



---

14

# Data Adjudication

<b>Introduction</b> .....	278
<b>General outcomes</b> .....	281



## INTRODUCTION

The PISA Technical Standards (see Annex F) specify the way in which PISA must be implemented in each country. The international contractor has to monitor the implementation in each country and adjudicate on the countries' adherence to the standards. This chapter describes the process used to adjudicate the implementation of PISA 2012 in each of the entities (i.e. the participating countries, economies and adjudicated regions<sup>1</sup>) and it gives the outcomes of data adjudication that are mainly based on the following aspects:

- the extent to which each adjudicated entity met PISA sampling standards;
- the outcomes of the adaptation, translation and verification process;
- the outcomes of the National Centre and PISA Quality Monitoring visits and interviews;
- the quality and completeness of the submitted data; and
- the outcomes of the international coding review.

The areas covered in the PISA 2012 Technical Standards include the following:

### **Data Standards**

- Target population and sampling
- Language of testing
- Field Trial participation
- Adaptation of tests, questionnaires and manuals
- Translation of tests, questionnaires and manuals
- Test administration
- Implementation of national options
- Security of the material
- Quality monitoring
- Printing of material
- Response coding
- Data submission

### **Management standards**

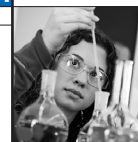
- Communication with the international contractors
- Notification of international and national options
- Schedule for submission of materials
- Drawing samples
- Management of data
- Archiving of materials

### **National involvement standards**

- National feedback

### **Implementing the standards – quality assurance**

National Project Managers of participating countries, economies and adjudicated regions are responsible for implementing the standards based on the international contractor's advice as contained in the various operational manuals and guidelines. Throughout the cycle of activities for each PISA survey the international contractor carried out quality assurance activities in two steps. The first step was to set up quality control using the operational manuals, as well as the agreement processes for national submissions on various aspects of the project. These processes give the international contractor staff the opportunity to ensure that PISA implementation was planned in accordance with the PISA 2012 Technical Standards, and to provide advice on taking rectifying action when required and before critical errors occurred. The second step was quality monitoring, which involved the systematic collection of data that monitored the implementation of the assessment in relation to the standards. For data adjudication it was the information collected



during both the quality control and quality monitoring activities that was used to determine the level of compliance with the standards.

### Information available for adjudication

The international contractor monitors a country's implementation of the data collection procedures from a range of perspectives and from processes occurring during many stages of the PISA cycle. These perspectives include monitoring a country's adherence to the deadlines, communication from the sampling contractor about each country's sampling plan, information from the language verification team, data from the PISA Quality Monitors and National Centre Quality Monitors, and information gleaned from direct interviews at National Project Manager and Marker Training meetings. The information is combined together in the database so that:

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

Many of these data collection procedures refer to specific milestone documents, specified in the National Project Manager's Manual and the Sampling Manual in particular. These are procedures that the international contractor requires for Field Trial and Main Survey preparation from each National Centre. The data adjudication process provides a motivation for collating and summarising the specific information relating to PISA standards collected in these documents, combined with information collected from specific quality monitoring procedures such as the national centre quality monitoring interview, PISA quality monitor visits and from information in the submitted data.

The quality monitoring information was collected from the following main administrative areas covering various quality monitoring instruments:

- International contractor administration and management: information relating to administration processes, agreement of adaptation spreadsheets, submission of information.
- Data analysis: information from the dodgy item reports, Field Trial sample, item information for cleaning.
- Field operations – manuals: information from the agreement of adaptations to test administration procedures and field operations.
- Final Optical Check team: information from the pre- and post-Main Survey Final Optical Checks of Main Survey booklets.
- Main Survey review: information provided by the National Project Managers in the Main Survey review process.
- National Centre quality monitoring: information gathered during the pre-Main Survey National Centre Quality Monitoring visits, and through interviews conducted during meetings of National Project Managers or at other times.
- PISA Quality Monitor country reports: information gathered via the test session reports from PISA Quality Monitors and through their interviews with School Co-ordinators.
- Sampling: information from the submitted data such as school and student response rates, exclusion rates and eligibility problems.
- Translation: information relating to the verification and translation process.
- PISA Quality Monitor co-ordination: information relating to the recruitment and selection of PISA Quality Monitors and National Quality Monitoring issues.
- Data cleaners: issues identified during the data cleaning checks and from data cleaners' reports.
- Item developers: issues identified in the coder query service and training of coders.
- Data processing: issues relating to the eligibility of students tested.
- Questionnaire data: issues relating to the questionnaire data in the national questionnaire reports provided by the international contractor.
- Questionnaire Final Optical Check: issues arising from the Final Optical Check of the questionnaires.

