Chapter 10: Field Operations

Pat Montalvan and Michael Lemay, Westat

10.1 Overview

As with all aspects of PIAAC, countries were asked to comply with a set of Technical Standards and Guidelines (TSG) for survey operations/data collection. These standards can be found in Chapters 2, 8, 9 and 10 of the TSG. Part of the TSG included a quality assurance (QA) and quality control (QC) program for survey operations covering the collection of a range of information about the design and implementation of PIAAC data collection in each country via written reports, phone conferences and some in-person meetings. (Chapter 11 provides a detailed description of the QA and QC program which facilitated the collection of this information.)

This chapter presents information about the 25 countries/territories that completed the PIAAC Main Study data collection: Australia, Austria, Canada, Cyprus,[1] the Czech Republic, Denmark, England (UK) and Northern Ireland (UK),[2] Estonia, Finland, Flanders (Belgium), France, Germany, Ireland, Italy, Japan, Korea, the Netherlands, Norway, Poland, the Russian Federation,[3] Slovakia, Spain, Sweden, and the United States.[4] All the information presented in this chapter is based on data self-reported by countries as of 31 July 2013.

Sections 10.2 through 10.7 of this chapter provide highlights of findings with respect to data collection timeline, production and response rates, interviewer training, field management practices, staffing and supervision, nonresponse reduction, and fieldwork quality control. Furthermore, at the end of data collection, interviewers were debriefed on their PIAAC experience. This feedback is summarized in section 10.8. Finally, section 10.9 concludes the chapter with recommendations for future PIAAC cycles.

Furthermore, it is also important to note that there were deviations from the TSG with regards to data collection in most countries. Whenever deviations were identified by the Consortium, be it during the planning, training or implementation stages, countries were notified quickly via email or telephone conference or both. If possible, acceptable alternatives were identified; otherwise both the country and the OECD were notified of the potential problem. However, for the most part, key TSG guidelines or acceptable alternatives were followed by most countries.

[1] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[2] England and Northern Ireland are reported on separately at the request of the United Kingdom.

[3] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

[4] Portugal and Chile, two countries which participated in the Round 1 Field Test, officially notified the OECD that they would not be conducting the Round 1 Main Study. They later rejoined PIAAC as part of Round 2.

10.2 Data collection timeline

Countries were expected to begin data collection on 1 August 2011 and complete fieldwork by 31 March 2012 (8 months or 243 days). Table 10.1 presents detailed information about each country's adherence to the data collection timeline.

Almost 60% of the countries completed the fieldwork by mid-April and the remainder by 24 November 2012. The actual length of the field period ranged from 79 days in France to 284 days in Sweden (average: 224 days).

The majority of countries did not start data collection on 1 August 2011 primarily because they believed that the vacation plans of many field staff and respondents would negatively impact production in this last month of summer. Seven countries (Austria, Estonia, Germany, Ireland, Poland, England (UK) and Northern Ireland (UK) began exactly on 1 August 2011. Four countries (Australia, Canada, Russian Federation[5] and Slovakia) began data collection in late fall for various reasons. Canada and Australia started in November and October, respectively, due to ongoing competing projects. Slovakia and the Russian Federation[6] began data collection in late October and late November, respectively, due to contractual and budgetary issues. France made the decision to begin data collection in September 2012.

Most countries concluded data collection by mid-April 2012. Nine countries ended data collection on or before 31 March 2012. Thirteen additional countries ended by 31 May, Sweden and Canada ended in June, and France ended in November 2012.

[5] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

[6] See above footnote.

Table 10.1: Data Collection Timeline

	Fieldwork Dates		Duration (Days)
	Start	End	
Australia	1 Oct 2011	31 Mar 2012	182
Austria	1 Aug 2011	31 Mar 2012	243
Canada**	1 Nov 2011	30 June 2012	242
Cyprus[7]	1 Sept 2011	31 Mar 2012	212
Czech Republic	15 Aug 2011	15 Apr 2012	244
Denmark	28 Aug 2011	17 Apr 2012	233
England (UK)	1 Aug 2011	31 Mar 2012	243
Estonia	1 Aug 2011	30 Apr 2012	273
Finland	30 Aug 2011	5 Apr 2012	219
Flanders (Belgium)	19 Aug 2011	31 Mar 2012	225
France	7 Sep 2012	24 Nov 2012	79
Germany	1 Aug 2011	31 Mar 2012	243
Ireland	1 Aug 2011	31 Mar 2012	243
Italy	1 Sept 2011	15 Apr 2012	227
Japan	30 July 2011	29 Feb 2012	214
Korea***	26 Sept 2011	24 Apr 2012	132
Netherlands	22 Aug 2011	11 May 2012	263
Northern Ireland (UK)	1 Aug 2011	13 Apr 2012	256
Norway	17 Aug 2011	30 Apr 2012	257
Poland*	1 Aug 2011	31 Mar 2012	243
Russian Federation* [8]	21 Nov 2011	29 May 2012	190
Slovakia	27 Oct 2011	24 Apr 2012	180
Spain	2 Sept 2011	30 Apr 2012	241
Sweden	22 Aug 2011	1 June 2012	284
United States	25 Aug 2011	3 Apr 2012	222

Source: Data Collection Form submitted after data cleaning unless otherwise noted.

* Based on Data Collection Form submitted after conclusion of data collection.

** Canadian PIAAC data collection was scheduled so as to not conflict with Census field activities.

*** Data collection was suspended due to administrative consideration between 23 December 2011 and 12 March 2012.

[7] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[8] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

10.3 Production and response rates

This section presents data on each country's production in terms of completes and response rates.

Table 10.2: Target and Actual Number of Completed Cases and Response Rates

	Number of Completed Cases		Overall Response Rate (without Reading Components)	
	Target	**Actual**	**Target**	**Actual**
Australia	7,928[3]	7,428	80	71
Austria	5,000	5,130	51	53
Canada	--[4]	27,285[3]	65	59
Cyprus[9]	4,500	5,053	67	73
Czech Republic	6,000[3]	6,102	70	66
Denmark	6,800[3]	7,328	60	50
England (UK)	5,000	5,131	56	59
Estonia	7,500[3]	7,632	61	63
Finland	5,150	5,464	64	66
Flanders (Belgium)	5,000	5,463	48	62
France	5,460	6,993	52	67
Germany	4,925	5,465	51	55
Ireland	5,600	5,983	64	72
Italy	4,455	4,621	51	56
Japan	5,000	5,278	50	50
Korea	5,000	6,667	80	75
Netherlands	5,000	5,170	50	51
Northern Ireland (UK)	3,600	3,761	56[2]	65
Norway	5,000	5,128	59	62
Poland	9,041[1,3]	9,366	57[1]	56
Russian Federation[10]	5,000[1]	3,892	54[1]	52
Slovakia	5,568[3]	5,723	65	66
Spain	5,876	6,055	45	48
Sweden	5,000	4,469	50	45
United States	5,000	5,010	68	70

Source: Data Collection Form submitted after data cleaning and Survey Design International File, unless otherwise noted.

[1] Based on Data Collection Form submitted after conclusion of data collection.
[2] A specific response rate target for Northern Ireland (UK) was not reported so it was assumed to be the same as for England.
[3] Country with oversamples and/or special populations.
[4] Not reported.

[9] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[10] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

(See Chapter 14 to understand why the targets for completes are different across the countries.) As shown in Table 10.2, all countries except Sweden met the PIAAC target number of completes required by the TSG. Sweden only completed 4,500 of the 5,000 completed cases normally required when administering problem solving.

The TSG requires countries to achieve a 70% overall response rate. However, the TSG also indicates that a response rate of 50% or better is acceptable if the results of a nonresponse bias analysis (when necessary) determine no significant bias is present. As noted above, at the end of this round of data collection, 22 countries had achieved a response rate of 50% or better.

In addition, for planning purposes, countries were asked to estimate their "expected" response rates at the beginning of the study. Seventeen countries met or exceeded their estimated target (Austria, Cyprus,[11] England-UK, Northern Ireland-UK Estonia, Finland, Flanders (Belgium), France, Germany, Ireland, Italy, Japan, Netherlands, Norway, Slovakia, Spain, and the United States).

To provide some context, Table 10.3 presents a comparison of PIAAC response rates to prior adult literacy surveys response rates – IALS and ALL – and to past rounds of the European Social Survey (ESS). This is provided for information purposes only as response rates are calculated differently across studies. PIAAC response rates are not directly comparable to past literacy studies such as IALS and ALL because of stringent restrictions on the level of exclusions permitted in PIAAC.

More details about sample sizes and response rates can be found in Chapter 16.

