

**STATISTICS AND DATA DIRECTORATE
COMMITTEE ON STATISTICS AND STATISTICAL POLICY**

Cancels & replaces the same document of 3 June 2020

CSSP Review of the Technical Standards for PISA 2021

**17th meeting of the Committee on Statistics and Statistical Policy
24-26 June 2020**

This meeting will take place in remote format.

This note will be presented FOR DISCUSSION AND APPROVAL under Agenda item 4.

This note summarises the key recommendations on the PISA Technical Standards that emerged from the review undertaken by CSSP following a request by the PISA Governing Board to CSSP in June 2019. This note builds on a more comprehensive review conducted by Rory Fitzgerald and Eric Harrison, City University London, which is included as Annex A to this note.

These recommendations were informed by comments made by members of an ad hoc reviewing group nominated by CSSP, which included representatives from Canada, the Czech Republic, France, Japan, Korea, Mexico, Romania, Turkey, the United Kingdom and the United States. Annex B includes the comments provided by the US member of the CSSP reviewing group, suggesting a specific model for re-purposing, completing and integrating the various pieces of guidance provided by a revised set of PISA Technical Standards.

CSSP Delegates are asked to:

- Approve the draft report attached to this note for transmission to the PISA Governing Board;
- Discuss CSSP's future role in OECD projects of a similar nature.

Contacts: Mr Marco Mira D'Ercole, SDD Head of Division (Marco.Mira@OECD.org),
Mr Fabrice Murthin, Well-being & Household Conditions, Head of Section
(Fabrice.Murthin@OECD.org)

JT03462571

1. Background

1. PISA is one of the most well-known and well-established instruments for the assessment of the skills and competences of students towards the end of their compulsory schooling. The tool combines an assessment of what students know in different fields and a number of questionnaires completed by students and the school principal (as well as, in some countries, by parents and teachers), to assess the context in which students' learning occurs. PISA, which is conducted every three years, was first fielded in 2000 and its country coverage has almost doubled since then. It covers three 'regular' assessments fields (reading, mathematics and sciences, one of them in more depth in each round) as well as 'additional' fields on a rotating basis. PISA is also one of the few OECD statistical activities in which the Secretariat (in partnership with other actors in the PISA process and under the authority of the PISA Governing Board, a "Level 1" body of the OECD Council) is responsible for managing the full cycle of data production, from the conception of the instrument to the collection and dissemination of data and the analysis of results.

2. PISA is a complex endeavour, with different actors involved at different phases of the data cycle. Within this process, Technical Standards help coordinate the activities of the Secretariat, its contractors and National Centres; they represent a mutual commitment by participating entities to implement PISA in ways that ensure the comparability of the resulting data. They complement other methodological documents such as the guidelines set by the PISA international contractors or the manuals for national project managers. The PISA Governing Board is responsible for setting the 22 entities of Technical Standards - grouped as 'data standards' (15), 'management standards' (5) and 'national involvement standards' (2) - that participating countries and other jurisdictions are asked to follow when participating in the project. PISA National Centres implement their data collection according to these standards, while PISA Technical Advisory Group (TAG) adjudicates the data collected as 'fit for purpose', based on the adherence to these standards of the collection practices followed by National Centres. Since 2006, a Sampling Referee participates in the TAG's data adjudication meeting to assess the extent to which violations of (sampling) standards may affect the comparability of the indicators drawn from the samples. Countries that fail to meet some of the PISA Technical Standards may still be included in the results disseminated by the OECD based on additional evidence they submit, for example on non-response biases.

3. In June 2019, the PISA Governing Board (GB) wrote to CSSP asking the committee to provide an opinion on the appropriateness of these Technical Standards in relation to those used for comparable large-scale surveys. This written evaluation, once approved by CSSP and transmitted to the PISA Governing Board, will inform the revision of existing standards and the preparation of new standards to guide the 2024 data collection. Following CSSP's positive response to this request, the Secretariat has undertaken this task with the support of two consultants and the creation of a Reviewing Group nominated by CSSP, made up of country representatives from NSOs (Canada, Mexico, Romania, Turkey) and educational institutions with responsibility in the collection, dissemination and analysis of educational statistics (Czech Republic, France, Japan, Korea, the United Kingdom, the United States). This Reviewing Group discussed the outline and the first draft of the Secretariat Review on three occasions. Annex A is the full-scale Review prepared by the consultants, while the timeline of this assessment is described in Appendix B of the Review. EDU staff supported this process by answering questions and by providing access to various internal documents, as well as through its own self-assessment of PISA conducted in the context of the regular OECD Quality Reviews undertaken for all OECD

statistical activities. The review was also informed by interviews conducted with members of the PISA National Centres in 10 countries. It was prepared over a compressed time-span (5 months) in order to allow discussion by CSSP at its June 2020 meeting.

2. Main conclusions

4. While PISA has a specific goal (i.e. assessing the competences of students towards the end of compulsory schooling, and their capacity to apply these competences in real world situations with a view to analysing education systems) it shares features common to most other cross-national data collections. Given these features, this review assessed the 22 entities that make up the PISA technical standards by re-organising them into a sequence that tracks the ‘survey life-cycle’, i.e. according to their chronological completion. Standards were evaluated in terms of their clarity, purpose and the degree to which they can be assured and verified; these standards were assessed against a benchmark provided by the 8th Task Force Report on Quality in Comparative Surveys recently completed under the aegis of the World Association for Public Opinion Research (WAPOR) and its US counterpart (AAPOR).

5. This review assessed each of the 22 entities of the PISA technical standards, providing comments and recommendations on their scope and language in the tables in Appendix A of the report. The overall assessment of this review is that PISA has standards that, when fully met, allow for cross-national inferences to be made with a high degree of confidence. In general, the data collected in accordance with the standards specified are likely to yield observations that are representative of the target population (e.g. in terms of probability sampling and response rates) and should also achieve high levels of cross-national equivalence (e.g. through an extensive translation processes). However, the technical standards do not cover all activities that determine the nature of inferences that can be drawn from the data. For example, confidence in point-estimate comparisons across countries also depends on the quality of the source instruments and on the use of appropriate methods to analyse and score the data collected; yet the technical standards are silent about the development of assessment and questionnaire instruments, or about the statistical methods used in analysing the data.

6. The review makes three general recommendations:

- *Clarify the role and usage of technical standards:* the Technical Standards document could be improved by clarifying its purpose and intended audience, and perhaps even its title.
- *Give more even coverage to all areas of the survey life-cycle:* the current standards are heavily skewed towards sampling, creating the impression that non-sampling areas such translation and coding, and compliance to them, are less important.
- *Sharpen the specification and improve the drafting of the standards:* in several places, criteria for standards fulfilment are vague and unspecified. The specification of technical standards can be sharpened by adding more details within the document, by explicitly cross-referencing other pieces of PISA documentation that provide such details, and by improving the drafting of some sections (e.g. par. 19 on ‘quality assurance’ in the Technical Standards document).

7. Beyond these general recommendations, the review identified further issues that will require the attention of the PISA Governing Board, including that the technical standards should illuminate the criteria for admitting a country more clearly. The review also recommends that specific steps should be considered within each of four phases of the lifecycle of the instrument. In particular, the review recommends changes or amendments to 12 of the 22 entities of the PISA technical standards.

- *Sampling.* Expand the sampling standards to include greater controls or checks over the process of school substitution and exclusion.
- *Instrument design and development.* Standards for the central design of the instruments should be added or specified separately. The aim of the field tests and pilots should also be clarified and the design of the pilot adapted as required. The standards should prescribe a preferred mode of administration, and offer guidance on minimising mode effects between computer- and paper-based modes. If computer-mode surveys become the benchmark, paper-mode surveys should be phased out rapidly.
- *Fieldwork and implementation.* The timetable in the standards should allow jurisdictions to conduct the tests at a point in the school year which can maximise participation and response rates. Policies on using incentives for participation should be clarified, and their use encouraged in contexts where response rates are low. The standards should set upper and lower boundaries for the length of tests, as variations in respondent burden represent a potential threat to data quality. The standards should also explicitly address issues of data protection.
- *Data coding and preparation.* Rules for setting data to ‘missing’ in cases of low response rates should be clarified.

8. While most of the above recommendations apply to the ‘data cycle’ of PISA, the most important concern raised by the review possibly relates to what is *not covered* by the existing standards. Three key issues have emerged:

- *Quality assurance of instrument:* further research work is warranted to establish the validity of the interpretations of test results in terms of the underlying latent construct(s) it aims to measure. This is especially relevant for instruments related to innovative fields. To the extent that the PISA technical standards have the ambition to guide the work of all actors involved in PISA, the review recommends that standards applying to the development of the instruments (i.e. psychometric and cross-national equivalence) and questionnaires should be added to the existing Technical Standards.
- *Sub-national surveys:* standards should also cover more detailed guidelines with regard to sub-national reporting. These should be developed in line with more general OECD guidance on the dissemination of sub-national data.
- *External review:* it is also recommended that the OECD considers establishing an external group to review regularly the Technical Standards in consultation with contractors and National Programme Managers, which could become a permanent part of the PISA review cycle.

Annex A. Review of PISA Technical Standards

1. Introduction and background

1. In June 2019, the Governing Board of the Programme for International Student Assessment (PISA) commissioned a quality review of the programme, to be undertaken by the OECD Committee for Statistics and Statistical Policy (CSSP). The CSSP review was supported by a reviewing group consisting of members of the OECD Secretariat (Statistics and Data Directorate/SDD), experts from OECD countries' National Statistical Offices (NSOs), agencies with responsibilities in educational statistics or academia, willing to contribute on a voluntary basis, as well as two external consultants (see Appendix A for more details on the composition of the Reviewing Group and on the consultation process that was followed). Members of the OECD Secretariat's PISA team (Education Directorate/EDU) and EDU's Lead Methodologist have participated in the group's meetings to provide all necessary clarification and explain the context of the review, but did not intervene in the preparation of the Review

2. This Review focuses on PISA Technical Standards, which play an important role within the complex data production cycle (see Appendix B for an overview of PISA key participants and timeline). For instance, PISA National Centres implement data collection according to these standards, while PISA Technical Advisory Group adjudicates the data collected as 'fit for purpose', based on the adherence of the collection practices followed by National Centres to these standards. Technical Standards were described in the PISA [Technical Report](#) (OECD, 2015) and were revised in 2018 for use in the PISA 2021 survey ([OECD, 2018](#)). The 2018 revision lists 77 standards organised in 22 entities belonging to three broad categories (Appendix C).