There were two types of PISA quality monitoring reports, one containing data for each observed session in each school and another detailing the general observations across all schools visited by each Quality Monitor. The PISA Quality Monitoring reports contain data related to Test Administration as well as a record of interview with School Co-ordinators. The Test Administrator session report was completed by the test administrator after each test session and also contained



data related to Test Administration. The data from this report were recorded by the National Centre and submitted as part of the national dataset to the international contractor. The National Centre quality monitor interview schedule contained information on all the standards, as did the Main Survey review.

The *National Centre Quality Monitor Interview Schedule* and the *Main Survey Review* were self-declared by the National Project Manager. The PISA Quality Monitoring data are collected independently of the National Project Manager.

### Data adjudication process

The main aim of the adjudication process is to make a judgement on each national dataset in a manner that is transparent, based on evidence and defensible. The data adjudication process achieved this through the following steps:

**Step 1:** Quality control and quality monitoring data were collected throughout the survey administration period.

**Step 2:** Data collected from both quality control and quality monitoring activities were entered into a single quality assurance database.

**Step 3:** Experts compiled country-by-country reports that contained quality assurance data for key areas of project implementation.

**Step 4:** Experts considered the quality assurance data that were collected from both the quality control and Quality Monitoring activities, to make a judgement. In this phase the experts collaborated with the project director and other international contractor staff to address any identified areas of concern. Where necessary, the relevant National Project Manager was contacted through the international project director. At the end of this phase experts constructed, for each adjudicated dataset, a summary detailing how the PISA Technical Standards had been met.

**Step 5:** The international contractor and the Technical Advisory Group reviewed the reports and made a determination with regard to the quality of the data from each adjudicated entity.

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

Information was collected in relation to these standards through a variety of mechanisms: through PISA Quality Monitor reports; through the Field Trial and Main Survey reviews; through information negotiated and stored on the MyPISA website (the portal which was used in PISA 2012) in relation to specific PISA implementation tasks; through communications and visits of international contractor staff to National Centres; through the formal and informal exchanges between the international contractor and National Centres over matters such as sampling, translation and verification, specially requested analyses (such as non-response bias analysis); through a detailed post-hoc inspection of all Main Survey assessment materials (test booklets); and through the data cleaning and data submission process.

For PISA 2012, an adjudication database was developed to capture, summarise and store the most important information derived from these various information sources. The staff members of the international contractor who lead each area of work were responsible for identifying relevant information, and entering it into the database. This means that at the time of data adjudication, relevant information is easily accessible for making recommendations about the fitness of use of data from each PISA adjudicated entity.

The adjudication database captures information related to the major phases of the data operation: field operations, sampling, computer-based problem solving and computer-based assessment of literacy (where applicable), questionnaires, cognitive tests. Within each of these phases, the specific activities are identified, and linked directly to the corresponding standards.