[11] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

Table 10.3: PIAAC Overall Response Rates Compared to Other Recent International Surveys

	European Social Survey					IALS[6]		ALL[7]	PIAAC
	2002[1]	2004[2]	2006[3]	2008[4]	2010[5]	1994	1998	2003	2012
Australia	--	--	--	--	--	96.0	--	--	71
Austria	60.4	62.4	64.0	--	--	--	--	--	53
Canada	--	--	--	--	--	69.0	--	66.0	59[3]
Cyprus[12]	--	--	67.3	78.7	69.7	--	--	--	73
Czech Republic	43.3	55.3	--	69.5	70.2	--	61.5	--	66
Denmark	67.6	64.2	50.8	53.9	55.4	--	65.7	--	50
England (UK)	55.5	50.6	54.6	55.8	56.3	63.0	--	--	59
N. Ireland (UK)	55.5	50.6	54.6	55.8	56.3	63.0	--	--	65
Estonia	--	79.1	65.0	57.4	56.2	--	--	--	63
Finland	73.2	70.7	64.4	68.4	59.5	--	69.1	--	66
Flanders (Belgium)	59.2	61.2	61.0	58.9	53.4	36.0	--	--	62
France	43.1	43.6	46.0	49.4	47.1	--	--	--	67
Germany	55.7	51.0	54.5	48.0	30.5	69.0	--	--	55
Ireland	64.5	62.5	56.8	51.6	58.3	60.0	--	--	72
Italy	43.7	59.3	--	--	--	--	--	44.0	56
Japan	--	--	--	--	--	--	--	--	50
Korea[2]	--	--	--	--	--	--	--	--	75
Netherlands	67.9	64.3	59.8	49.8	60.0	45.0	--	--	51
Norway	65.0	66.2	65.5	60.4	58.0	--	60.9	56.0	62
Poland	73.2	73.7	70.2	71.2	70.3	75.0	--	--	56
Russian Federation[13]	--	--	69.5	67.9	66.6	--	--	--	52
Slovakia	--	62.7	73.2	72.6	74.7	--	--	--	66
Spain	53.2	54.9	65.9	66.8	68.5	--	--	--	48
Sweden	69.5	65.4	65.9	62.2	51.0	60.0	--	--	45
United States	--	--	--	--	--	60.0	--	66.0	70

Source: Data Collection Form submitted after data cleaning and Survey Design International File, unless otherwise noted.

[1] ESS1 - 2002 Summary and deviations. http://ess.nsd.uib.no/ess/round1/deviations.html.
[2] ESS2 - 2004 Summary and deviations. http://ess.nsd.uib.no/ess/round2/deviations.html.
[3] ESS3 - 2006 Summary and deviations. http://ess.nsd.uib.no/ess/round3/deviations.html.
[4] ESS4 - 2008 Summary and deviations. http://ess.nsd.uib.no/ess/round4/deviations.html.
[5] ESS5 - 2010 Summary and deviations, http://ess.nsd.uib.no/ess/round5/deviations.html.
[6] "Literacy in the Information Age." Final report of the International Adult Literacy Survey. Table B.6a and B.6b, p. 119, http://www.oecd.org/education/highereducationandadultlearning/41529765.pdf.
[7] "Learning a Living." First results of the Adult Literacy and Lifeskills Survey, Table B8, p. 327, http://www.oecd.org/education/educationeconomyandsociety/34867438.pdf .

[12] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[13] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

10.4 Interviewer training

Providing timely, adequate, standardized training to interviewers is an important tool in promoting the collection of quality data. Interviewers need to be very familiar with the survey procedures to administer them consistently across respondents and to produce data as error-free as possible. Familiarity with survey procedures allows interviewers to focus on gaining respondent cooperation, which in turn should help maximize response rates.

Chapter 9 of the TSG is dedicated entirely to training. It covers a variety of aspects associated with a successful training approach. Countries were, at a minimum, expected to:

- Conduct interviewer training in person, very close to the start of data collection.
- Train interviewers in small groups of 15-20.
- Assemble training staff to include a knowledgeable lead trainer, assistant(s), and technical support.
- Offer an adequate level of training. Although the Consortium recommended that countries should offer the same amount of training to all interviewers regardless of their level of experience, guidance was provided to tailor training to the level of experience of interviewers. (About 15 hours was recommended for experienced interviewers and 30 hours for interviewers new to survey research.)
- Provide sufficient hours of in-person training on **BQ and Direct Assessment administration** in the form of scripted mock interviews in which interviewers take turns reading the questions and a respondent (trainer or other interviewer) provides scripted answers. (About four hours recommended for each.)
- Provide sufficient hours of in-person training on **gaining cooperation** in the form of lectures and roundtable exercises where experienced interviewers are placed in groups with less experienced interviewers to discuss effective strategies for dealing with reluctant respondents. (About four hours recommended.)

10.4.1 Training logistics

The Consortium's recommendation was to conduct interviewer training the week before the start of data collection so interviewers could quickly apply the techniques learned and minimize learning loss. As shown in Table 10.4, most countries (68%) conducted interviewer training one to two weeks prior to the beginning of data collection. A significant proportion of countries (32%) held interviewer training sessions three weeks or more prior to data collection (Canada, France, Ireland, Italy, Japan, Netherlands, Norway and Poland). These countries typically organized several training sessions staggered in time so that only a fraction of interviewers received their training just before beginning work; for the first groups of interviewers to be trained, there was a considerable lag between training and data collection.

Six countries (24%) continued to train interviewers long after the start of data collection (more than four months) by organizing supplemental training sessions to compensate either for interviewer attrition or insufficient initial staffing.

A total of 380 interviewer training sessions were held worldwide, with numbers of sessions per country ranging from two in the Russian Federation[14] to 72 in Canada. The duration of training sessions varied significantly within and across countries. For example, the Netherlands held training that lasted between one and two days, while sessions held by Ireland lasted six to seven days.

Table 10.4: Summary of Main Study Interviewer Training Logistics

	Interviewer Training				
	Date Began	**Date Ended**	**Data Collection Start Date**	**Number of Sessions Held**	**Number of Days Per Event**
Australia	28 Sept 2011	25 Jan 2012	1 Oct 2011	15	3
Austria	11 July 2011	11 Nov 2011	1 Aug 2011	8	2-3
Canada	3 Oct 2011	6 Apr 2012	1 Nov 2011	72	4-5
Cyprus[15]	23 Aug 2011	7 Dec 2011	1 Sept 2011	9	2
Czech Republic	12 Aug 2011	14 Jan 2012	15 Aug 2011	15	2-3
Denmark	25 Aug 2011	11 Sept 2011	28 Aug 2011	4	2-4
England (UK)	18 July 2011	18 Nov 2011	1 Aug 2011	26	2
Estonia	12 July 2011	15 Dec 2011	1 Aug 2011	11	2-4
Finland	16 Aug 2011	7 Sept 2011	30 Aug 2011	7	2
Flanders (Belgium)	16 Aug 2011	18 Nov 2011	19 Aug 2011	7	3
France	4 July 2012	5 Sept 2012	7 Sept 2012	63	3
Germany	18 July 2011	12 Aug 2011	1 Aug 2011	5	3-5
Ireland	23 June 2011	28 July 2011	1 Aug 2011	3	6-7
Italy	22 June 2011	29 Sept 2011	1 Sept 2011	10	2-3
Japan	4 July 2011	29 July 2011	30 July 2011	14	4
Korea	15 Sept 2011	9 Mar 2012	26 Sept 2011	13	5
Netherlands	27 June 2011	12 Aug 2011	22 Aug 2011	16	1-2
Northern Ireland (UK)	25 July 2011	4 Nov 2011	1 Aug 2011	14	2
Norway	20 June 2011	30 Sept 2011	17 Aug 2011	12	2-5
Poland	6 July 2011	8 Feb 2012	1 Aug 2011	7	3
Russian Federation[16]	7 Nov 2011	2 Dec 2011	21 Nov 2011	2	3-4
Slovakia	6 Oct 2011	31 Jan 2012	27 Oct 2011	8	2
Spain	29 Aug 2011	2 Feb 2012	2 Sept 2011	29	3-4
Sweden	16 Aug 2011	2 Sept 2011	22 Aug 2011	6	1-3
United States	18 Aug 2011	13 Jan 2012	25 Aug 2011	4	4-6

Source: Interviewer Training Forms

[14] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

[15] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[16] Please refer to the above note regarding the Russian Federation.

At each training session, countries were required to have at least one lead trainer, one assistant trainer and one person responsible for technical support. The lead trainer requirement was met by all countries. However, 52% of countries conducted some training sessions without an assistant and/or technical support staff (Table 10.5).

In addition, 17 countries (68%) exceeded the maximum number of 20 trainees per training room in some sessions.

Table 10.5: Interviewer Training Staffing and Class Sizes for the Main Study

	Number of Training Staff Per Session/Room			
	Lead	**Assist**	**Tech**	**Number of Trainees Per Session/Room***
Australia	1-2	0-4	1	2-22
Austria	2	1-4	2-3	9-26
Canada	1-2	0-1	0	1-26
Cyprus[17]	1-2	2-3	1-2	8-39
Czech Republic	3-4	0-2	1	8-21
Denmark	3-6	3-6	1	35-66
England (UK)	1-3	0-4	0-2	8-17
Estonia	5	3	5-6	7-20
Finland	2	5	1	11-23
Flanders (Belgium)	2-3	1	1	5-19
France	1-2	0-1	0-3	3-10
Germany	1-3	2-3	1	18-31
Ireland	2	1-2	2	15-23
Italy	2-3	0-4	0-3	14-22
Japan	1	0-2	1-2	9-23
Korea	2	1-2	0	2-58
Netherlands	4	1	4	Not reported
Northern Ireland (UK)	1	0-1	1	9-15
Norway	2-3	1-6	1-2	8-29
Poland	1	0-2	2-4	12-74
Russian Federation[18]	2	4	1	83-87
Slovakia	1-3	0-3	0-1	2-38
Spain	1-2	0-2	0-1	1-9
Sweden	5	1	1	20-24
United States	1-2	0-1	1-2	15-17

Source: Interviewer Training Form.