3. Against this background, this report has three principal aims:

- To examine the technical standards to assess the extent to which they are fit for the purpose of PISA;
- To place these standards within the context of best practice in cross-national data collection worldwide;
- To examine the PISA organisational and procedural structures to assess the extent to which they support and help to deliver those technical standards.

4. In producing the report, we have drawn on a number of sources of information:

- Discussions with and advice from members of the CSSP review group and the OECD Education Directorate;
- The PISA Technical Standards for 2021 produced by OECD;
- The National Project Manager's Manual;
- The 2015 Technical Report;

- Skype Interviews with contractors and national project managers;
- Other background documents, e.g. the draft PISA Quality Review self-assessment;
- Report of the AAPOR/WAPOR Task Force Report on Quality in Comparative Surveys.
- Specifications for cross-national surveys such as the European Social Survey

5. Much of the analysis conducted for this report focuses on individual PISA standards, individual omissions and the overall collective impact of those standards. However, the standards are contained within a larger Technical Standards document, which is itself just one of many PISA documents dealing with the methodological issues associated with the overall programme. This report reviews and evaluates the value of the Technical Standards document as a whole, not just the individual standards.

6. The report is organised as follows. Section 2 presents the main characteristics of PISA, identifies features it shares with all exercises in cross-national data collection, and highlights its own distinctive technical and management challenges. Sections 3 to 6 review the standards by re-organising them into a sequence that tracks the ‘survey life-cycle’, in other words according to their chronological completion. The report identifies four broad areas: i) sampling; ii) instrument design and development; iii) fieldwork implementation; and iv) data preparation and deposit. In every case the report evaluates a standard in terms of its clarity, its purpose and the degree to which it can be assured and verified. Section 7 (Structure of Delivery) takes a more holistic approach by looking at the way that all the players in the PISA structure interact to ensure compliance with standards. Section 8 presents some conclusions, drawing attention to those standards that might benefit from clarification, those where requirements might be loosened/tightened, and areas which appear to be underdeveloped or absent in the current PISA standards. A summary health-check of the standards is included in Appendix A along with a series of recommendations for further consideration.

2. The purpose of PISA

7. PISA is a survey of 15-year-old students carried out every three years around the world. The survey assesses directly student ability in reading, science and mathematics, while also collecting contextual information about students’ backgrounds, their attitudes to learning and life in general. It also gathers information about the school context through optional questionnaires for teachers and/or parents. The survey was first fielded in 2000 (in 32 countries) and in 2021 will cover more than 80 countries. Initially a pencil and paper test, the PISA assessment is now overwhelmingly administered through a computer-based delivery platform. PISA’s main goal is “to provide valid and reliable indicators of student skills at the level of national educational systems, which can be compared across systems and over time and related to key factors which shape their development”.

8. Each wave of PISA has its own unique goals and characteristics. PISA has many distinctive features that mark it out from other cross-national data exercises.

- Firstly, it is not solely gathering information from respondents in the conventional sense of inviting them to offer opinions, report behaviour and relate their preferences. Its primary function is as a ‘test’, an exercise to gauge pupils’ ability across a range of disciplines, and to assess (by implication) the effectiveness of the educational system within which they are studying.

- Secondly it is based on a sample of school children. For both practical and ethical reasons, these respondents are recruited through the schools they attend. This involves a nested sample design (pupils within schools) and makes full analysis of the resulting data more complex.
 - Thirdly, the data collected are explicitly measuring performance on the part of pupils, the schools they attend, and the system within which those schools are embedded. This makes comparisons over time and across countries even more political than most other international comparisons. High levels of validity and cross-national comparability are therefore essential.
9. And yet in other ways PISA is very like many other cross-national surveys with which we are familiar. It identifies a target population of interest; it then identifies a sampling frame that captures a very high proportion of that population; a sample is drawn and those within it are approached and invited (or instructed!) to take part. The co-operating respondents are then in receipt of a series of instruments, which can be lists of questions, statements, or other stimuli such as, in this case, a cognitive task for them to perform. The data then passes through a process of coding, cleaning and checking, before being made available for analysis. Like most data collection exercises of comparable size and complexity, it has a well-defined organisational structure, comprising a governing body, a management group, a consortium of fieldwork contractors, national representatives responsible for implementation on the ground, and a set of expert advisors drawn from all the relevant fields. Given these characteristics, it seems well justified to describe, evaluate and benchmark PISA within the frame of comparative social survey research.

3. Sampling

10. All cross-national, and for that matter national, surveys, rely on taking samples from a larger population of interest, collecting data from that sample, and using them to draw inferences about the nature of that population. PISA is no different in this regard, though it faces particular design challenges. In cross-national studies, a balance needs to be struck between ensuring that sampling procedures are compliant with the central specification (and therefore equivalent), and that they best reflect the specifics of the situation of different jurisdictions. These procedures are detailed in Chapter 4 of the PISA Technical Report.

11. Overall the PISA Technical Standards are rather detailed on sampling underlining the strong emphasis put on this area by OECD. In general, the schools-based approach is seen as appropriate to meet the measurement aims of PISA. However, some of the implementation should be tightened and where necessary reflected in the standards.

3.1. Assessment of individual standards

- **Standard 1.1.** *The PISA Desired Target Population is agreed upon through negotiation between the National Project Manager and the international contractors within the constraints imposed by the definition of the PISA Target Population. The Target Population for PISA starts with students attending all educational institutions located within the country, and in grade 7 or higher. The “standard” PISA target population is further refined to its age basis: students between 15 years and 3 (completed) months and 16 years and 2 (completed) months at the beginning of the testing period*

- **Standard 1.2.** *Unless otherwise agreed upon only PISA-eligible students participate in the test.*

12. These two standards define the PISA target population (i.e. eligible students). TS1.1 is clear and does not need amendment. However, we draw attention to the following. There is a clear rationale for harmonising the upper age boundary for participants - to capture students prior to “lower secondary” exams. This is most straightforward when country educational systems are broadly aligned with each other. However, insisting on a specific age (the ‘standard’ PISA target population is further refined to its age basis, i.e. students between 15 years and 3 completed months, and 16 years and 2 completed months at the beginning of the testing period) rather than equivalent school years causes implementation challenges. To participate in PISA, schools may have to assemble a group of students from two or more grades. Other educational surveys that do not adopt this approach. However, such a fundamental shift in the definition of the target population would probably do more damage to the time series than it would offer in improved operations. We therefore recommend no change.

13. TS1.2 is clear as written, but the term “eligible” needs more precision. Given the enormous variation across the participating countries, we feel that the criteria for deciding eligibility, for instance in relation to recently arrived migrants and those whose first language is not the native tongue, should be specified and implemented in a more consistent manner.

- **Standard 1.4.** *Schools are sampled using agreed upon, established and professionally recognised principles of scientific sampling.*

14. This standard identifies school as the basic unit of the PISA sampling. What is not explicitly recognised by the standard is that there are trade-offs between sample design and the final response rate - the degree to which a reported response rate includes ‘substitutes’ should be more transparent when reporting. The use of incentives to favour participation may need greater codifying, as it appears to vary between countries. Care should also be taken to ensure there is not excessive flexibility in terms of school selection in nested designs. Greater efforts should be made to reduce school substitutions in some countries. School and student-level incentives could facilitate participation but this must be handled carefully.

- **Standard 1.5.** *Student lists should not be collected more than 8 weeks prior to the start of data collection, unless otherwise agreed upon.*
- **Standard 1.6.** *Students are sampled using agreed upon, established and professionally recognised principles of scientific sampling and in a way that represents the full population of PISA-Eligible students.*

15. These standards define the period between the moment when participating students are identified and when the testing takes places, and provide general guidance on the sampling of students. This review does not suggest any amendments to these two standards, but given the devolution of the sampling frame construction to country-level it recommends that this division of labour between the contractor and the National Project Manager should be clearly specified and enforced.

- **Standard 1.7.** *The PISA Defined Target Population covers 95% or more of the PISA Desired Target Population. That is, school level exclusions and within school exclusions combined do not exceed 5%.*

16. This standard sets an upper limit to the number of students selected for participation in PISA that do not take the testing. The 5% exclusion value (like all such targets) is somewhat arbitrary and it does not take full account of the wide variations in the way education is organised across countries. Sampling from an educational system is quite different from drawing from a register or an address file. Greater thought should be given to how to realise this standard in a functionally equivalent way, with particular reference to the proportions of different excluded categories e. g. those with a language barrier, those with special educational needs, etc.

- **Standard 1.8.** *The student sample size for the computer-based mode is a minimum of 6 300 assessed students, and 2 100 for additional adjudicated entities, or the entire PISA Defined Target Population where the PISA Defined Target Population is below 6 300 and 2 100 respectively. The student sample size of assessed students for the paper-based mode is a minimum of 5 250.*

17. This standard sets the minimal size of the sample of participating students for different modes of the survey. It is not possible to evaluate adherence to these targets based on the Technical Standards document per se. These sample sizes are presumably set so as to produce error margins for particular point estimates of a given size. It would be useful to specify in the Technical Standards what is the desired margin of error, rather than simply provide the target sample number.

- **Standard 1.9.** *The school sample size needs to result in a minimum of 150 participating schools, and 50 participating schools for additional adjudicated entities, or all schools that have students in the PISA Defined Target Population where the number of schools with students in the PISA De-fined Target Population is below 150 and 50 respectively. Countries not having at least 150 schools, but which have more students than the required minimum student sample size, can be permitted, if agreed upon, to take a smaller sample of schools while still ensuring enough sampled PISA students overall.*
- **Standard 1.10.** *The minimum acceptable sample size in each school is 25 students per school (all students in the case of school with fewer than 25 eligible students enrolled).*

18. Additional adjudicated entities refer to any geographical units below the level of the nation-state which are treated as self-contained units for the purpose of implementing PISA. Consideration should be given to varying this number of participating schools based upon the complexity of the sample design, for instance the nature of the school system, the type of stratification, etc. 150 schools per country might not always have the same meaning in terms of ‘effective’ representation. In any event the technical standards should justify (in-text or by reference) why this figure has been set.