Within each section of the database, specific comments are entered that describe the situation of concern, the source of the evidence about that situation, and the recommended action. Each entry is classified as serious, minor or is rated as being of no importance for adjudication. Typically, events classified as serious would warrant very close expert scrutiny, and possibly action affecting adjudication outcomes. For example, the reliability of parental occupation data from Albania was subject to scrutiny, resulting in a recommendation that all data dependant on Albania's parental occupation data (in particular, all data that use the *PISA index of economic, social and cultural status* [ESCS]) should be deleted from the database and relevant tables. Events classified as minor would typically not directly affect adjudication outcomes, but will be reported back to National Centres to assist them in reviewing their national procedures. It was expected that



the data adjudication would result in a range of possible recommendations. Some possible, foreseen recommendations included:

- that the data be declared fit for use;
- that some data be removed for a particular country, for example the removal of data for some items such as open-ended items, or the removal of data for some schools;
- that rectifying action be performed by the National Project Manager, for example; providing additional evidence to demonstrate that there is no non-response bias, or rescore open-ended items;
- that the data not be endorsed for use in certain types of analyses; and
- that the data not be endorsed for inclusion in the PISA 2012 Database.

Throughout PISA 2012, the international contractor concentrated its quality control activities to ensure that the highest scientific standards were met. However, during data adjudication a wider definition of quality was used especially when considering data that were at risk. In particular the underlying criterion used in adjudication was fitness for use. That is, data were endorsed for use if they were deemed to be fit for meeting the major intended purposes of PISA.

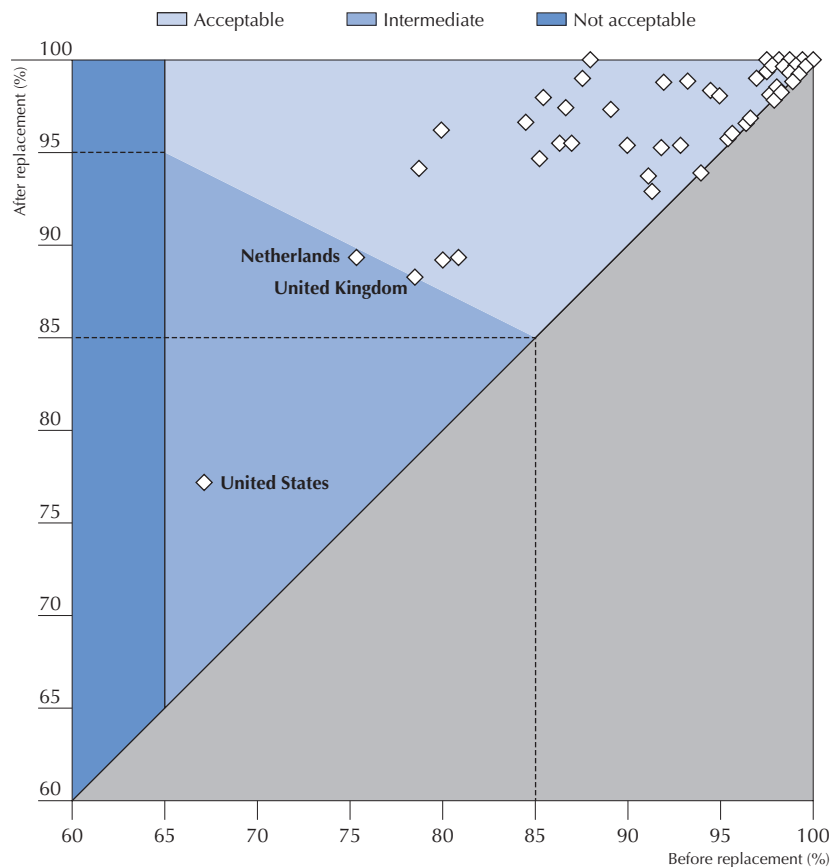
## GENERAL OUTCOMES

### Overview of response rate issues

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

■ Figure 14.1 ■

**Attained school response rates**





One country – the United States – failed to meet the school response rate requirements. One other country, the Netherlands, had a response rate below the 85% level before the replacement but cleared the acceptable level after the replacement schools were included.

After reviewing the sampling outcomes, the international contractor asked the United States to provide additional data that would assist it in making a balanced judgement about the threat of non-response to the accuracy of inferences that could be made from their PISA data.

The international contractor determined that the data were acceptable.

