discrimination parameters across modes)	1	0	0
Items not delivered on computer (paper-based only)	9	4	6
Items not delivered on paper (computer-based only)	172	152	76

Note: The table reports the number of scalar-invariant, metric-invariant and unique items based on international parameters. In any particular country, items that receive country-specific item parameters (see Annex A6) must also be considered.

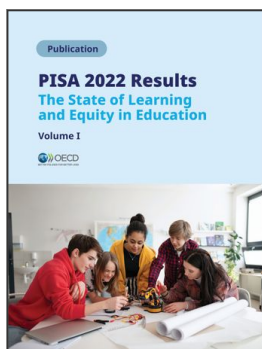
Source: OECD, PISA 2022 Database; *PISA 2022 Technical Report* (OECD, Forthcoming^[1]).

Notes

¹ For trend items included in the new PISA 2022 PBA instrument, the equivalence across modes was tested in the context of PISA for Development by using CBA item parameters as starting values in scaling.

References

OECD (Forthcoming), *PISA 2022 Technical Report*, PISA, 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), “How comparable are the PISA 2022 computer- and paper-based tests?”, in *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/69e64ebb-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>.