- **Standard 1.16.** *Unless otherwise agreed upon, the international contractors will draw the school sample for the Main Survey.*
- **Standard 1.17.** *Unless otherwise agreed upon, the National Centre will use the sampling contractor’s software to draw the student sample, using the list of eligible students provided for each school.*

19. These standards set limits to the number of participating schools for each jurisdiction covered by PISA, as well as on the student size of each school. In order to ensure a fully input-harmonised and consistent approach, it is important that both the samples of schools and the samples of students should always be drawn by the international

contactor with oversight of sampling across PISA. If this is not possible, very clear guidelines or procedures should be issued to countries; if countries are to do this themselves, there must be the capacity for central verification and back checking.

20. Consideration should also be given to mandating the centralisation of this process, if only to remove any doubt about the potential for manipulation of the process, and to ensure the validity of results. In this respect, the AAPOR-WAPOR guidelines state in regard to household level surveys: “those surveys that involve interviewer selection of respondents are more likely to produce a gender biased sample (with disproportionately more females), compared to samples based on registers” (p. 34).

3.2. Summary

21. Sampling is the most important area of the technical standards. First, because the principle that one can generalise to a wider population with confidence is the cornerstone of the survey method. Second, sampling is the easiest area of the standards for which to quantify ‘quality thresholds’. It is therefore easier to use as a threshold for inclusion in the final data release. Other areas should be brought up to the level of sampling specification in order to improve the technical standards rather than reducing the levels required here.

22. Serious consideration should be given to tightening implementation especially regarding the selection of schools and in terms of definitions regarding exclusions in an equivalent manner across countries. Where sub-regional comparisons are important adjudicated region procedures should be adopted where official reporting will be at that level.

4. Instrument design and development

23. The second stage of the survey lifecycle is the development of the instruments to be administered to respondents. This review uses the term ‘instruments’ in the same way as it is used in the PISA Technical Standards document, i.e. as a generic term to refer to any material used in the field to gather data. In the case of PISA, this comprises the student tests, the student questionnaires, and the questionnaire for teachers.

24. Overall the technical standards in this area might become more comprehensive. The process of developing the instruments could be included and the purpose and scope of the field tests defined more thoroughly.

4.1. Assessment of individual standards

- **Standard 2.1.** *The PISA test is administered to a student in a language of instruction provided by the sampled school to that sampled student in the major domain (Mathematics) of the test.*

If the language of instruction in the major domain is not well defined across the set of sampled students then, if agreed upon, a choice of language can be provided, with the decision being made at the student, school, or National Centre level. Agreement with the international contractor will be subject to the principle that the language options provided should be languages that are common in the community and are common languages of instruction in schools in that adjudicated entity.

If the language of instruction differs across domains then, if agreed upon, students may be tested using assessment instruments in more than one language on the condition that the test language of each domain matches the language of instruction for that domain. Information obtained from the Field Trial will be used to gauge the suitability of using assessment instruments with more than one language in the Main Survey.

In all cases the choice of test language(s) in the assessment instruments is made prior to the administration of the test.

25. No amendments are suggested to TS2.1.

- **Standard 3.1.** *PISA participants participating in the PISA 2021 Main Survey will have successfully implemented the Field Trial. Unless otherwise agreed upon:*
 - *A Field Trial should occur in an assessment language if that language group represents more than 5% of the target population.*
 - *For the largest language group among the target population, the Field Trial student sample should be a minimum of 200 students per item.*
 - *For all other assessment languages that apply to at least 5% of the target population, the Field Trial student sample should be a minimum of 100 students per item.*
 - *For additional adjudicated entities, where the assessment language applies to at least 5% of the target population in the entity, the Field Trial student sample should be a minimum of 100 students per item.*

26. The aim of a field trial is generally to troubleshoot and ensure readiness for the main stage. In this respect, the AAPOR-WAPOR guidelines state “When designing the questionnaire and other survey materials, researchers must attempt to identify and be informed by ways in which members of different cultures may differ systematically in how questions are understood and answered. Understanding the population of interest and thorough pretesting are essential for the identification of potential problems with design considerations and instruments in order to avoid results plagued by measurement and nonresponse error”.

27. The technical standards should define the meaning of “successfully implemented”. Is this simply a feasibility check or a full-blown experiment assessing the quality of items (particularly new modules)? The former could be defined as the ‘weak test’, where the trial is judged a success if there are no major hitches in the implementation and where no major errors in items are identified. The latter would describe a formal ‘control’ trial, where different versions of items are fielded or their running order is randomised, with the express objective of identifying the best approach. There are trade-offs between limiting field test length (to reduce burden) and undertaking a realistic rehearsal (keeping length closer to the eventual main stage). The overall length of the field tests should be kept under review and limited to what is strictly necessary.

- **Standard 4.1.** *The majority of test items used in previous cycles will be administered unchanged from their previous administration, unless amendments have been made to source versions, or outright errors have been identified in the national versions.*

28. This reads more like a statement of policy or a description of an outcome, rather than a technical standard. It could be reworded as “should be administered”, or it may be superfluous here.

- **Standard 4.2.** *All assessment instruments are equivalent to the source versions. Agreed upon adaptations to the local context are made if needed.*

29. The AAPOR-WAPOR guidelines state: “Approximately 20 years ago, a review of this multi-disciplinary literature concluded that literally dozens of forms of equivalence were being discussed in practice (Johnson, 1998). These discussions often employed different terms to denote the same underlying concept, and also used similar terms to reference differing equivalence concepts. Since that time, the variety of conceptualizations of equivalence in this literature has continued to expand, to more than 90 as of today” (see Johnson (2019), p. 14).

30. Again, this should be worded as an aspiration. Equivalence cannot be fully evaluated until after the fact and in relation to data. Equivalence is an ambiguous term, as noted in the AAPOR-WAPOR guidelines. Many Multinational, Multiregional, and Multicultural surveys (3MC) use the shorthand “ask the same question” to indicate equivalence in translations. Language differences can result in variability in the time taken to administer instruments. It should be made clearer what is intended. A fuller indication of the level of adaptations permitted could be included as a footnote or an external cross-reference. Care should be taken to ensure adaptations to local circumstances are not disproportionate and are centrally monitored.

- **Standard 4.3.** *National versions of questionnaire items used in previous cycles will be administered unchanged from their previous administration, unless amendments have been made to source versions, outright errors have been identified in the national versions, or a change in the national context calls for an adjustment.*
- **Standard 4.4.** *The questionnaire instruments are equivalent to the source versions. Agreed upon adaptations to the local context are made if as needed.*

31. The AAPOR-WAPOR guidelines state “Cross-cultural validity should be established for questionnaires designed to compare data (Fitzgerald & Zavala-Rojas, 2020; Smith, 2004). However, common practice frequently avoids measurement equivalence testing, or equivalence is only tested for a limited selection of items of a questionnaire” (p. 41).

- **Standard 4.5.** *School level materials are equivalent to the source versions. Agreed upon adaptations to the local context are made as needed.*

32. No amendments are suggested, although TS4.4 seems very similar to TS4.2. In general (as above), these seem to be statements of policy rather than standards for compliance, given that the criteria for their evaluation are not presented here.

- **Standard 5.1.** *The following documents are translated into the assessment language in order to be linguistically equivalent to the international source versions.*
 - *All administered assessment instruments*
 - *All administered questionnaires*
 - *The Test Administrator script from the Test Administrator (or School Associate) Manual*
 - *The Coding Guides (unless otherwise agreed upon)*

- **Standard 5.2.** *Unless otherwise agreed upon, school level materials are translated/adapted into the assessment language to make them functionally equivalent to the international source versions.*
33. No amendments are suggested here, though we note the inconsistency of terminology. TS5.1 talks of “linguistic equivalence” while TS5.2 uses “functional equivalence”. The latter would seem to be essential in both cases.
- **Standard 6.1.** *The international contractors must test all national software versions prior to their release to ensure that they were assembled correctly and have no technical problems.*
 - **Standard 6.2.** *Once released, countries must test the national software versions following testing plans to ensure the correct implementation of national adaptations and extensions, display of national languages, and proper functioning on computers typically found in schools in each country. Testing results must be submitted to the international contractors so that any errors can be promptly resolved.*
34. No amendments are suggested to TS6.1 and TS6.2. Further amplification could be provided in a footnote or cross-reference.
- **Standard 10.1.** *Only national options that are agreed upon between the National Centre and the international contractors are implemented.*
 - **Standard 10.2.** *Any national option instruments that are not part of the core components of PISA are administered after all the test and questionnaire instruments of the core component of PISA have been administered to students that are part of the international PISA sample, unless other-wise agreed upon.*
35. The ‘mix and match’ approach to the construction of the PISA instrument menu may result in variation in the overall duration of the instruments – and this can affect both respondent engagement/fatigue as well as potentially drop-outs. Some concerns are being expressed by countries about the length of the PISA tests. Given that there is already a core and a series of opt-ins, there is a trade-off between increased national ‘ownership’ of the study and the increased burden on respondents.
- **Standard 13.1.** *All paper-based student assessment material will be centrally assembled by the international contractors and must be printed using the final print-ready file and agreed upon paper and print quality. New countries/entities must submit a printed copy of all Field Trial instruments (booklets and questionnaires) for approval of the printing quality for the Main Survey. The same printing standard must be used for both the Field Trial and the Main Survey.*
 - **Standard 13.2.** *The cover page of all national PISA test paper-based materials used for students and schools must contain all titles and approved logos in a standard format provided in the international version.*
 - **Standard 13.3.** *The layout and pagination of all test paper-based material is the same as in the source versions, unless otherwise agreed upon.*
 - **Standard 13.4.** *The layout and formatting of the paper-based questionnaire material is equivalent to the source versions, with the exception of changes made necessary by national adaptations.*

36. While the transition from paper to computer-based tests is generally viewed as a success among those we have spoken to, it is also clear that any test in computer format introduces and assumes a background level of ICT proficiency. The need for access to computer stations or devices can also introduce an element of ‘stagecraft’ into the administration of the test. Given the extensive literature on potential mode effects in data collection, and that it is not possible to unpick these from “country effects” in cross-national surveys, we strongly recommend the phasing out of paper-based tests.