*A range indicates that a country conducted multiple training sessions with varying numbers of training staff and trainees. Only the minimum and maximum are reported here.

[17] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[18] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

10.4.2 Content covered

As shown in Table 10.6, the Consortium proposed an interviewer training program of approximately 15 hours for returning Field Test interviewers with good reviews (Profile 1) and 36 hours for new interviewers without any interviewing experience (Profile 4).[19] For interviewers with some experience on other surveys (Profile 3) or those with less than favorable Field Test reviews (Profile 2), the requirements were to essentially train interviewers as if they were new except that they could be exempted from training on administrative procedures and the case management system.

As countries were allowed to tailor their training program to their interviewers' particular needs, it is somewhat challenging to evaluate the adequacy of training offered. However, there were a certain number of topics for which virtually no tailoring was allowed for interviewers without PIAAC Field Test experience. These are BQ and assessment administration and gaining cooperation. For these topics (as well as others), the Consortium had provided detailed training materials that countries were required to use. As can be seen in Table 10.6, the time requirements in hours for these topics were essentially the same for Profiles 2, 3 and 4, that is, seven to 10 hours on BQ administration and 5.5 hours for assessment administration.

After completion of the Main Study, the Consortium realized that the BQ training materials required only four to five hours to be administered. Consequently, for the purpose of this report, countries were evaluated against this revised requirement. The duration of the assessment administration training was revised slightly to four hours.

[19] For countries using a screener, an additional two hours of training on screener administration was recommended.

Table 10.6: Required and Optional Components of Interviewer Training by Interviewer Profile (Main Study)

Interviewer Training Topic	Profile 1 PIAAC Field Test interviewers with good reviews	Profile 2 Interviewers with less than favorable PIAAC	Profile 3 Interviewers with some experience on other surveys	Profile 4 Interviewers without any interviewing experience
Introduction	0.50	0.50	1.75	1.75
Preparing for the field	0.50	0.50	0.50	0.50
CAPI training			1.50	1.50
Locating households/respondents		1.50	1.50	1.50
Case management system		(1.50)	(1.50)	1.50
Screener interactive, **if applicable**	1.25	1.25	2.00	2.00
BQ	4.00	7.00	8.00	10.00
Disposition codes	1.00	1.00	1.00	1.00
Direct Assessment	4.75	5.50	5.50	5.50
Core scoring	0.50	0.50	0.50	0.50
QC/Admin	(0.50)	(2.50)	(2.50)	2.50
Gaining respondent cooperation		4.50	2.50	4.50
Practice interview (role play)	1.75	1.75	3.25	3.25
Live respondent practice	2.00	2.00	2.00	2.00
Total hours for countries with list samples	**15.50 (15)**	**28.75**	**32 (28)**	**36**

Source: Clarifications Regarding Main Study Interviewer Training Requirements, 30 March 2011.

As shown in Table 10.7, fewer than half of countries (40%) met or exceeded the number of hours dedicated to gaining cooperation training (about four hours for new interviewers and two hours for those with prior experience; no gaining cooperation training was necessary for returning PIAAC Field Test interviewers with good reviews). Another 16% of countries met the requirement for some of the interviewers.

Only about half of countries (15, or 60%) spent the minimum recommended amount of time on BQ administration (four hours, regardless of level of experience). Another 17% met the requirement for some of their interviewers only. Only 14 (56%) of countries met the minimum number of hours required for assessment administration (about four hours). Another 17% of countries met the requirement partially (i.e., for some of their interviewers only).

Table 10.7: Actual Training Time Spent on BQ Administration, Assessment Administration and Gaining Cooperation, by Training Type and Interviewer Profile (Main Study)

	Training Type*	Number of Trainees	Majority Profile in Group	Hours In Person	BQ Total**	DA Total***	Coop Total
Australia	N/A	220	varied	19.75	4.5	3.6	.7
Austria	Full	50	3	28	5.5	6	1.5
	Reduced	97	varied	16.5	2	3.5	.5
Canada	N/A	759	varied	37.5	9	8.5	7
Cyprus[20]	N/A	150	4	18	2.5	2.8	1.5
Czech Republic	Full	159	varied	16.1	2.7	2.2	3
	Reduced	49	3	12	2.2	2.5	1.5
Denmark	Full	155	3	26	5	4	3.5
	Reduced	56	1	15	2.5	2	1
England (UK)	N/A	328	varied	10	1.3	3.5	.8
Estonia	Full	70	varied	33	7	4	4
	Reduced	43	3	24	5	4	1
	Reduced	19	3	17	4	3	0
Finland	N/A	124	1	15	4	2	2
Flanders (Belgium)	N/A	101	3	24	5	4	3
France	N/A	508	varied	18	4	3	.3
Germany	Full	91	3	31	6	7.8	3
	Reduced	38	1	22.3	3.5	6.3	3
Ireland	Full	38	3	44.5	6	8	5.5
	Reduced	18	3	38	4	7	4.5
Italy	N/A	170	varied	27	7	5	0
Japan	N/A	205	varied	23.8	2.8	3.3	1
Korea	N/A	229	3	30	7	6	3
Netherlands	Full	165	3	14.5	3.5	2.5	1
	Reduced	100	1	7.5	1	1.5	.5
Northern Ireland (UK)	N/A	165	3	10	1.3	3.5	.8
Norway	Full	42	4	19	2	5	0
	Reduced	15	1	14	0	6	4
	Reduced	98	1	16	1	3	1
Poland	N/A	236	3	25	6	4.5	2
Russian Federation[21]	Full	87	3	34.6	7.1	6.1	2.2
	Reduced	83	3	31.2	6.3	5.6	1.5
Slovakia	N/A	105	varied	20.4	5.5	3.5	.9
Spain	N/A	113	varied	18	3.0	4.0	1.7

[20] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[21] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

Table 10.7 (cont.): Actual Training Time Spent on BQ Administration, Assessment Administration and Gaining Cooperation, by Training Type and Interviewer Profile (Main Study)

	Training Type*	Number of Trainees	Majority Profile in Group	Hours In Person	BQ Total**	DA Total***	Coop Total
Sweden	Full	64	3	17.9	4	3.8	1.5
	Reduced	68	1	7.6	0	2.3	1.5
United States	Full	186	3	38.3	4.8	5.3	5.5
	Full	3	3	32	4.3	4.5	4
	Full	6	3	31.8	4.5	5	2.5

Source: Interviewer Training Form.

* Several groups of interviewers may have been offered the same training type. For conciseness, groups offered the same training are not listed separately in this table.
** Includes time spent at in-person training on Introduction to CI/BQ administration, BQ interactives, and BQ exercises.
*** Includes time spent at in-person training on Introduction to Direct Assessment, Direct Assessment interactives, and Core Scoring.

The data in Table 10.7 suggest that several countries made significant adaptations to interviewer training scripts provided by the Consortium. Countries were permitted to make adaptations to Consortium training materials to fit their specific situation (mostly BQ adaptations), but these adaptations were not expected to dramatically affect the time spent on training.

The recommended amount of time to spend on BQ and assessment administration was deemed necessary for interviewers to get exposure to each question and become comfortable with the instruments. Interviewers must be unhindered by the technical aspects of survey administration to be able to focus on one of the most challenging part of their job – obtaining and maintaining cooperation from respondents. Spending significantly less time than recommended on these critical topics may have negatively affected response rate and/or data quality in many countries.

10.5 Staffing and field management

Hiring a sufficient number of fieldworkers (supervisors and interviewers), close supervision, and monitoring of production goals and response rates are fundamentals of successful fieldwork.

10.5.1 Interviewer hiring and attrition

Each country was required to hire a sufficient number of interviewers to achieve that country's production goals in eight months (see Table 10.2 for production targets). Because the optimal number of interviewers depends on numerous country-specific factors, the Consortium could not determine the exact number each country needed. However, TSG 8.3.1 provided specific considerations for countries. National teams were advised to use the best information available from similar national surveys conducted in their country as well as interviewers' PIAAC Field Test experience. Countries with compressed data collection schedules were advised to adjust their staffing needs accordingly.

Table 10.8 provides detailed information about staffing levels and attrition and suggests that countries learned from their Field Test experience and adjusted their staffing for the Main Study.

Twenty-three countries hired more than 100 interviewers (between 102 to 786 interviewers). Only two geographically small countries hired fewer than 100 – Ireland (61) and Cyprus[22] (84).

About 40% of countries experienced substantial levels of interviewer attrition (above 20%). All but four countries (88%) had some interviewer resignations. About 10 countries (40%) laid off interviewers, and 64% dismissed interviewers due to poor productivity or quality control issues.