### **Computer-Based Assessment**

The PISA 2012 Technical Standards did not include any specification of the response rate needed for students undertaking the computer-based assessment. In the absence of agreed technical standards for the response rate of students undertaking the computer-based assessment, the Technical Advisory Group advised for the PISA 2009 Digital Reading Assessment that the desired response rate was 0.8 of the response rate of students undertaking the paper-based assessment. This response rate requirement has been carried into PISA 2012. Comments for the response rates of the countries that implemented the computer-based assessment are discussed based on those criteria.

### **Detailed country comments**

It is important to recognise that PISA data adjudication is a late but not necessarily final step in the quality assurance process. By the time each country was adjudicated at the Technical Advisory Group meeting that took place in Melbourne in April 2013, the quality assurance and monitoring processes outlined earlier in this chapter and in Chapter 7 had been implemented. Data adjudication focused on residual issues that remained after these quality assurance processes had been carried out.

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

#### ***Departures from standard procedures in the national data collection plan***

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

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

#### ***Departures from standards arising from implementation***

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

A point in the preparation stage that led to significant errors in several countries was in the final stages of the preparation of the test booklets and questionnaire instruments at the national centre, following the Final Optical Check of these materials by the international verification team (see Chapter 5). These errors included a failure to correct errors that had been identified by the international verifiers as part of the final optical check, or the introduction of completely new errors



to the booklets and/or questionnaires following the Final Optical Check. An obvious example of such an error (which was emphatically warned against, but nevertheless unfortunately occurred in a number of countries) is in the repagination of the booklets, so that the location of the item components (e.g. stimulus material and multiple-choice responses) would differ from the materials approved internationally. The nature and extent of such errors, the estimated impact on data quality, and actions taken with regard to the international database, are reported in the country summaries below.

A small number of countries failed to reach the required minimum sample sizes of 4 500 students and 150 schools. Such cases were considered as part of the data adjudication process. Even a minor deviation in sample size might be considered a substantive enough issue to report, for example in countries where standard errors tend to be higher for a given sample size. On the other hand, minor deviations from these minimal sample sizes (i.e. shortfalls of fewer than 50 students or 5 schools, and in countries that nevertheless achieved comparable standard errors on the major survey estimates) are not reported below.

A component of the data adjudication process was to consider the cases of multiple, or more complex departures from the PISA standard procedures, as well as to consider the impact of errors or shortfalls across all aspects of each country's data collection plan and implementation, and make an evaluation with respect to the quality and international comparability of the PISA results. Notable departures from the standards are reported in the country summaries below. If a country is not listed below then it fully met the PISA standards. Further, in the case of minor deviations from the standards, unless otherwise noted, additional data were usually available to suggest the data were suitable for use.

Particular attention has been paid to the achievement of the specified response rates of 85% for schools, 80% for students within schools and no more than 5% of students excluded from the assessment.

## Albania

Analysis of the data for Albania suggest that the PISA Technical Standards may not have been fully met for the following four main reasons: (i) the rate of missing data for parental occupation was systematically related to (sampling) stratum membership of each student (students from higher strata had more missing data); (ii) 80.3% of the School Questionnaires returned contained impossible total school enrolment data, as there were more students in the Student Questionnaire data file than the number of students listed as enrolled at the school; (iii) the coding of items in Albania is at an extremely high level of agreement between independent coders; and (iv) the expected relationships between student achievement and student background characteristics were inconsistent with previous experience. Further investigation of the survey instruments, the procedures for test implementation and coding of student responses at the national level provided sufficient evidence of systematic errors and violations of the PISA Technical Standards for the parental occupation data and school enrolment data. Albania's data for parental occupation (*ESCS*, *HISEI*, *BMMJ* and *BMFJ*) and school enrolment (*SC07Q01*, *SC07Q02*, *SCHSIZE*, *PCGIRLS*, *STRATIO*, *SMRATIO* and *IRATCOMP*) are, therefore, deleted from the PISA 2012 international dataset.