- **Standard 17.1.** *National options are agreed upon with the international contractors before 1 December in the year preceding the Field Trial and confirmed before 1 November in the year pre-eding the Main Survey.*
- **Standard 17.2.** *The National Centre notifies the OECD Secretariat of its intention to participate in specific international options three months prior to the start of the translation period. International options can only be dropped between the Field Trial and the Main Survey, not added.*

37. The PISA assessment is lengthy, with accompanying risks of cognitive burden on students and possibly causing lower data quality in later parts of the tests. A review of the costs and benefits of adding additional modules would help inform a debate on this issue. Other models of administration could be considered: greater use of split samples, and the possibility of core and follow-on tests would be two options.

- **Standard 18.1.** *An agreed upon Translation Plan will be negotiated between each National Centre and the international contractors.*
- **Standard 18.3.** *Questionnaire materials are submitted for linguistic verification only after all adaptations have been agreed upon.*
- **Standard 18.4.** *All adaptations to those elements of the school level materials that are required to be functionally equivalent to the source as specified in Standard 5.2, need to be agreed upon.*
- **Standard 21.2.** *National Centres provide feedback to the international contractors on the development of instruments, domain frameworks, the adaptation of instruments, and other domain-related matters that represent the perspectives of the relevant national stakeholders.*

38. All the above standards relate to sign-off processes as part of quality assurance. It is questionable whether they really constitute technical standards.

4.2. Summary

39. Compared to sampling, this area is under specified and it is recommended to make this more comprehensive in later drafts. The process of initial instrument design is not specified nor are methods to measure equivalence. Both are clear omissions. It is strongly recommended to boost equivalence by using the same mode in all countries and phasing out paper-based methods.

5. Fieldwork and implementation

40. Fieldwork is often one of the most difficult areas to control in cross-national data collection exercises necessarily devolved from central teams and yet it can be a major source of error in case of poorly specified areas or deviations from agreed standards. Again, compared to sampling, this area is under developed and some additional flexibility is recommended in other areas.

41. This section includes all standards which relate to, or are achieved during, the field stage. Regardless of how well prior decisions about the design of the sampling strategies and the development of data collection instruments have been taken, they will amount to little if they are not correctly and consistently implemented.

5.1. Assessment of individual standards

- **Standard 1.3.** *Unless otherwise agreed upon, the testing period:*
 - *is no longer than eight consecutive weeks in duration for computer-based testing participants,*
 - *is no longer than six consecutive weeks in duration for paper-based testing participants,*
 - *does not coincide with the first six weeks of the academic year, and*
 - *begins exactly three years from the beginning of the testing period in the previous PISA cycle.*

42. The first three points here are unproblematic. However, the last bullet point, and more generally the timing of the PISA main stage, can be problematic in some countries if the testing falls just before exams or other important milestones. Consideration should be given to being more flexible as this could increase response rates at both the school and pupil level. While some flexibility on the choice of the testing period seems already offered, it could be helpful if this were applied more consistently.

- **Standard 1.11.** *The final weighted school response rate is at least 85% of sampled eligible and non-excluded schools. If a response rate is below 85% then an acceptable response rate can still be achieved through agreed upon use of replacement schools.*
- **Standard 1.12.** *The final weighted student response rate is at least 80% of all sampled students across responding schools.*
- **Standard 1.13.** *The final weighted teacher response rate is at least 75% of all sampled teachers across responding schools.*
- **Standard 1.14.** *The final weighted sampling unit response rate for any optional cognitive assessment is at least 80% of all sampled students across responding schools.*

43. A weighted response rate is referred to at 1.11-1.14 but the method of calculation is not stated in the Technical Standards. This should be specified and the location of the calculation cross-referenced. The Technical Standards should also specify the maximum amount of replacement that is allowable. Greater efforts should be made to avoid school level refusals, with replacement seen as a last resort and round on round reductions aimed for, rather than this becoming a repeated structural feature. The process of re-placement

should be monitored within strata, and within strata limits should be set (e.g. if almost all refusals are within a single strata this could have a very serious impact; even replacements within that strata might not be equivalent if there are very high refusals, for example with private schools).

- **Standard 7.1.** *Each country should have a designated PISA helpdesk with contact information provided to each of its test administrators and school coordinators.*
44. It is not clear if this is a technical standard. It is rather a technical requirement.
- **Standard 7.2.** *In countries that administer the computer-based version of PISA, the helpdesk staff must be familiar with the PISA computer system requirements applications and training materials; be familiar with all national software standards and procedures; and attend the test administrator training sessions to become familiar with the computer-based assessments and appreciate the challenges faced by schools and test administrators.*
45. The AAPOR-WAPOR guidelines state “Already, many international companies with a presence in the EU have adopted General Data Protection Regulation (GDPR) compliant policies not only for their EU branches, but across all of their countries of operation” (p. 64).
46. It is not clear if data protection issues are properly communicated and certainly this is missing from the technical standards (see below for more on GDPR).
- **Standard 8.1.** *All test sessions follow international procedures as specified in the PISA school level materials, particularly the procedures that relate to:*
 - *test session timing,*
 - *maintaining test conditions,*
 - *responding to students’ questions,*
 - *student tracking, and*
 - *assigning assessment materials.*
47. This standard is appropriate and no change is suggested.
- **Standard 8.2.** *The relationship between Test Administrators and participating students must not compromise the credibility of the test session. In particular, the Test Administrator should not be the reading, mathematics, or science instructor, a relative, or a personal acquaintance of any student in the assessment sessions he or she will administer for PISA.*
48. The standard is well specified and required. However, it is not clear how it is monitored in field implementation, and this should be clarified. Reference should also be made to forbidding coaching of students under any circumstances. Tough sanctions should be specified for infringements, up to and including exclusion from the dataset and from sub-sequent rounds of PISA.
- **Standard 8.3.** *National Centres must not offer rewards or incentives that are related to student achievement in the PISA test to students, teachers, or schools.*
49. The reason TS8.3 is important is that results are not just analysed for their social scientific interest; they can also be viewed as performance measures of students, schools and whole education systems. Gatekeepers and participants might potentially be motivated to ‘game’ or manipulate certain parts of the PISA process. While this standard forbids the use of incentives to different types of participants that related to students’ performances in the test, it glosses over the more general issues of using other types of incentives to increase

participation. Incentives to participation should be encouraged especially in countries failing to reach response rate targets although of course they should not be related to student outcomes. Clearer guidelines should be specified in the Technical Standards about the use of rewards and incentives. It appears some countries are already using them, while others are not. In some cases, they could be helpful to boost response.

- **Standard 9.1.** *Qualified contractor staff will conduct trainer training sessions with NPMs or designees on PISA materials and procedures to prepare them to train PISA test administrators.*
- **Standard 9.2.** *NPMs or designees shall use the comprehensive training materials and approach developed by the contractors and provided on the PISA Portal to train PISA test administrators.*
- **Standard 9.3.** *All test administrator training sessions should be scripted to ensure consistency of presentations across training sessions and across countries. Failure to do so could cause errors in data collection and make results less comparable.*
- **Standard 9.4.** *In-person and/or web-based test administrator trainings should be conducted by the NPMs or designees unless a suitable alternative is agreed upon.*

50. All the above standards are about training of the staff administering the tests. Training is a critical condition for the production and accurate results. Standards 9.1-9.4 are appropriate and no change is suggested.

- **Standard 9.5.** *PQMs need to successfully complete self-training materials, attend webinars to review and enhance the self-training, and attend the test administrator training, unless otherwise agreed upon.*

51. Whilst Technical Standard 9.5 is fit for purpose, the implementation is not always even and tightening of compliance should be considered.

- **Standard 11.1.** *PISA materials designated as secure are kept confidential at all times. Secure materials include all test materials, data, and draft materials. In particular: no-one other than approved project staff and participating students during the test session is able to access and view the test materials, no one other than approved project staff will have access to secure PISA data and embargoed material, and formal confidentiality arrangements will be in place for all approved project staff.*

52. This standard aims to ensure that the testing material remains confidential until the test is actually administered, which is crucial to avoid that students could receive specific training before taking the test. While this is critically important for credibility of PISA, greater details regarding protection provisions should be included in the Technical Standards. As the OECD is based in Europe, the data collection would be subject to the GDPR with the OECD (Paris) listed as the Data Controller (since OECD defines the scope and purpose of the data collection). A Data Protection Officer should also be named in the Technical Standards.

53. Other comparative social surveys rely on tighter standards in this field. For example, the European Social Survey Specification has a detailed section which outlines how compliance with the General Data Protection Regulation (GDPR) and national law is to be ensured. It outlines the requirements for its national teams and survey agencies. This includes reference to Data Processing Agreements, which must be signed before any handling of personal data takes place. It also foresees that the Data Management Plan must

be submitted for approval by the Director of ESS ERIC, and includes references to a Data Protection Officer and central record of local DPOs and details of how data breaches are to be handled and how information should be provided to respondents containing a harmonised data protection statement. In addition, details of data subject rights and how to exercise them are noted.

54. The PISA Technical Standards should be updated to reflect the GDPR.

- **Standard 12.1.** *PISA Main Survey test administration is monitored using site visits by trained in-dependent quality monitors.*
- **Standard 12.2.** *Fifteen site visits to observe test administration sessions are conducted in each PISA participating country/economy, and five site visits in each adjudicated region.*
- **Standard 12.3.** *Test administration sessions that are the subject of a site visit are selected by the international contractors to be representative of a variety of schools in a country/economy.*

55. These standards guide the monitoring of test administration sessions to ensure that students' tests are conducted in line with the general rules governing PISA. Standards 12.1-12.3 are adequate and do not need to be amended.

5.2. Summary

56. Data protection standards should be specified in more detail to ensure legal compliance. The issue of incentives should be revisited, procedures for weighted response rates better defined, and consideration of timing flexibility discussed. Furthermore, the issue of the length of the PISA instrument and its impact on respondent engagement should also be discussed for future iterations of the standards.