Table 10.8: Data Collection Staffing and Attrition

	Number of Interviewers					Causes of Attrition				
	Attended Training	Received Assignment	Working at the End of Study	Typical Hours Worked Per Week	Attrition Rate (%)	Quit	Laid off	Productivity	Quality Control	Other
Australia	229	229	189	15-30	17	x				x
Austria	151	150	142	15	5	x	x	x		
Canada	810	786	274	5-25	65	x	x	x		
Cyprus[23]	150	84	5	20-40	94	x	x	x	x	
Czech Republic	194	194	74	20-40	62	x	x	x	x	
Denmark	216	216	192	8-20	11	x	x	x		x
England (UK)	343	328	243	10-25	26	x	x			
Estonia	127	124	75	30-40	40	x		x		
Finland	124	124	122	15-20	2	x				
Flanders (Belgium)	102	102	35	20	66	x		x		x
France	508	508	506	2-20	≈0					x
Germany	129	129	125	--[**]	3		x			
Ireland	70	61	40	25	34	x		x		
Italy	170	170	159	25-35	6	x		x	x	
Japan	228	226	224	5-35	1			x		x
Korea	220	220	216	40	2	x				
Netherlands (The)	275	275	167	10-15	39	x	x	x		
Northern Ireland (UK)	186	186	181	10	3	x				
Norway	140	140	134	10-25	4	x				
Poland[*]	286	286	196	18	31	x				x
Russian Federation[*] [24]	170	140	140	15-42	0	x			x	
Slovakia	107	107	97	8	9	x	x	x		x
Spain	144	139	117	30-40	16	x		x	x	
Sweden	145	137	135	10-15	2				x	x
United States	195	192	50	25-40	74	x	x	x	x	

Source: Data Collection Form submitted after data cleaning unless otherwise noted.

[*] Based on Data Collection Form submitted after conclusion of data collection.
[**] Not reported.

[22] The number of interviewers hired by countries depended on several factors. For example, most countries had interviewers working part time while others had interviewers working full time on PIAAC (see Table 10.8 for the typical number of hours worked by PIAAC interviewers in each country).
[23] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[24] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

10.5.2 Field management

Two key indicators of adequate field management are: (1) the supervisor-interviewer ratio and (2) the frequency/regularity of supervisor-interviewer meetings.

In terms of the **interviewer-supervisor ratio,** countries were advised to assign one supervisor for every 15-20 interviewers to support the close supervision and mentoring of data collection. Table 10.9 indicates that 16 countries (64%) adhered to the recommended ratio of 20:1. However, when the ratio is increased to 30:1, only one country (Netherlands) stands out as far exceeding the Consortium recommendation with a ratio of 55:1.

Table 10.9: Number of Interviewers per Supervisor

	Number of Interviewers Who Received Assignments	Number of Supervisors	Size of Supervisor Assignment
Australia	229	10	15-22
Austria	150	6	27
Canada	786	80	8-10
Cyprus[25]	84	4	10-20
Czech Republic	194	6	15-25
Denmark	216	8	20-30
England (UK)	328	63	1-20
Estonia	124	8	11-15
Finland	124	6	20-30
Flanders (Belgium)	102	4	25
France	508	44	6-20
Germany	129	8	15-25
Ireland	61	4	12-14
Italy	170	10	10-20
Japan	226	31	2-20
Korea	220	61	2-5
Netherlands (The)	275	5	55
Northern Ireland (UK)	186	20	10
Norway	140	7	15-20
Poland*	286	50	2-6
Russian Federation* [26]	140	24	5-20
Slovakia	107	6	12-16
Spain	139	18	4-12
Sweden	137	6	23
United States	192	11	16-19

Source: Data Collection Form submitted after data cleaning unless otherwise noted.

*Based on Data Collection Form submitted after conclusion of data collection.

[25] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[26] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

With regard to adequate *communication between field staff*, the TSG calls for weekly phone or in-person communication among the various levels of field staff and email communication as necessary. In particular, field supervisors should have weekly calls with their interviewers to ensure steady and adequate progress in data collection by keeping all staff on task, and making them accountable for their progress or lack thereof. Discussion during the meetings should focus on progress through caseload, response rates, problems encountered, and strategies/solutions for the completion of their remaining cases. Meeting sporadically can result in failures to meet data quality and production goals.

The majority of countries (16, or 64%) followed communication recommendations. Another six countries either had meetings every other week or less often (Finland, Poland, the Russian Federation[27]) or had variation across regions (Canada, Slovakia, and Spain). Only three countries had no scheduled meetings and opted to have meetings only as needed (Austria, Czech Republic, the Netherlands).

Countries used a variety of modes to communicate with their field staff. All countries used phone and all countries, but two used email (Denmark and Slovakia). Other strategies such as in-person meetings and newsletters were used by slightly more than half of countries. Some countries mentioned the use of newer technologies such as an online forum and video conferencing.

Details regarding the modes and frequency of communication are presented in Table 10.10.

Table 10.10: Modes of Communication Used between Field Staff during Data Collection

	Modes of Communication Used					
Country	**In Person**	**Phone**	**Email**	**Newsletter**	**Other**	**Frequency**
Australia		**x**	**x**	**x**	Lotus Notes database	Weekly
Austria	**x**	**x**	**x**			As needed
Canada	**x**	**x**	**x**	**x**		Varies
Cyprus[28]	**x**	**x**	**x**		Secure FTP Server, web service	Daily
Czech Republic		**x**	**x**	**x**		As needed
Denmark		**x**				As needed, weekly
England (UK)	**x**	**x**	**x**	**x**		As needed, weekly
Estonia	**x**	**x**	**x**		Online forum	As needed, weekly
Finland	**x**	**x**	**x**	**x**	Online forum	As needed, biweekly
Flanders (Belgium)	**x**	**x**	**x**			As needed, weekly
France	x	**x**	**x**			As needed, weekly
Germany		**x**	**x**	**x**		As needed, weekly
Ireland	**x**	**x**	**x**	**x**	Group briefing every 2 months	As needed, weekly

[27] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

[28] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

Table 10.10 (cont.): Modes of Communication Used between Field Staff during Data Collection

	Modes of Communication Used					
Country	In Person	Phone	Email	Newsletter	Other	Frequency
Italy	x	x	x		Video conferencing	As needed, weekly
Japan	x	x	x		Fax, message of Main Study	As needed, weekly
Korea	x	x	x		Q&A on the website	As needed, 2-3 times a week
Netherlands (The)		x	x	x		As needed, daily if necessary
Northern Ireland (UK)		x	x	x		As needed, weekly
Norway	x	x	x	x		As needed, weekly
Poland*	x	x	x			As needed, biweekly
Russian Federation* [29]		x	x	x	Video conferencing	Biweekly
Slovakia		x		x		Varies
Spain	x	x	x	x	Agency website	Varies
Sweden		x	x	x		Weekly
United States	x	x	x			As needed, weekly

Source: Data Collection Form submitted after data cleaning unless otherwise noted.

* Based on Data Collection Form submitted after conclusion of data collection.

10.6 Nonresponse reduction

Implementation of a comprehensive strategy to promote survey participation is a key element in obtaining acceptable and/or high response rates. Doing so requires the use of a variety of outreach materials and tools, the collection of information on contact attempts and nonresponse, and strategies specifically aimed at minimizing nonresponse. In addition, countries were strongly advised, but not required, to offer a respondent incentive as a way to increase participation.

10.6.1 Use of introductory materials and outreach tools

Countries were required to send an introductory letter to households/respondents in advance of the interviewer visit and were advised to use a variety of tools to increase the visibility and legitimacy of the study. Table 10.11 shows that virtually all countries used an introductory letter, a study brochure, a study-specific website, and a respondent help line.[30] Endorsement letters, newspaper articles and press releases were used by about half of countries. Few countries made use of radio or TV advertisements.

[29] Please refer to the above note regarding the Russian Federation.

[30] This is a telephone line that potential respondents can call to receive additional information about the survey. The number for this line is usually provided in the introductory letter or the study brochure.

With regard to the use of respondent help lines by potential respondents, Table 10.11 shows that countries received widely varying numbers of calls. Among countries providing counts, Estonia received the fewest, with 20 calls, and Korea received the most, with 1,739 calls.

In addition, some countries participated in TV shows, held press conferences, and placed ads on the web and social media.

Table 10.11: Introductory Materials Used in Presenting the Study to Respondents/Households

	Intro. Letter	Study Brochure	Endorsement Letter	Newspaper Article	TV Ads	Radio Ads	Press Release	Study-Specific Website	Respondent Helpline (# calls)	Other
Australia	x	x						x	x (n.r.)	
Austria	x	x						x	x (400)	
Canada	x	x						x	x (1491)	
Cyprus[31]	x	x	x	x			x	x	x (133)	
Czech Republic	x	x	x	x	x	x	x	x	x (386)	
Denmark	x	x		x		x	x	x	x (505)	
England (UK)	x	x						x	x (823)	
Estonia	x	x		x	x	x	x	x	x (20)	posters, video, web ads
Finland	x	x	x	x			x	x		TV show, social network ads
Flanders (Belgium)	x	x						x	x (375)	
France	x	x							x (500)	letters to mayor's office and police stations
Germany	x	x	x	x			x	x	x (307)	flyers
Ireland	x	x	x	x			x	x	x (115)	
Italy		x	x	x			x	x	x (168)	press conference
Japan	x	x	x	x	x		x	x	x (1644)	
Korea	x	x	x					x	x (1739)	posters, banners
Netherlands (The)	x	x						x	x (400)	
Northern Ireland (UK)	x	x						x	x (242)	
Norway	x	x		x			x	x	x (912)	Main Study messages
Poland*	x	x		x			x	x	x (90)	refrigerator magnet
Russian Federation*[32]	x	x	x					x		
Slovakia	x	x		x		x	x	x	x (90)	call-back cards
Spain	x	x						x	x (198)	letters to local councils/condos
Sweden	x	x		x			x	x	x (n.r.)	radio/TV interviews
United States	x	x	x					x	x (183)	refrigerator magnet, tailored flyers, pens

Source: Data Collection Form submitted after data cleaning unless otherwise noted.