## Argentina

### **CABA (Ciudad Autónoma de Buenos Aires)**

There were fewer than the 1500 students specified in the standards for an adjudicated region assessed (1316) and fewer than 50 schools participating (48). Data were included in the final database.

## Belgium

Belgium had a school response rate before replacements of 83.67%. After replacement the response rate was 95.92% which was above the PISA standard.

### **Flemish Community**

The Flemish Community had a school response rate before replacements of 79.56%. After replacement the response rate was 95.58% which was above the PISA standard.

## Canada

There was a total of 6.38% exclusions in Canada. A bias analysis showed that the non-response bias would be negligible. It was thought that the extra students excluded were special needs students.

Canada's data were, therefore, included in the final database.



## Denmark

Overall exclusions were greater than 5% (6.18%). Data were fully explained – there was a difficulty in defining the school population – some international schools were not included when they should have been. Denmark also had a school response rate before replacements of 84.97%. After replacement the response rate was 92.62% which was above the PISA standard. Denmark's data were included in the international database.

## Estonia

There was a total of 5.80% exclusions in Estonia. A bias analysis showed that the non-response bias would be negligible. It was thought that the extra students excluded were special needs students.

Estonia's data were, therefore, included in the final database.

## Hong Kong-China

Hong Kong had a school response rate before replacements of 78.85%. After replacement the response rate was 94.23% which was above the PISA standard. Fewer than 150 schools participated (147), but this was deemed to be acceptable and Hong Kong's data were included in the international database.

## Iceland

Fewer than 150 schools (133) participated for Iceland. This was deemed to be acceptable, as Iceland was a full census country for the paper-based assessment, and Iceland's data were included in the international database.

## Luxembourg

There was a total of 8.40% exclusions in Luxembourg. Further analysis indicated that the non-response bias would be negligible. The data from Luxembourg, therefore, were included in the international database.

## Netherlands

The Netherlands had a school response rate before replacements of 74.37%. After replacement the response rate was 88.94% which was above the PISA standard. There were also fewer than 4500 students assessed (4434). Data were included in the international database.

## New Zealand

New Zealand had a school response rate before replacements of 79.19%. After replacement the response rate was 89.85% which was above the PISA standard. Data were included in the international database.

## Norway

There was a total of 6.11% exclusions in Norway. Data were included in the final database.

## Poland

Poland had a school response rate before replacements of 84.57%. After replacement the response rate was 96.81% which was above the PISA standard. Data were included in the international database.

## Qatar

The tests were incorrectly printed. After completing the Final Optical Check the printer made changes which resulted in questions being presented to the students in non-standard ways. Item difficulty was calculated and no systematic influence was observed in these cases. The data from Qatar, therefore, were included in the international database.

## Spain

### *Balearic Islands*

There was a total of 7.41% exclusions, and fewer than 1500 students assessed (1455) in the Balearic Islands. Data were included in the final database.

### *Catalonia*

There was a total of 6.42% exclusions in Catalonia. Data were included in the final database.



**Aragon**

There were fewer than 1500 students assessed (1427). Data were included in the final database.

**Murcia**

Murcia had fewer than 1500 students assessed (1411). Data were included in the final database.

**Sweden**

There was a total of 5.44% exclusions in Sweden. Data were included in the final database.

**Tunisia**

Tunisia had fewer than 4500 students assessed (4391). Data were included in the final database.

**United Kingdom**

There was a total of 5.49% exclusions in the United Kingdom. Data were included in the final database.

The United Kingdom had a school response rate before replacements of 84.44%. After replacement the response rate was 90.16% which was above the PISA standard.

**United States**

There was a total of 5.35% exclusions in the United States. Additional analysis supported the case that no notable bias would result from non-response. It was thought that the extra students excluded were special needs students.

The United States had a school response rate of 77.78%. Additional analysis supported the case that no notable bias would result from non-response. The data from the United States, therefore, were included in the international database.

**Florida**

There was a total of 8.29% exclusions in Florida. Data were included in the final database.

**Massachusetts**

There were fewer than 50 schools assessed (49). Data were included in the final database.

**Note**

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