6. Data coding and preparation

57. Data coding and preparation is an area that often receives less attention in cross-national surveys and can be the source of major errors. For example, in early stages of the European Social Survey measures of highest qualification were poorly harmonised leading to difficulties in making cross-national comparisons; this was rectified in later waves by much greater coordination of the process. In general PISA handles this area well but the technical standards themselves are not very detailed. Some suggestions for tightening definitions and ensuring transparency to end users are made too.

58. In order to be authoritative, a survey should produce not only high quality data but be accompanied by thorough and transparent documentation for users. This is normally achieved in two ways. Firstly, there is a painstaking process of cleaning and checking the data prior to deposit. Secondly there is a set of procedures for alerting users to any variables where there are deviations from the agreed standards for harmonisation. These deviations could relate to the formulation of an item, some problem in its administration, the way it was subsequently coded, and so on. At a higher level, there might also be national deviations from any of the Technical Standards that, while not warranting exclusion from the dataset, are worth users being alerted to.

6.1. Assessment of individual standards

- **Standard 1.15.** *Analyses based on questionnaire data that do not link to a weighted 75% of the target population shall be flagged or replaced by a missing code in OECD reports.*
59. This Technical Standard defines rules for attaching ‘flags’ to specific data that are deemed to be affected by problems or deviations from the PISA standards. The standard should be clearer on the criteria that would require a ‘flag’, and what the threshold would be for a value being set to ‘missing’. This is a major distinction that the current standard fails to convey. Users may – indeed will often – choose to include flagged data regardless, but they cannot include data that is missing altogether. If the distinction between flagged and missing is not implemented consistently, this may have an effect on the overall data quality.
- **Standard 14.1.** *The coding scheme described in the coding guides is implemented according to instructions from the international contractors’ item developers.*
60. Coding is an area of considerable importance and yet the PISA Technical Standards has only a single sentence devoted to it, amounting to the requirement to follow instructions. The standards should be better specified or more specific guidance should be is-sued to NPMs and international contractors.
- **Standard 15.1.** *Each PISA participant submits its data in a single complete database, unless otherwise agreed upon.*
 - **Standard 15.2.** *All data collected for PISA will be imported into a national database using the Data Management Expert (DME) data integration software provided by the international contractors following specifications in the corresponding operational manuals and international/national record layouts (codebooks). Data are submitted in the DME format.*
 - **Standard 15.3.** *Data for all instruments are submitted. This includes the assessment data, questionnaires data, and tracking data as described in the PISA operations manuals.*
61. These standards are about the rules for submitting PISA test results and other data into national databases. We have no suggestions to changes, except to repeat previous remarks about the need for potential amplifications in footnotes and/or cross-references to other sources.
- **Standard 15.4.** *Unless agreed upon, all data are submitted without recoding any of the original response variables.*
 - **Standard 15.5.** *Each PISA participating country’s database is submitted with full documentation as specified in the PISA operations manuals.*
62. These standards set rules for reporting all PISA data in their original format supported with the required documentation. While adequate in their current formulation, the Technical Standards should provide some parameters for situations when such recoding might be permitted (e.g. for data protection purposes) - and by whom.
- **Standard 18.2.** *The following items are submitted to the international contractors in accordance with agreed timelines: the Translation Plan, a print sample of booklets prior to final printing, for new countries/entities using the paper-based*

instruments (where this is required, see Standard 13.1), results from the national checking of adapted computer-based assessment materials and questionnaires, adaptations to school level materials, sampling forms (see Standard 1), demographic tables, Completed Field Trial and Main Survey Review Forms, documents related to PISA Quality Monitors: nomination information, Test Administrator training schedules, translated school level materials, school contact information, test dates, and other documents as specified in the PISA operations manuals.

63. This standard summarises requirements for exchange of materials and information between countries and contractors, which are mentioned in other standards. In the absence of any timeframes, it is essentially a description of steps that need to be taken when coordinating a data collection exercise of this type. Consideration should be given to specifying the timeframes for the transmission of this material.

- **Standard 19.1.** *The timeline for submission of national databases to the international contractors is within eight weeks of the last day of testing for the Field Trial and within eight weeks of the last day of testing for the Main Survey, unless otherwise agreed upon.*
- **Standard 19.2.** *National Centres execute data checking procedures as specified in the PISA operations manuals before submitting the database.*
- **Standard 19.3.** *National Centres make a data manager available upon submission of the database. The data manager: is authorised to respond to international contractor data queries, is available for a three month period immediately after the database is submitted unless otherwise agreed upon, is able to communicate in English, is able to respond to international contractor queries within three working days, and is able to resolve data discrepancies.*
- **Standard 19.4.** *A complete set of PISA paper-based instruments as administered and including any national options, is forwarded to the international contractors on or before the first day of testing. The submission must include the electronic PDF and/or Word versions of all instruments.*
- **Standard 19.5.** *To enable the PISA participant to submit a single dataset, all instruments for all additional adjudicated entities will contain the same variables as the primary adjudicated entity of the PISA participant.*

64. These standards relate to data preparation by National Centres and deposit to international contractors. We make no suggestions for amendments to these standards. However, TS19.4 does not seem to fit well in this group, as it relates to, the delivery of data collection instruments prior to fieldwork, i.e. to an earlier stage of the cycle.

- **Standard 20.1.** *The international contractors will maintain a permanent electronic archive of all assessment materials, school level materials and coding guides, including all national versions. For documents that are finalised by countries, they are required to upload the latest version to the PISA Portal.*
- **Standard 20.2.** *The National Project Manager must submit one copy of each of the following adapted and translated Main Survey materials to the international contractors: electronic versions (Word and/or PDF) of all administered Test Instruments, including international and national options, electronic versions (Word and/or PDF) of all administered Questionnaires, including international*

and national options (paper-based countries only); electronic versions of the school level materials; and electronic versions of the Coding Guides.

- **Standard 20.3.** *Unless otherwise requested, National Centres will archive all Field Trial materials until the beginning of the Main Survey, and all Main Survey materials until the publication of the international report. Materials to be archived include: all respondents' paper-based test booklets and questionnaires (PBA countries) or USB drives containing completed SDS and all associated data (CBA countries) sampling forms, student lists, student tracking instruments, and all data submitted to the international contractors. After completion of a survey, the National Centre will transfer final versions of all national materials to the international contractors who will compile the national archives from all participants and transfer them to OECD after completion of the Main Survey.*

65. These standards define the reporting obligations falling upon international contractors, national Project managers and national centres. These standards are clear and require no amendments.

6.2. Summary

66. Compared to other areas of the data lifecycle, coding requirements should be made more comprehensive. The issue of analyses based on questionnaire data that do not link to a weighted 75% of the target population and how they are flagged should be attended to in later drafts. Technical Standards would become more meaningful in some cases if they had timelines attached.

7. The Structure of Delivery

67. Cross-national data collection of this type inevitably succeeds or fails based upon the cooperation between central coordinators and local deliverers. In recent decades the importance of a strong central hub has been emphasised in 3MC data collection exercises and OECD plays that role in PISA. However, it is important to ensure adequate bottom up input too to ensure wide input into the design and facilitate a common sense of ownership.

68. In addition to stating requirements for each stage of the survey lifecycle, most specifications of cross-national surveys set out procedures for its overall management, and the nature of communications between the principal players. This is of considerable importance, given the large number of 'moving parts' in complex cross-national data collection exercises. The following standards refer to this overarching aspect of quality.

7.1. Assessment of individual standards

- **Standard 16.1.** *The international contractors ensure that qualified staff are available to respond in English to requests by the National Centres during all stages of the project. The qualified staff: Are authorised to respond to National Centre queries, Acknowledge receipt of National Centre queries within one working day, Respond to coder queries from National Centres within one working day, Respond to other queries from National Centres within five working days, or, if processing the query takes longer, give an indication of the amount of time required to respond to the query.*

- **Standard 16.2.** *The National Centre ensures that qualified staff are available to respond to requests in English by the international contractors during all stages of the project. The qualified staff: Are authorised to respond to queries, Are able to communicate in English, Acknowledge receipt of queries within one working day, Respond to queries from the international contractors within five working days, or, if processing the query takes longer, give an indication of the amount of time required to respond to the query.*

69. These standards set rules for international contractors and national centres in terms of type of information that their staff should be able to provide on various aspects of the data collection exercise. They are clear and (unlike some others elsewhere in the document) quantified. It is important that compliance with these standards is firmly enforced, given the knock-on effects on other countries and contractors, and potentially on the overall quality of PISA. This is particularly pressing when the overall three year cycle is so constraining.

- **Standard 21.1.** *National Centres develop appropriate mechanisms in order to promote participation, effective implementation, and dissemination of results amongst all relevant national stakeholders.*

70. This standard describes the activities that National Centres should undertake in order to achieve the successful implementation of the project. These activities are particularly useful in countries where participation is not de facto compulsory and/or where teams are struggling to reach targets.

- **Standard 21.2.** *National Centres provide feedback to the international contractors on the development of instruments, domain frameworks, the adaptation of instruments, and other domain-related matters that represent the perspectives of the relevant national stakeholders.*
- **Standard 22.1.** *Representatives from each National Centre are required to attend all PISA international meetings including National Project Manager meetings, coder training, and any separate within school sampling training, and data management training, as necessary. Up to 6 international meetings are planned per cycle.*
- **Standard 22.2.** *Representatives from each National Centre who attend international meetings must be able to work and communicate in English.*

71. These standards regulate some very important aspects of real-time quality monitoring. As formulated, however, they read more like contract specifications than technical standards. We note the need for ongoing review of how effectively these operate. In countries where either the NPM is the member of the Governing Board, or where there is an active communication channel between the two, these standards are likely to work well. However, arrangements may not be equally joined up in other countries. Given the number of representatives, PISA face to face meetings should be supplemented with smaller, more interactive forums, possibly organised at the level of the statistical region. This requirement could be incorporated into the Technical Standards themselves, or initially trialled during the implementation of PISA 2021.

7.2. Summary

72. A review of PISA meetings and communications might be considered to ensure they function as two way processes rather than only being information giving. It should be noted that NPMs felt well supported by contractors. Ways to ensure NPMs who are not members of the Governing Board are more connected to the overall PISA process could be considered.