* Based on Data Collection Form submitted after conclusion of data collection.

[31] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[32] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

10.6.2 Documenting contact attempts

Countries were advised to require interviewers to thoroughly document each contact attempt with sample persons/households and to record as much information as possible on nonresponse cases. The purpose was to provide supervisors with the information necessary to manage the work of interviewers effectively and help them be productive.

The information recorded for each contact attempt had to include, at a minimum, the date, time and outcome of each visit. Interviewers were also supposed to provide comments that might prove helpful in obtaining respondent cooperation during future contacts.

Table 10.12 provides a summary of the information recorded by countries about each contact attempt and nonresponse cases. All countries recorded all elements recommended about contact attempts. However, a significant number of countries (n=7) did not provide an opportunity for interviewers to write comments about the case, which can be very helpful when planning nonresponse work.

Table 10.12: Information Collected by Interviewers about Contact Attempts during Main Study Data Collection

	Day	Date	Time	Mode	Outcome	Comments	Other
Australia	x	x	x	x	x		
Austria	x	x	x	x	x		
Belgium	x	x	x	x	x	x	
Canada	x	x	x	x	x	x	Several other
Cyprus[33]	x	x	x	x	x	x	
Czech Republic	x	x	x	x	x	x	
Denmark	x	x	x	x	x		
England (UK)	x	x	x	x	x	x	
Estonia	x	x	x	x	x	x	
Finland	x	x	x	x	x	x	
France	x	x	x	x	x	x	
Germany	x	x	x	x	x	x	
Ireland	x	x	x	x	x	x	
Italy	x	x	x	x	x	x	
Japan	x	x	x	x	x		
Korea	x	x	x	x	x	x	
Netherlands	x	x	x	x	x		
Northern Ireland (UK)	x	x	x	x	x	x	
Norway	x	x	x	x	x	x	Interviewer ID
Poland*	x	x	x	x	x	x	
Russian Federation*[34]	x	x	x	x	x	x	
Slovakia	x	x	x	x	x		
Spain	x	x	x		x	x	
Sweden	x	x	x	x	x		
United States	x	x	x	x	x	x	Interviewer ID

Source: Data Collection Form submitted after data cleaning unless otherwise noted.

* Based on Data Collection Form submitted after conclusion of data collection.

[33] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[34] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

10.6.3 Monitoring contact attempts

At a minimum, countries were required to ensure that a minimum number of contact attempts were made to each respondent/household. Countries were strongly advised to attempt initial contacts in person and to make at least three subsequent contact attempts. Countries for which telephone initial contacts are customary were allowed to deviate from this standard but were required to make at least six subsequent attempts.

Table 10.13 presents details of the contact procedures used by participating countries. It shows that 21 countries (84%) used in-person initial contacts, either exclusively or in combination with the telephone. An additional three countries (France, Italy and the Russian Federation[35]) used a hybrid strategy in which some individuals were initially contacted by personal visits and others by telephone. All countries met the minimum number of contacts required with respect to their mode choice.

Table 10.13: Strategy for Contacting Potential Respondents/Households during Main Study Data Collection

	Mode of Initial Contact		Minimum Number of Subsequent Contacts	
	In Person	**Telephone**	**In Person**	**Telephone**
Australia	**x**		5	5
Austria	**x**		4	0
Canada	**x**		5	20
Cyprus[36]	**x**		4	5
Czech Republic	**x**		5	0
Denmark	**x**		5	0
England (UK)	**x**		6	0
Estonia	**x**		7	2
Finland		**x**	4	0
Flanders (Belgium)	**x**		5	0
France	**x**	x	5	7
Germany	**x**		4	0
Ireland	**x**		4	0
Italy	**x**	**x**	4	7
Japan	**x**		4	0
Korea	**x**		4	7
Netherlands	**x**		6	0
Northern Ireland (UK)	**x**		3	0
Norway		**x**	3	7
Poland*	**x**		4	0
Russian Federation* [37]	**x**	**x**	4	7
Slovakia	**x**		4	0
Spain	**x**		6	4
Sweden		**x**	0	10
United States	**x**		4	0

Source: Data Collection Form submitted after data cleaning unless otherwise noted.

[35] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

[36] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[37] Please refer to the above note regarding the Russian Federation.

[*] Based on Data Collection Form submitted after conclusion of data collection.

Finland, Norway and Sweden used the telephone as the sole mode for initial contact, although in-person visits were used by Finland and Norway to supplement the telephone strategy. Sweden made no in-person visits, except when interviewers could do so on the way to an appointment.

10.6.4 Documenting nonresponse

In addition to recording information about each contact attempt, countries were also required to record details about each case finalized as nonresponse. These included basic demographics about the person who refused, the strength of the refusal, the likelihood of conversion, any problems encountered, and any relevant information that might facilitate future contact with a potential respondent.

The level of detail recorded varied from country to country. However, all countries recorded basic information about nonrespondents, as shown in Table 10.14.

Table 10.14: Information Collected by Interviewers on Nonresponse Cases during Main Study Data Collection

	Demographics	Refusal Strength	Problems Encountered	Conversion Likelihood	Comments	Other
Australia	x[2]	x	x	x		
Austria	x	x	x	x	x	
Canada		x	x	x	x	x[3]
Cyprus[38]	x	x	x	x	x	
Czech Republic	x	x	x	x	x	
Denmark	x	x	x			
England (UK)	x	x	x	x	x	x[6]
Estonia	x	x	x	x	x	
Finland	x	x		x	x	x[4]
Flanders (Belgium)	x	x	x	x	x	
France	x	x	x	x		
Germany		x	x		x	x[5]
Ireland		x	x	x	x	
Italy	x	x	x		x	
Japan	x	x	x	x	x	
Korea	x	x	x	x	x	
Netherlands	x	x	x	x	x	
Northern Ireland (UK)	x	x	x	x	x	
Norway	x				x	
Poland[1]	x		x		x	
Russian Federation[1] [39]		x	x	x	x	
Slovakia	x	x	x	x	x	
Spain	x	x	x	x		
Sweden	x	x	x	x	x	
United States	x	x	x	x	x	x[7]

Source: Data Collection Form submitted after data cleaning unless otherwise noted.

[1] Based on Data Collection Form submitted after conclusion of data collection.
[2] Only if screener has been completed.
[3] Refusals are escalated to supervisor and manager level for resolution, and these steps are recorded in the case management system.
[4] Type of refusal.
[5] Presence of an intercom, house type, condition of the house, respondent's social class and education as appraised by the interviewer prior to first contact attempt.
[6] In one of the data collection agencies: recommendation for profile of interviewer who is more likely to be successful at converting the case.
[7] Name and phone number of a contact person.

In addition, countries were asked to report some of the most common reasons for refusal to do the BQ (Table 10.15) and assessment (Table 10.16). For the BQ, lack of interest was the most

[38] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[39] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

often cited reason across all countries, followed by lack of time ("too busy"). For the assessment, the excessive length ("too long") and lack of time were the most often cited refusal reasons.

Table 10.15: Most Common Reasons for Refusal to BQ (Main Study)

	Not Interested	Too Long	Don't Want To Be Bothered	Don't Trust Surveys	Too Busy	Other
Australia		**x**			**x**	
Austria	**x**					
Canada		**x**	**x**		**x**	
Cyprus[40]	**x**	**x**			**x**	
Czech Republic	**x**		**x**		**x**	
Denmark	**x**	**x**	**x**			
England (UK)	**x**		**x**		**x**	
Estonia	**x**		**x**		**x**	
Finland	**x**		**x**		**x**	
Flanders (Belgium)	**x**	**x**			**x**	
France	x	x	x			
Germany						**x**[2]
Ireland	**x**		**x**		**x**	
Italy	**x**	**x**		**x**	**x**	**x**[3]
Japan	**x**				**x**	**x**[4]
Korea	**x**	**x**	**x**			
Netherlands	**x**	**x**	**x**			
Northern Ireland (UK)	**x**		**x**		**x**	
Norway	**x**	**x**			**x**	
Poland[1*]	**x**		**x**		**x**	
Russian Federation[1*] [41]	**x**				**x**	
Slovakia	**x**		**x**		**x**	
Spain	**x**		**x**		**x**	
Sweden	**x**				**x**	**x**[5]
United States	**x**	**x**			**x**	

Source: Data Collection Form submitted after data cleaning unless otherwise noted.

[1*] Based on Data Collection Form submitted after conclusion of data collection.
[2] Legal guardian refused respondents' participation; respondent doesn't want to give more information (altogether three refusals).
[3] Literacy-related problems.
[4] Sickness, poor physical condition.
[5] Voluntary nature of the survey.

[40] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[41] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

Table 10.16: Most Common Reasons for Refusal to Assessment (Main Study)

	Not Interest-ed	Too Long	Don't Want To Be Bothered	Waste of Time and Money	Too Busy	Don't Want To Do Exercise	Too Complicated	Other
Australia		x			x			
Austria		x				x	x	
Canada		x			x	x		
Cyprus[42]		x			x			
Czech Republic						x		
Denmark		x				x	x	
England (UK)	x	x			x			
Estonia		x			x	x		
Finland		x				x	x	
Flanders (Belgium)		x			x	x		
France		x			x	x		
Germany			x			x	x	
Ireland		x			x		x	
Italy		x			x	x	x	
Japan								
Korea		x				x	x	
Netherlands		x					x	
Northern Ireland (UK)					x			
Norway								
Poland*		x		x	x			
Russian Federation* [43]	x	x			x			
Slovakia	x		x		x			
Spain	x		x		x			
Sweden	x				x			x**
United States	x	x			x			

Source: Data Collection Form submitted after data cleaning unless otherwise noted.