8. Conclusions and recommendations

73. The stated purpose of the PISA technical standards is to: “list the set of standards upon which the PISA 2021 data collection activities will be based, as was the case for previous PISA. In following the procedures specified in the standards, the partners involved in the data collection activities contribute to creating an international dataset of a quality that allows for valid cross-national inferences to be made”. More specifically they aim to ensure “consistency, precision and generalisability of the data”.

74. The overall assessment of this review is that PISA has standards that, when fully met, allow for cross-national inferences to be made with a high degree of confidence. In general, the data collected in accordance with the standards specified are likely to yield observations that are representative of the target population (e.g. in terms of probability sampling and response rates) and should also achieve high levels of cross-national equivalence (e.g. through an extensive translation processes). The extent to which such inferences can be made to the level required by the OECD (for example point estimate comparisons between countries) is however not supported by the Technical Standards as currently drafted. A number of areas should be reviewed or improved in a future draft of the standards to better align them with their stated purpose.

75. The purpose of the Technical Standards document itself should be clarified so that its intended audience is clearer. Is the primary purpose of the document to convey information about the entire PISA operation to external stakeholders and general readers? Or is to act as a set of guidelines, aimed at participating countries and their National Project Managers and national contractors regarding their data collection activities? It is currently more oriented towards the latter but this leaves key areas of the data lifecycle (notably instrument design and data reporting) without sufficiently clear standards. Two courses of action are possible. First, the document could be redrafted and renamed ‘Guidelines for Data Collection’, and the reader referred to additional documents which deal with other issues not covered. Second, the existing Technical Standards document can be re-balanced to include standards relating to every area of the PISA data lifecycle.

76. The coverage of the data collection life-cycle provided by the 2021 technical standards is uneven, and heavily skewed towards sampling. This unevenness can create an impression that non-sampling areas are less important in PISA, especially as they seem to be reviewed less often at the adjudication stage. A future draft of the standards should cover other areas such as translation or coding in more detail. Furthermore, greater effort should in future be put into developing ways to verify compliance with standards in non-sampling errors areas, allowing them to be given greater weight at the evaluation stage. This remains a challenge in cross-national survey and assessment measurement overall.

77. The completeness of the Technical Standards should be improved with a view to ensuring evenness across all areas currently included. This review has highlighted specific

cases where more detail should be provided (e.g. on the calculation of weighted response rates) or cross-referencing to other documents made (e.g. the NPM manual). While this unevenness may reflect the gradual development of the standards over time, steps should be taken review to ensure a more even coverage of the different phases of the survey life-cycle.

78. The blocks of descriptive text and bullet points between the sets of standards do not add very much value. The document could be more effective if it simply presented groups of (measurable) standards, with a short accompanying commentary and links to further and fuller background information.

79. All vague language should be removed from the document and replaced with specified targets based on either thresholds or acceptable ranges, operationalised in measurable form and expressed as numbers.

80. The document should be redesigned in an ‘online first’ format with live links to other materials in a broader compendium of PISA resources. Where there is extensive reference to external documents (e.g. the NPM manual, coding guidelines), these documents should undergo a regular process of review to ensure they remain aligned with the Technical Standards and vice versa.

81. While earlier sections have already commented on individual Technical Standards, within each of five phases of the lifecycle of the instrument there are specific steps that should be considered as a priority.

82. In the *sampling* area, greater control or checks over the process of school sampling should be introduced. In particular, we recommend greater clarity regarding school exclusion procedures. Moreover, the technical standards should be more explicit on the types of acceptable sample designs for school selection, as well as the within school process for selecting students.

83. Standards relating to instrument *design and development* should either be included in the document or made available in a linked document. This is particularly important in relation to the educational assessments themselves, their design and delivery. The purpose and desired outcomes of the field trial should be clarified, and a target number of items specified, in order to standardise the length. Perhaps most importantly to maximise comparability at the data collection and reporting stages, a swift phasing out of paper-based data collection for the student instruments should be considered, with a view to move, as soon as possible, to a single mode of test administration.

84. The standards should be clearer about translation protocols - how equivalence is to be maximised, checked and verified. Procedures for dealing with bi- or multi-lingual instruction should be specified.

85. For *fieldwork and implementation* phase, the timetable should allow countries to conduct the tests at a point in the school year when fewer competing demands might increase response at the school and student level. The issue of incentives for school and students should be clarified and encouraged at a reasonable level in countries where response rates are currently too low. As the PISA assessment and questionnaire are rather demanding for students, efforts should be made to reduce the overall burden on the respondents. Particularly of concern is the fact that variations in length due to the optional modules and national additions could reduce cross-national comparability. Finally, the Technical Standards should refer more specifically to data protection standards.

86. In the *data coding and preparation* area, rules for setting data to missing in cases of low response rates should be further clarified.

87. The Technical Standards document has sections on ‘quality assurance’, but the information therein is often vague. The document could be improved by removing these sections. Instead, each set of standards should have a link to an external inventory of quality assurance mechanisms that explain how the relevant standards can best be achieved.

88. In addition to these core parts of the data lifecycle, the Technical Standards also refer to the *structure of delivery*. Given the complexity of the PISA operation, the document should identify all relevant stakeholders, their roles and responsibilities throughout the entire process. One emerging theme in this review was a sense that, over time, the OECD has assumed a more direct management role rather than a facilitator role. This reflects a similar process in the field of cross-national survey organisation where a strong central management hub is seen as essential. However, this more top-down approach needs to be balanced by bottom-up communication and routes for proper consultation.

89. It is also important that the technical standards are driven by the science and seen to be independent from any political considerations. This review hence recommends that the OECD consider establishing an external group to regularly review the Technical Standards in consultation with contractors and NPMs. This could perhaps become a permanent part of the review cycle of PISA. It should also be noted that the interviews with NPMS conducted in the context of this review generally saw respondents praise the high level of technical support and guidance provided by the contractors.

90. A number of further issues that emerged in the course of this review should be considered.

- Ideally data should be made freely available to the scientific community with appropriate checks in place to ensure confidentiality and with documentation needed to support secondary analysis. PISA findings might receive greater scrutiny if a selection of outputs were subject to independent external replication. Transparency and replicability are now generally expected in the field of cross-national measurement.
- PISA has provision for ‘adjudicated’ areas to provide quality assurance for regional analysis. This should be applied if the data is to be reported at the sub-national level. Other regional use of the data should either be avoided or more clearly flagged as non-adjudicated. If sub-national analysis is to become a more important feature of PISA, then the sampling design should be amended to facilitate that.
- There should be clearer codification of all adjudication processes, the parties involved and their responsibilities.
- The PISA cycle is rather compressed, leaving limited time for round on round improvement and lesson learning. Since change over time in the core measures is likely to be rather gradual, a longer interval between rounds could be considered.
- The process of expanding PISA has been impressive but also difficult and sometimes controversial. The Technical Standards should illuminate the criteria for admitting a country more clearly. Organisationally further expansion could perhaps be slowed to allow a more cautious and considered approach to expansion.
- Whilst overall the quality of PSIA was felt to be high, especially in respect of the core domains, some concerns were expressed by National Project Managers about

whether new domains were being measured to the same level of quality as core areas.

91. The PISA Technical Standards are an ambitious set of technical requirements and targets, which are crucial for ensuring that PISA can meet its overall aims and objectives. At the same time, their scope and purpose should be clarified and a more balanced document produced covering the entire data lifecycle. Future review informed by consultation is recommended alongside benchmarking both within the assessment field and cross-national survey domain. A permanent review group could be one way to achieve this goal in the future.

Appendix A. PISA technical standards recommended for review

Table 1. PISA technical standards recommended for review: Sampling

Standard	Area	Comments
1.4	Sampling	There are trade-offs between sample design and the final response rate achieved. The degree to which a reported response rate includes “substitutes” should be more transparent when reporting. The use of incentives may need greater codifying, as it appears to vary between countries. Care should be taken to limit flexibility in terms of school selection in nested designs. Greater efforts should be made to reduce school substitutions in some countries, possibly by encouraging school and student level incentives.
1.7	Sampling	The 5% exclusion figure (like all such targets) is somewhat arbitrary and it does not take full account of the wide variations in the way education is organised across countries. Sampling from an educational system is quite different to drawing from a register or an address file. Greater thought might be given to how to realise this standard in a functionally equivalent way, with particular reference to the proportions of different excluded categories e. g. those with a language barrier, those with special educational needs, etc.
1.8	Sampling	It is not possible to evaluate these targets in the Technical Standards document per se. We presume they are calculated to produce error margins for particular point estimates. The Technical Standards should elaborate on the choice of this target, through cross-referencing to other documentation.
1.9	Sampling	Consideration should be given to varying the required number of schools based upon the complexity of school system, the type of stratification used, etc. 150 might not always have the same meaning in terms of ‘effective’ impact. In any event the technical guidelines should justify why this figure has been set.
1.10	Sampling	Consideration should be given to varying this number based upon the complexity of the sample design, for instance the nature of the school system, the type of stratification etc. 150 schools might not always have the same meaning in terms of ‘effective’ representation. In any event the technical guidelines should justify (in text or by reference) why this figure has been set.
1.11	Sampling	The technical standards should specify the amount of replacement that is acceptable. Greater efforts should be made to avoid school level refusal, with replacement seen as a last resort and round on round reductions aimed for. The process should be monitored within strata and within strata limits should be set.
1.17	Sampling	In order to ensure a fully input-harmonised and consistent approach, it is really important that both the samples of schools and the samples of students should always be drawn by the international contactors. If this is not possible, very clear guidelines or procedures should be issued to countries, and if countries are to do this themselves, there must be the capacity for central verification and back-checking. Consideration should be given to mandating the centralisation of this process, if only to remove any doubt about the potential for the manipulation of the process, and the eventual validity of the results.

**Table 2. PISA Technical Standards recommended for review:
Instrument design and development**

Standard	Area	Comments
3.1	Instrument design and development	The aim of a field trial is generally to troubleshoot and ensure readiness for the main stage. The technical standards should define the meaning of 'successfully implemented'. Is this simply a feasibility check or a full-blown experiment assessing the quality of items (particularly new modules?) There are trade-offs between limiting field test length (to reduce burden) and undertaking a realistic rehearsal (keeping length closer to the eventual main stage). The overall length of the field tests should be kept under review and limited to what is strictly necessary.
4.2	Instrument design and development	Again, this should be worded as an aspiration. Equivalence cannot be fully evaluated until after the fact and in relation to data. Equivalence is an ambiguous term, as noted in the AAPOR-WAPOR guidelines. Many 3MC surveys use the shorthand "ask the same question" to indicate equivalence in translations. Language differences can result in variability in the time taken to administer instruments. It should be made clearer what is intended. A fuller indication of the level of adaptations permitted could be included as a footnote or an external cross-reference. Care should be taken to ensure adaptations to local circumstances are not disproportionate and are centrally monitored.
4.4/5	Instrument design and development	No amendments are suggested, although TS4.4 seems very similar to TS4.2. In general (as above), these seem to be statements of policy rather than standards for compliance, given that the criteria for their evaluation are not presented here.
10.1/2	Instrument design and development	The "mix and match" approach to the construction of the PISA instrument "menu" may result in variation in the overall duration of the instruments – and this can affect both respondent engagement/fatigue as well as potentially drop-outs. Some concerns are being expressed by countries about the length of the PISA tests. Given that there is already a core and a series of opt-ins, there is a trade-off between increased national "ownership" of the study and the increased burden on respondents.
13.1/4	Instrument design and development	It is not clear that paper-based assessment tests are fully equivalent to the computer tests. Ideally paper-based measurement should be phased out as quickly as possible.
17.2	Instrument design and development	The PISA assessment is lengthy, with accompanying risks of cognitive burden on students and possibly causing lower data quality in later parts of the tests. A review of the costs and benefits of adding additional modules would help inform a debate on this issue. Other models of administration could be considered: greater use of split samples, and the possibility of core and follow-on tests would be two options.

Table 3. PISA Technical Standards recommended for review: Fieldwork implementation

Standard	Area	Comments
1.3	Fieldwork implementation	The timing of the PISA main stage can be problematic in some countries if the testing falls just before exams or other important milestones. Consideration should be given to being more flexible on timing as this could increase response rates at both the school and pupil level.
1.12	Fieldwork implementation	Consideration should be given to incentivising schools and students to maximise participation.
1.14	Fieldwork implementation	A weighted response rate is referred to at 1.11-1.14 but the method of calculation is not stated in the Technical Standards. This should be added as a footnote or the location of the calculation cross-referenced. The Technical Standards should specify the maximum amount of replacement that is allowable. Greater efforts should be made to avoid school level refusal with replacement seen as a last resort and round on round reductions aimed for, rather than this becoming a repeat structural feature. The process should be monitored within strata and within strata limits should be set (e.g. if almost all refusals are within a single strata this could have a very serious impact and even replacements within that strata might then not be equivalent if there are very high refusals for example with private schools).
7.2	Fieldwork implementation	It is not clear if data protection issues are properly communicated and certainly this is missing from the technical standards (see below for more on GDPR).
8.2	Fieldwork implementation	The standard is well specified and required. However, it is not clear how it is monitored, and this might be clarified. Reference should also be made to forbidding coaching of students under any circumstances. Tough sanctions should be specified for infringements, up to and including exclusion from the dataset and from subsequent rounds of PISA.
8.3	Fieldwork implementation	Use of incentives should be encouraged although of course not be related to student <i>achievement</i> , only their participation. Uneven use of incentives across countries should be addressed.
11.1	Fieldwork implementation	Greater details regarding the data protection provisions should be included in the Technical Standards. As OECD is based in Europe it would appear the data collection would be subject to the GDPR, with OECD in Paris designated as the Data Controller (since OECD defines the scope and purpose of the data collection). A DPO should also be named in the Technical Standards.

Table 4. PISA technical standards recommended for review: Data preparation and deposit

Standard	Area	Comments
1.15	Data Preparation and Deposit	The Technical Standard should be clearer on the criteria that would require a “flag” and what the threshold would be for a value being set to missing. This is a major distinction. Users may – indeed will often – choose to include flagged data, with or without footnotes, but missing data is exactly that.
14.1	Data Preparation and Deposit	This is an area of considerable effort and yet has only a single sentence in the Technical Standards, amounting to the requirement to follow instructions. Either it could be better specified or just left in guidance issued to NPMs and international contractors.
15.4/5	Data Preparation and Deposit	The technical standards should provide some parameters on when such recoding might be permitted (e.g. for data protection purposes) - and by whom

Appendix B: PISA reviewing group and reviewing process

92. This report has been drafted by Rory Fitzgerald and Eric Harrison, respectively Director and Deputy Director of the European Social Survey – ERIC, who acted as OECD consultants. The review was carried out under the supervision of Marco Mira d’Ercole and Fabrice Murtin (SDD). Information on PISA has been kindly provided by Tue Halgreen and Francesco Avvisati (EDU). The Reviewing Group included 11 national experts representing 10 countries, who provided written and/or oral comments at various stages of the Review. The list of country experts is described in Table 5.

Table 5. List of Country Representatives in the Reviewing Group

Canada	Tamara Knighton	STATCAN, Canadian Centre for Education Statistics	Expert
Czech Republic	Jakub Lysek	Czech School Inspectorate	Expert
France	Fabrice Murat	Education Ministry	Expert
Japan	Satoshi Usami	University of Tokyo	Associate Professor
Mexico	José Paul Carrasco Escoba	INEGI (Statistical Innovation)	Expert
	Armando Ibarra Medina Noriki	INEGI (Samples Selection and Control)	Expert
Korea	Namwook Koo	Korea Institute for Curriculum and Evaluation (http://www.kice.re.kr/main.d)	Associate Research Fellow
Romania	Sultana Elena Stan	National Institute of Statistics	Expert
Turkey	Derya TUNCER	Turkish Statistical Institute	Expert
UK	Barbara Donahue	Department for Education (Standards and Testing Agency)	Head of Assessment Research and Psychometrics
US	Enis Dogan	National Center for Education Statistics	Senior Psychometrician

93. In terms of process and timeline, three video-conferences with the Reviewing Group were organised; the first one kicked off the discussion, the second one focused on the outline of the report, the last one discussed the first draft (Table). Written comments were received at various stages of the review. In addition, ten interviews with PISA contractors or National Project Managers were conducted in April.

Table 6. Timeline of the review

01 January 2020	Selection of external consultant and CSSP experts
03 February 2020	Video-conference with consultants and EDU to exchange key information
28 February 2010	Video-conference with the Reviewing Group to discuss key aspects to be covered by the report
12 March 2020	Outline of the report circulated by the consultant to the Secretariat.
20 March 2020	Second video-conference with the Reviewing Group to discuss outline
03 April 2020	Deadline for written comments on outline and PISA Technical Standards 2021 by Reviewing Group
01 April 2020	Conduct of interviews and drafting of the report
27 April 2020	First draft circulated internally to the Secretariat
07 May 2020	Draft report sent out to the Reviewing Group
15 May 2020	Deadline for reception of comments by Reviewing Group
19 May 2020	Final video-conference with Reviewing Group
02 June 2020	Final draft sent to CSSP
24-25 June 2020	Presentation at CSSP
13 July 2020	Revision of the report to incorporate CSSP comments
End-July	Submission to PISA Governing Board

Appendix C: PISA management structure and timeline

94. The PISA management structure utilises the strengths of countries with full membership status in PISA and existing infrastructures at the national and international levels and encourages co-operation and development with national educational and statistical agencies. The structure combines access to political structures at the level of the OECD, ownership by fully participating countries during project design, implementation and reporting and the necessary power of decision making on a day to day basis.

95. Countries with full membership status in PISA (currently the 35 OECD countries plus Brazil) are responsible for PISA at the policy level. Through the PISA Governing Board, they determine PISA's policy priorities and oversee adherence to these priorities during its implementation. This includes the setting of priorities for indicator development and reporting, the development of data collection instruments, and the determination of the scope of work that shall afterwards be translated into a tender for the international contractors.

96. The OECD Secretariat is responsible for PISA's overall management. It serves as the Secretariat of the PISA Governing Board and as the interface between the PISA Governing Board and the contractors.

97. The design and implementation of the surveys, within the framework established by the PISA Governing Board, is the responsibility of contractors selected through an approved and transparent tendering process, working closely with the OECD Secretariat.

98. Experts from participating countries serve on Subject Matter Expert Groups that provide input to the development of the assessment instruments to ensure that the participants' diverse cultural and curricular contexts are reflected in the assessments.

Participating countries

99. Participating countries shape the project in three principal ways:
- As members of the PISA Governing Board, they determine the policy objectives and design parameters for PISA. OECD countries and Partner countries that are full participants in the program (Associate Partners) are represented in the Governing Board. Brazil is currently an Associate Partner with full membership of the PISA Governing Board.
 - Through National Project Managers (NPMs), they implement the project at the national level subject to agreed-upon administrative procedures.
 - Through experts represented in the Subject Matter Expert Groups (SMEG), they contribute to the development of the assessment frameworks and assessment instruments under the governance and auspices of the contractors.

The PISA Governing Board (PGB)

100. The PISA Governing Board is composed of representatives of OECD countries and partner countries with full membership in PISA (Associate partners).

101. In consultation with the OECD Secretariat, the PISA Governing Board: (i) determines PISA's policy objectives and the content domains to be tested; (ii) establishes priorities for indicators, analysis and data collection instrument development; (iii) specifies the scope of work that shall be translated into a tender; and (iv) guides the preparation, review, and completion of all programme-related reports.

102. The PISA Governing Board also works with the OECD Secretariat to ensure compliance with the policy objectives and design parameters at milestones during PISA's implementation. In addition to enabling participating countries to share substantively with one another the programme's decision making and policy direction,

103. The PISA Governing Board currently considers two forms of cooperation with OECD partner countries as part of its global relations strategy: (i) Participant status, which gives access to participation in the PISA survey and observer status at the PISA Governing Board, and (ii) Associate status, which covers full participation in the PISA survey and in the PISA Governing Board. Associates in PISA take on the same obligations and rights as OECD countries, including the right to vote in the Governing Board. Brazil is currently an Associate.