* Based on Data Collection Form submitted after conclusion of data collection.
** Voluntary nature of the survey.

10.6.5 Nonresponse strategy

Countries were strongly advised to implement a comprehensive strategy to deal with nonresponse cases. Most countries (92%) implemented a strategy involving a combination of techniques, such as case reassignment, senior interviewer follow-up and the use of tailored letters. Two countries had strategies involving only the use of case reassignment (UK-Northern Ireland) or supervisor follow-up combined with tailored letters (Korea). However, Korea and Northern Ireland (UK) offered substantial monetary incentives (64 and 37 Euros, respectively), and secured response rates at or above 65%. Table 10.17 presents each strategy in detail.

[42] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[43] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

Table 10.17: Strategies to Deal with Difficult/Nonresponse Cases during Main Study Data Collection

	Case Re-Assign.	Follow-Up Senior FIs	Follow-Up Supervisors	Traveling Reassignment	Tailored Letters	Refusal Conversion Letters	Other
Australia	x	x			x	x	
Austria	x	x				x	x[2]
Canada	x	x	x		x	x	
Cyprus[44]	x		x	x	x	x	
Czech Republic	x	x		x	x	x	
Denmark	x	x	x			x	
England (UK)	x	x	x	x		x	
Estonia	x				x	x	
Finland	x	x			x	x	
Flanders (Belgium)	x	x	x			x	
France			x		x	x	
Germany	x	x			x	x	x[3]
Ireland	x		x	x	x	x	
Italy	x		x	x			x[4]
Japan	x	x	x		x	x	x[5]
Korea			x		x	x	x[6]
Netherlands	x	x		x	x	x	
Northern Ireland (UK)	x						
Norway	x	x		x	x	x	
Poland[1*]	x	x	x			x	
Russian Federation[1* 45]	x		x		x		x[7]
Slovakia	x	x				x	x[8]
Spain	x		x		x	x	
Sweden	x	x	x	x	x	x	x[9]
United States	x	x	x	x	x	x	x[10]

Source: Data Collection Form submitted after data cleaning unless otherwise noted.

[1*] Based on Data Collection Form submitted after conclusion of data collection.
[2] Motivation calls to nonrespondents of certain groups identified in order to reduce bias.
[3] In certain cases, interviewers had access to funding for discretionary incentives (only symbolic) or received day rates (e.g., for refusal conversion).
[4] Refusal conversion calls.
[5] Offered option to conduct interview at home or out of home such as community hall. Designed mobile and PC websites to allow respondents to schedule appointment for interview.
[6] Field managers or field directors tried to persuade some respondents.
[7] Contact leaders of local communities and ethnic diasporas; contact building managers.
[8] Telephone calls to the households by field managers, supervisors.
[9] Group of interviewers dedicated to refusal conversion.
[10] Tailored flyers, mail-in screener forms sent to sampled households yet to be screened.

[44] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[45] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

10.6.6 Use of incentives

The vast majority of countries (23, or 92%) offered some form of incentive. Two countries (8%), Australia and Canada, have rules preventing the use of incentives in government surveys. Among countries that offered an incentive, 18 (78%) used a monetary incentive. Details regarding the nature of each country's incentive are provided in Table 10.18.

Table 10.18: Respondent Incentives Used During Main Study Data Collection

	Incentive type		
	Monetary**	Non-Monetary	Description
Australia			None
Austria	x		50 EUR shopping voucher
Canada			None
Cyprus[46]	x		50 EUR shopping voucher from popular furniture store
Czech Republic	x		500 CZK (approx. 20 EUR)
Denmark		x	100 DKR (approx. 13 EUR) to respondents who participated in live practice· Lottery ticket in the last phase of the data collection period
England (UK)	x	x	30 GBP (approx. 37 EUR) voucher; booklet of stamps as a refusal conversion tool in some areas.
Estonia		x	Magazine subscription
Finland		x	USB flash drive; lottery of popular tablet computer
Flanders (Belgium)		x	Lottery ticket (3 EUR)
France		x	Numeracy kit
Germany	x	x	Study-specific adhesive notepad sent to all with introductory letter and brochure; 50 EUR upon completion
Ireland	x		30 EUR shopping voucher
Italy	x		30 EUR shopping coupon (increased to 40 EUR in the last 10 weeks of fieldwork).
Japan		x	Book voucher
Korea	x		4 EUR for completed screener + 20 EUR for completed BQ + 40 EUR for completed assessment
Netherlands	x		20 EUR voucher (increased to 40 EUR in the final stage of data collection)
N. Ireland (UK)	x		30 GBP (approx. 37 EUR) voucher
Norway	x	x	Refrigerator magnet to all; 500 NOK (approx. 66 EUR) gift card upon completion
Poland*	x	x	8 EUR shopping voucher; lottery ticket
Russian Federation* [47]	x		300 RUB (approx. 7 EUR) or 500 RUB (approx. 12 EUR) depending on regions
Slovakia	x		10 EUR
Spain	x	x	Choice of 20 EUR voucher or equivalent donation to NGO
Sweden	x	x	Refrigerator magnet to all; 10 EUR check upon completion
United States	x	x	Study-specific refrigerator magnet and pen to all; 50 USD upon completion (approx. 40 EUR)

Source: Data Collection Form submitted after data cleaning unless otherwise noted.

* Based on Data Collection Form submitted after conclusion of data collection.

[46] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[47] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

[**] The distinction between monetary and nonmonetary incentive is somewhat subjective. Strictly speaking, anything other than cash or a check is not monetary. However, "shopping vouchers" were considered to be monetary incentives unless they could only be exchanged for very specific goods such as books or magazine subscriptions.

10.7 Fieldwork quality control

Each country was required to implement a comprehensive fieldwork quality control plan. This plan had to include:

- an evaluation of interviewing skills at the end of training and during data collection
- the implementation of a validation (callback) effort to detect falsification
- the review of survey and process data through the analysis of automated management reports

10.7.1 Audio recording/observation of interviews

Countries were strongly advised to monitor at least two interviews per interviewer during the early stages of data collection and provide feedback. Monitoring could either be done by audio recording interviews, observing the interviews in person, or a combination of both.

The vast majority of countries (22, or 88%) did some form of monitoring. Among these 22 countries, 15 monitored at least one interview per interviewer on average, but few reached the recommended level of two interviews per interviewer (see ratio of interviews monitored to number of interviewers assigned in Table 10.19). The Consortium's recommendation was to monitor the second and 10th complete achieved by each interviewer. However, some interviewers may not have been productive enough to allow for a country to monitor a second interview. Therefore, countries are considered to have met the standard if they have monitored at least one interview per interviewer on average. Ten countries did not meet this reduced standard. Australia, Austria and Northern Ireland (UK) did not monitor any interviews. Canada, Finland, France, Italy, the Netherlands, Norway and England (UK) did not monitor enough interviews given the number of interviewers they assigned to PIAAC.

Table 10.19 shows the number of interviewers assigned to PIAAC, the number of interviews that were audio recorded or observed in each country, and the ratio of interviews monitored to the number of interviewers assigned to PIAAC work.

Table 10.19: Number of Interviews Monitored by Mode during the Main Study Data Collection

	Number of Interviewers Assigned	Number of Interviews Monitored				Ratio of Interviews Monitored to Number of Interviewers Assigned
		Taping Full Interview	Taping Snippets	Observation	Total	
Australia	229	0	0	0	0	0
Austria	150	0	0	0	0	0
Canada	786	0	0	385	385	0.49
Cyprus[48]	84	121	0	0	121	1.44
Czech Republic	194	0	199	0	199	1.03
Denmark	216	440	0	0	440	2.04
England (UK)	328	0	0	41	41	0.13
Estonia	124	503	0	0	503	4.05
Finland	124	101	0	0	101	0.81
Flanders (Belgium)	102	135	0	0	135	1.32
France	508	0	0	400	400	0.79
Germany	129	245	4	0	249	1.93
Ireland	61	100	0	40	140	2.29
Italy	170	0	0	165	165	0.97
Japan	226	0	425	0	425	1.88
Korea	220	682	0	218	900	4.09
Netherlands	275	36	0	0	36	0.13
Northern Ireland (UK)	186	0	0	0	0	0
Norway	140	0	0	120	120	0.86
Poland*	286	1800	0	0	1800	6.29
Russian Federation* [49]	140	1250	0	0	1250	8.93
Slovakia	107	0	306	0	306	2.86
Spain	139	176	44	0	220	1.58
Sweden	137	274	0	0	274	2.00
United States	192	298	0	0	298	1.55

Source: Data Collection Form submitted after data cleaning unless otherwise noted.

* Based on Data Collection Form submitted after conclusion of data collection.