National Project Managers

104. Participating countries and economies nominate National Project Managers to implement the surveys in the national context. National Project Managers are the primary means of day to day contact between participating countries and the contractors for the implementation of PISA and shall communicate with the contractors on all issues related to the implementation of the assessments in their country. National Project Managers play a vital role in ensuring that PISA is a high-quality project with results that can be verified and evaluated. They can also play an important role in the development and review of PISA's reports and publications, in consultation with their respective PISA Governing Board member.

105. The National Project Manager decides how best to facilitate the communication and coordination needed at the national level for implementing data collection responsibilities as well as for interacting with international contractors.

106. A major risk in the implementation of PISA is the possibility of deviations introduced at the national level during the course of the implementation. Seemingly unimportant decisions, taken alone or in combination, can undermine the integrity of the entire survey in a particular country. The international contractors seek an open and collegial process with National Project Managers that stresses the role that key design parameters shall play in assuring final quality. The strict adherence to PISA standards ensures that the assessment is carried out with a high degree of uniformity in all participating countries and economies.

Subject Matter Expert Groups and Questionnaire Expert Groups

107. A strong cognitive core and a coherent theoretical underpinning are of key importance in the validity of PISA and other international comparative assessments. Furthermore, such assessments require a wide range of internationally available technical expertise in the different assessment domains. Substantive input from countries is also imperative in ensuring that the assessments will be internationally valid and reflect the cultural and curricular context of participating countries. Subject matter expert groups in PISA are usually established for each content domain to be tested as well as for context questionnaires. The expert groups are key players in the development of the assessment frameworks and instruments. In addition, these groups assist the international contractors in linking PISA's policy objectives with substantive and technical expertise. The contractors are required to establish processes by which the expert groups, the contractors and the Secretariat can best communicate with each other and with the participating countries.

Technical Advisory Group

108. The Technical Advisory Group constitutes a forum through which the main actors implementing the project interact both among themselves and with those whose additional technical expertise is sought. Unlike the Subject Matter Expert Groups, the Technical Advisory Group has a permanent role across survey cycles and include, among other experts, those individuals who have a leading operational role in PISA 2021. The members of the group were, at the outset of the first PISA cycle, appointed by the OECD Secretariat in consultation with the PISA Governing Board, based on nominations by the contractors for the first cycle. Some of the membership of the group has remained constant since the first cycle, and new members have been appointed as required. The Technical Advisory Group is managed by the OECD Secretariat.

109. The Technical Advisory Group explores technical issues that have policy implications. During programme implementation, the OECD Secretariat brings proposals for modifications of PISA's programme of work or its implementation procedures that have cost implications at either international or national levels to the PISA Governing Board for review and decision (e.g. modifications to data collection instrument development procedures, the sample and test design, the translation procedures, or marking and verification procedures). When appropriate, the PISA Governing Board seeks the advice of the Technical Advisory Group on these or other matters, either directly or through the OECD Secretariat. The detail of the agenda for TAG meetings is drawn up by the OECD Secretariat, working closely with the International Survey Director (Core A).

Sampling referee

110. To ensure the integrity of national samples, the PISA Governing Board appoints a sampling referee for each cycle of PISA. Based on evidence about the quality of the samples for PISA 2021 and in consultation with the Technical Advisory Group or other experts and with the contractor for Core C, the sampling referee will assess the implications for the use of country results in the international and thematic reports and will make recommendations to the PISA Governing Board regarding the use of individual countries' data in the reporting process. In addition, the sampling referee will inform participating

countries and the PISA Governing Board as early as possible of problems with sampling or response rates that may or will jeopardise countries' compliance with sampling guidelines for PISA 2021, providing an explanation for the problems or concerns and, when possible, suggesting remedies for them.

The OECD Secretariat

111. The OECD Secretariat is responsible for PISA's overall management. This entails preparing the terms of reference for each survey cycle under the guidance of the PISA Governing Board, engaging contractors to implement specified activities, and monitoring the project on a day to day basis. The OECD Secretariat is also responsible for building consensus at the policy level among countries with full membership in the PISA Governing Board, both during the preparation of the terms of references and at milestone points of the surveys.

112. The OECD Secretariat serves as the Secretariat of the PISA Governing Board and as the interface between the PISA Governing Board and the contractors during all stages of PISA 2021. It is a further responsibility of the OECD Secretariat to provide the PISA Governing Board with a progress report on no less than a biannual basis as well as with a report on financial and contractual management on an annual basis.

113. The OECD Secretariat produces indicators and analyses based on statistical components provided by the contractors, and is fully responsible for preparing the international PISA reports in collaboration with the countries with full membership status in PISA, through the PISA Governing Board. It is also responsible for overseeing thematic reports or other reports or research papers related to the PISA cycle, which are published under the responsibility of the OECD.

114. Countries with full membership status in PISA shall agree on a set of general rules for the inclusion/exclusion of country results in international reports. The OECD Secretariat shall arbitrate disagreements between participating countries and the sampling referee under guidelines established by the PISA Governing Board.

115. Additionally, it is the OECD Secretariat's role to participate actively during the development of all documents and reports and to oversee the documentation process of the project including approval of all documents before they are provided to participating countries. This applies, in particular, to meeting documents, manuals and survey instruments.

Timeline

116. PISA surveys are run every 3 years and each of them is implemented over a 5-year cycle. Figure 1 illustrates the data cycle for PISA 2024, as based on PISA 2021 cycle. Key activities are planned as follow:

- 2021: Development and review of all frameworks as well as development and review of cognitive items and questionnaires.
- 2022: Preparation for Field Trial data collection including sampling, school materials, translation, adaptation, verification, testing of Student Delivery System, etc.

- 2023: Implementation of the Field Trial, coding of open constructed response questions, Field Trial data preparation and submission, and Field Trial data analysis
- 2023-2024: Preparation for the Main Survey data collection including sampling, school materials, corrections, translation, adaptation, verification, testing of Student Delivery System, etc.
- 2024: Main Survey data collection, coding of open-constructed response questions, Main Survey data preparation and submission, and Main Survey data analysis.
- 2025: Analysis of Main Survey results, preparation and review of national and international databases, and preparation of reports and dissemination products.

Figure 1. PISA 2024 data cycle based on PISA 2021



Appendix D: PISA list of technical standards

117. Table describes the full list of Technical Standards for PISA 2021. There are 77 standards (not disclosed here for the sake of space) organised in 22 entities. Those that are subject to a specific recommendation made in this report are highlighted in bold.

Table 7. Overview of technical standards

Data standards	Management standards
1. Target population and sampling	16. Communication with the international contractors
2. Language of testing	17. Notification of international and national options
3. Field Trial participation	18. Schedule for submission of materials
4. Adaptation of tests, questionnaires and school-level materials	19. Management of data
5. Translation of assessment instruments, questionnaires and school-level materials	20. Archiving of materials
6. Testing of national software versions	
7. Technical support	National involvement standards
8. Test administration	21. National feedback
9. Training support	22. Meeting attendance
10. Implementation of national options	
11. Security of the material	
12. Quality monitoring	
13. Assembling and printing paper-based materials	
14. Response coding	
15. Data submission	

Annex B. Comments from the US Member of the CSSP Group Reviewing the PISA Technical Standards

Comments provided by Enis Dogan, Senior Psychometrician, National Center for Education Statistics, United States

1. The PISA program can maintain its successful implementation only if its assessments are guided by precise and comprehensive technical standards. The current version of the *2021 Standards* needs to be revised in order to provide such guidance. We offer a number of recommendations below to bring the existing document to an acceptable level of depth in terms of its current scope. However, the standards need to be either greatly expanded in scope or a separate set of technical standards should be developed to provide comprehensive guidance in all aspects of assessment design, development, and implementation. This need is discussed further following the more immediate recommendations we offer:

- Intended uses of the ‘standards’ should be clarified and document should be retitled accordingly (e.g. as “*Data Collection Standards*”).
- Standards should require development of various documents regarding design, development, and implementation of assessments, including assessment frameworks, assessment plan, data analysis plan, and technical report and specify the scope, and development and approval process for each.
- Standards should specify, in greater detail, the adjudication process to be followed in case any one of the standards are not met by any of the participating countries or economies.
- Parameters around negotiations between the NPMs and contractors should be better defined. Not all aspects of assessment development and implementation can be left open to negotiation.
- All vague language (e.g. “The size of this representative sample should not be too small”, p. 4) should be eliminated from the standards.
- Quality assurance (QA) sections should not read as merely a list of bullet points, and standards should specify how is each QA step is accomplished, by whom and when.
- Standards should describe the Technical Advisory Group in term of its role and responsibilities, qualifications of its members, and whether and how often its members should be refreshed.

Need for comprehensive technical standards

2. The purpose of the *2021 Standards* is stated as “to list the set of standards upon which the PISA 2021 data collection activities will be based” (p. 2). Technical standards should be comprehensive enough to serve as a yardstick against which to appraise the current PISA assessments and a guide to good practice for future ones. Data collection can only be one aspect of such appraisal. Therefore, there is a great need for a comprehensive set of technical standards. The Standards for Educational and Psychological Testing (APA, AERA, NCME, 2014) can provide guidance in shaping this activity. These standards are organized in three foundational aspects: Validity, Reliability/Precision, and Fairness. Unfortunately, these key concepts are not even mentioned in the *2021 Standards*. Other aspects in the APA, AERA, NCME standards include test design and development, scoring, scaling and score interpretation, which are also, largely, neglected in the *2021 Standards*. The guidelines established by the International Test Commission (ITC) (n.d.) would also be a valuable resource in revision PISA standards. The ITC guidelines cover aspects such as translating and adapting tests, computer-based testing, test use, and assessment of diverse populations etc. The technical Standards for IEA Studies (Martin, Rust & Adams, 1999) is another example that covers several aspects such as design, development of instruments, data collection and processing and data analysis. These documents should be used as resources in building a comprehensive set of technical standards for future PISA assessments.

References

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (U.S.) (2014), *Standards for educational and psychological testing*. Washington, DC: AERA, American Educational Research.
- International Test Commission (n.d.). *International Guidelines for Test Use*, retrieved from <https://www.intestcom.org/page/17>.
- Martin, M.O., K. Rust and R.J. Adams (1999), *Technical standards for IEA Studies*. Delft, The Netherlands: International Association for the Evaluation of Educational Achievement.