10.7.2 Validation

Validation (back-checks) is critical to data validity; it is the most important QC feature of household data collection. As stated in the TSG, the validation procedure required the verification of "10% of an interviewer's finalized work, including cases finalized as nonresponse." The validation plan had to ensure that:

- validation cases were selected randomly;
- at least 10 percent of each interviewer's cases were validated; and
- all dispositions were validated, not just completes

[48] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[49] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

Table 10.20 presents a summary of the validation procedures implemented by countries.

Table 10.20: Summary of Adherence to Validation Standards by Countries (Main Study)

Requirements	Implementation
Cases must be selected randomly, not ad hoc	20 countries (80%) selected most or all cases randomly
At least 10% of each interviewer's work must be validated	11 countries (44%) reached the 10% threshold for each interviewer 17 countries (68%) reached a threshold of 7% for 90% of interviewer
All types of disposition must be validated (completes, non-contacts, refusals, ineligibles)	22 countries (88%) validated all dispositions

Source: Data Collection Form submitted after data cleaning. For Poland and the Russian Federation,[50] the source is the Data Collection Form submitted after conclusion of data collection.

The requirement to validate *each* interviewer at the 10% level appears to have been the most challenging for countries to meet, as only 11 countries did so. Even when setting the threshold lower (7% of cases validated for 90% of interviewers), only 17 countries met this requirement.

Regarding other validation requirements, 20 countries selected most or all validation cases randomly (Germany, Japan, Poland and England (UK) only selected *some* cases randomly; France didn't select any cases randomly) and 22 countries (88%) validated all dispositions (Australia and Japan did not validate cases finalized as ineligible; France only validated cases finalized as completes).[51]

Details about each country's validation procedure are presented in Table 10.21.

[50] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

[51] Based on information provided on QC forms and during monthly QC conference calls, the Russian Federation followed validation requirements. However, analysis of the data revealed evidence of falsification affecting a significant proportion of cases. This level of falsification should have been detected by validation. The fact that it was not suggests that validation was not conducted in a manner sufficiently adequate to uncover falsification.

Table 10.21: Summary of Validation Procedure for Main Study

	Percentage of interviewers…		Validation Mode				Dispositions Validated					
	validated at 10%	validated at 7%	Phone	In Person	Mail	Other	Completes	Non-Contacts	Refusal	Ineligible	Random Selection	Consortium Form Used
Australia	0	0	793	0	0	0	x	x	x		All	No
Austria	90	94	1122	0	0	0	x	x	x	x	All	Adapted
Canada	65	85	5357	0	0	0	x	x	x	x	All	Adapted
Cyprus[52]	100	100	637	2	379	0	x	x	x	x	All	As is
Czech Republic	100	100	2189	877	0	0	x	x	x	x	Most	Adapted
Denmark	100	100	990	0	7	0	x	x	x	x	Most	As is
England (UK)	12	20	524	2872	33	0	x	x	x	x	Some	Adapted
Estonia	98	100	1138	588	620	0	x	x	x	x	All	As is
Finland	16	46	559	0	0	0	x	x	x	x	All	Adapted
Flanders (Belgium)	75	84	1006	0	0	0	x	x	x	x	All	Adapted
France	100	100	0	0	6,684	0	x				No	No
Germany	100	100	175	176	3400	39	x	x	x	x	Some	Adapted
Ireland	100	100	918	275	12	0	x	x	x	x	Most	As is
Italy	96	99	1450	0	0	0	x	x	x	x	Most	Adapted
Japan	100	100	996	171	589	0	x	x	x		Some	Adapted
Korea	100	100	745	134	0	0	x	x	x	x	All	As is
Netherlands (The)	76	86	584	0	665	0	x	x	x	x	Most	Adapted
Northern Ireland (UK)	91	95	219	1124	2133	0	x	x	x	x	Most	Adapted
Norway	100	100	830	0	0	0	x	x	x	x	All	Adapted
Poland*	36	40	0	1499	0	0	x	x	x	x	Some	Adapted
Russian Federation*[53]	100	100	2500	0	0	0	x	x	x	x	All	As is
Slovakia	97	97	1708	140	0	0	x	x	x	x	Most	As is
Spain	100	100	1045	320	0	0	x	x	x	x	Most	Adapted
Sweden	80	91	860	0	230	0	x	x	x	x	All	Adapted
United States	100	100	1611	228	54	0	x	x	x	x	Most	As is

Source: Data Collection Form submitted after data cleaning unless otherwise noted.

* Based on Data Collection Form submitted after conclusion of data collection.

[52] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[53] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

Thirteen countries (52%) uncovered instances of falsification involving one to 22 interviewers. The nature of falsifications was as follows:

- underreporting the number of household members in the screener
- completing the BQ over the telephone
- using a proxy respondent for the BQ
- misusing the disposition codes
- leaving the exercise booklets with the respondent overnight
- having someone other than the selected respondent complete the exercise
- making up answers to the BQ and the exercise

This emphasizes the critical importance of validation for in-person studies in which interviewers are working independently in the field. A rigorous validation procedure is critical to substantiating data quality.

10.7.3 Other quality control checks

Countries were advised to use automated management reports (proposed by the Consortium) dealing with process data as well as any other means of detecting falsification available to them. The majority of countries (88%) used some of the reports proposed by the Consortium to monitor administration length, time lapse between interviews, and the number of interviews completed per day. Three countries (France, Japan and Russian Federation[54]) did not. Details are provided in Table 10.22.

Table 10.22: Use of Fieldwork Quality Control Reports During the Main Study Data Collection

	Interview Duration	Individual Instrument Duration	Time Between Interviews	Interviews Conducted Very Late/Very Early	Number of Interviews Per Day	Other
Australia	x	x				x[2]
Austria	x	x	x	x	x	
Canada	x	x	x	x	x	
Cyprus[55]	x	x	x	x	x	
Czech Republic	x	x	x	x	x	
Denmark	x	x	x	x	x	
England (UK)	x	x	x	x	x	
Estonia	x	x	x	x	x	
Finland	x		x	x	x	
Flanders (Belgium)	x	x	x	x	x	
France						
Germany	x	x	x	x	x	x[3]
Ireland	x	x	x	x	x	x[4]
Italy	x		x	x		
Japan						
Korea	x	x	x	x	x	
Netherlands	x			x	x	
Northern Ireland (UK)	x	x	x	x	x	
Norway	x	x	x	x	x	x[5]
Poland[1*]	x	x	x	x	x	x[6]
Russian Federation[1* 56]						
Slovakia	x				x	
Spain	x	x	x	x	x	
Sweden	x	x			x	
United States	x	x	x	x	x	

Source: Data Collection Form submitted after data cleaning unless otherwise noted.

[1*] Based on Data Collection Form submitted after conclusion of data collection.
[2] Number of calls and spread of days and times.
[3] Consistency checks of interview and register data (age, gender, nationality).
[4] Review of interviews conducted over 2 days.
[5] Population register checks.
[6] Inconsistency between some BQ items; respondent's actual and declared birthdate.

[54] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

[55] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[56] Please refer to the above note regarding the Russian Federation.

In addition, countries were advised to monitor the quality of data throughout the Main Study data collection. Most countries (88%) reviewed data frequencies and missing data rates. All countries but France reviewed the quality of open-ended responses. Details are presented in Table 10.23.

Table 10.23: Procedures to Monitor Quality of Data During the Main Study Data Collection

	Data Frequencies	Review of 'Other-Specify' Responses	Review of Open-Ended Responses	Missing Data Rates	Other
Australia	x	x	x		
Austria	x		x	x	x[2]
Canada	x	x	x	x	x[3]
Cyprus[57]	x	x	x	x	
Czech Republic	x	x	x	x	
Denmark	x	x	x	x	
England (UK)		x	x	x	
Estonia	x	x	x	x	
Finland	x	x	x	x	
Flanders (Belgium)	x	x	x	x	
France	x			x	
Germany	x		x	x	
Ireland	x	x	x	x	
Italy			x		
Japan		x	x	x	
Korea	x	x	x	x	
Netherlands		x	x	x	
Northern Ireland (UK)	x	x	x	x	
Norway	x	x	x	x	
Poland[1*]	x	x	x	x	
Russian Federation[1*] [58]	x	x	x	x	
Slovakia	x		x		
Spain	x	x	x	x	
Sweden	x	x	x	x	x[4]
United States	x	x	x	x	

Source: Data Collection Form submitted after data cleaning unless otherwise noted.

[1] Based on Data Collection Form submitted after conclusion of data collection.
[2] Consistency checks.
[3] Cross-tabulations, merging of files for consistency checks, fixing data discrepancies.
[4] Macro checks of data; distributions of select background variables have been checked against distribution of corresponding variable from population register and Labor Force Survey.

[57] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[58] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

10.7.4 Interviewer productivity

Unusually high interviewer productivity (measured in number of completed interviews) can be an indication of falsification. Countries were asked to monitor the minimum, mean and maximum number of completes per interviewer and to increase the validation rate of interviews with high production.

The mean number of completes per interviewer ranged from 15 in England (UK) to 102 in Ireland. These countries were characterized by an unusually large and unusually small interviewer workforce, respectively, which is reflected in these numbers. In most countries, the mean number of completes per interviewer was in the 30-40 range. The maximum ranges were extremely varied from 51 in the Russian Federation[59] to 317 in Spain. Details about interviewer productivity are presented in Table 10.24.[60]

Table 10.24: Summary Statistics of the Number of Completes Achieved by Interviewers for the Main Study Data Collection

	Mean (rounded to nearest unit)	Minimum	Maximum	Range
Australia	34	1	123	122
Austria	35	10	116	106
Canada	33	1	132	131
Cyprus[61]	47	0	199	199
Czech Republic	34	1	177	176
Denmark	36	1	130	129
England (UK)	15	1	52	51
Estonia	60	3	195	192
Finland	45	14	91	77
Flanders (Belgium)	56	1	272	271
France	--[**]	--[**]	--[**]	--[**]
Germany	41	8	82	74
Ireland	102	11	156	145
Italy	26	1	97	96
Japan	23	3	73	70
Korea	34	15	58	43
N. Ireland (UK)	20	1	64	63
Netherlands	20	1	137	136
Norway	45	3	143	140
Poland[*]	39	1	138	137
Russian Fed.[*] [62]	35	5	51	46
Slovakia	56	1	159	158

[59] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

[60] Interviewer productivity may have been influenced by the number of hours worked (see Table 10.8 for the typical number of hours worked by PIAAC interviewers in each country).

[61] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[62] Please refer to the above note regarding the Russian Federation.

Table 10.24 (cont.): Summary Statistics of the Number of Completes Achieved by Interviewers for the Main Study Data Collection

	Mean (rounded to nearest unit)	Minimum	Maximum	Range
Spain	42	1	317	316
Sweden	35	4	89	85
United States	25	1	123	122

Source: Data Collection Form submitted after data cleaning unless otherwise noted.

[1] Based on Data Collection Form submitted after conclusion of data collection.
[2] Not reported.

10.8 Interviewer debriefing

Countries were required to administer a debriefing questionnaire and/or hold debriefing sessions with interviewers to gain insights into the problems they encountered, comments made by respondents, and suggestions for improving procedures for future cycles of PIAAC. Countries were required to provide a report to the Consortium summarizing the key findings. An international summary covering interviewer training, BQ and Exercise administration, and the virtual machine was provided to all Consortium members and OECD.

10.9 Recommendations for future cycles

The Field Test and Main Study provided opportunities for countries, the Consortium and the OECD to understand country compliance with the implementation of PIAAC according to a set of agreed-upon standards and to meet production goals. Based on the Field Test and Main Study experience of PIAAC Round 1, the Consortium is proposing a series of recommendations for future cycles of PIAAC.

1. **Study planning is crucial to success, and timely submission of the National Survey Design and Planning Report (NSDPR) must be a nonnegotiable requirement for participation.** Completing a thorough and timely NSDPR ensures: (1) that countries have thought through the study requirements/challenges and are prepared to assure the Consortium that they are fully committed to PIAAC; and (2) allows the Consortium to provide timely feedback on areas of concern.

   In particular, countries hiring data collection organizations separate from the survey institute must be confident that their contractor intends to meet the TSG and can provide all the information necessary to submit a complete and timely NSDPR.

2. **Countries must conduct a rigorous survey institute selection and monitoring process.** Countries should start the search and selection process for the organization that will undertake data collection as early as possible. Final selection should occur no later than six months prior to the start of data collection.

Countries should provide candidate organizations with clear experience and performance guidelines based on the TSG. Final selection should be based on demonstrated experience and the ability to perform the work following the PIAAC TSG.

Countries must monitor the data collection entity closely during the period of performance, requiring at least monthly meetings with key organization staff as well as monthly reports. During the data collection period, countries should also require weekly production status reports.

3. **All cycles of PIAAC must include a Field Test.** Cycle 1 countries learned a great deal from the Field Test experience, which allowed them to adjust their data collection process in preparation for the Main Study. Due to expected changes in future cycles, especially in terms of content and country staffing, all future cycles of PIAAC should require a Field Test, even for countries having implemented a successful Cycle 1.

4. **The Main Study data collection period should be extended one to two months.** This will accommodate countries dealing with populations that have fairly rigid holiday observances (e.g., August vacations, winter breaks, religious periods) and experience weather/climate challenges to plan accordingly and meet the data collection timeframe.

5. **Countries should adhere to the training program produced by the Consortium and train field staff following the TSG hours specified.** Successful data collection requires interviewers that are well trained on the importance of the survey, instrument administration and procedures, and obtaining high response rates. The guidelines below help ensure that country interviewers receive sufficient training. These guidelines for training hours have been revised based on the Cycle 1 experience (i.e., reduced by six hours) and are displayed in Table 10.25.

Table 10.25: Revised Interviewer Training Requirements in Hours

Topics	Inexperienced Interviewers	Experienced Interviewers
General Interviewing Techniques	4	--
Protocol		
- procedures	8	8
- screener administration	4[1]	4[1]
- BQ+EX administration	6	6
- role-plays/practice interviews	7	7
Gaining Cooperation	3	3
Total	28-32 hours	24-28 hours

[1] Not applicable for countries with registry samples.

6. **Countries should adopt a rigorous field management style as specified by the TSG.** Close management of fieldwork is crucial to the success of data collection. Countries must require continual monitoring of field staff and an adequate supervisor to interviewer ratio. NPMs should require country data collection managers to communicate weekly with them and their field staff to ensure adequate monitoring of production and response

rates. Data collection staff at all levels, from supervisors to interviewers, must be held accountable for their performance. This can be best achieved through frequent communication and monitoring.

7. **All validation TSG must be followed.** Countries must be required to agree to adhere to these standards, with no exceptions. This is the most important quality control activity undertaken in household studies. Thus, validation cases must be randomly selected from a sample of all finalized cases and must be conducted at the 10% level on all interviewers working on PIAAC.

Chapter 11: Quality Control Monitoring Activities

Pat Montalvan and Michael Lemay, Westat

11.1 Overview

This chapter presents the details of the survey operations' quality assurance (QA) and quality control (QC) activities conducted by the Consortium as part of the Main Study.

This program was designed to: (1) support the collection of data and results that are valid, reliable and comparable across countries and over time, and satisfy accepted quality assurance goals; (2) keep the OECD and the Consortium informed about the progress of data collection; and (3) provide advice to countries needing assistance. The aim was to implement a program that represents the best tradeoff of quality and cost within the constraints of the project. The plan was presented to the OECD and the BPC and approved by the PIAAC Technical Advisory Group (TAG) in July 2008.

The principal objectives of the QA and QC program for survey operations/data collection were the following:

- Undertake activities which monitor the implementation of the PIAAC Technical Standards and Guidelines (TSG) for survey operations during the Field Test and Main Study.
- Review the progress of data collection and identify potential problems and solutions on a timely basis during the Field Test and Main Study.
- Make recommendations to enhance the Main Study based on the Field Test experience.
- Identify sources of nonsampling error to inform analysis.
- Make recommendations for the next wave/cycle of PIAAC.

The PIAAC QC process provided continuous support to countries in following the TSG before, during and after data collection. It assisted countries by answering questions and communicating areas of concern in a timely manner. Furthermore, it informed OECD and the Consortium of the status of data collection in each country on a regular basis throughout the process. The level of cooperation from countries was superior to the Field Test and was very good overall.

The process described in this chapter allowed collection of most of the information presented in Chapter 10.

A description of the QA and QC activities for survey operations follows in section 11.2. Section 11.3 looks at country compliance with these activities.

11.2 Quality assurance and quality control activities

11.2.1 Quality assurance (QA)

The QA process for survey operations consisted of the development of standards and guidelines, including the QC process, a QC and management manual, and the training of national teams on QC activities.

Development of standards and guidelines for survey operations

The first step in the implementation of the PIAAC quality assurance program was the specification of standards and guidelines covering all aspects of the survey life cycle, from sample and instrumentation design to data processing and analysis. A significant portion of the TSG (Chapters 8 through 10) deals specifically with survey operations concerns such as field staff recruitment, management and training, and field management practices. The PIAAC standards are based on generally agreed upon policies or best practices to be adhered to in the conduct of the survey.

Development of survey operations QC and management manual

The purpose of this manual was to: (1) provide national teams with details on important survey operations standards with practical suggestions on how to implement them (e.g., field management reports, fieldwork quality control, tools to increase respondent cooperation); and (2) provide national teams with details on the logistics of the PIAAC quality control program (e.g., forms to be submitted, quality control call schedule).

International training on survey operations QC

The international training on survey operations QC control took place prior to the Field Test international interviewer training in February 2010 and covered the essential points in the QC manual. Key points were covered again at the June 2011 NPM meeting prior to the Main Study data collection.

11.2.2 Quality control (QC)

The QC process consisted in regular communication in the form of reports, conference calls and ad hoc email exchanges. This section provides a summary description of each activity.

National Survey Design and Planning Report (NSDPR) review

Each country was required to submit an NSDPR covering all aspects of the survey implementation at least six months prior to the beginning of data collection. The Consortium reviewed the survey operations chapters (four chapters, covering 70 standards) of the NSDPR for each country and reported on any serious deviations from the TSG.

Data collection QC conference calls

The Consortium conducted conference calls (see Chapter 5 in QC and Management Manual for more details) with each of the PIAAC countries on a regular basis throughout the critical Field Test and Main Study data collection periods. The goals of the calls were to: (1) review the content of the monthly data collection QC forms submitted by countries (see below); (2) give countries the opportunity to ask questions in real time; and (3) discuss any survey operations issues that may have arisen in each country.

Calls were held prior to the start of data collection, during data collection, and one month after data collection. Calls were held monthly with each country during Field Test data collection and reduced to every other month during the Main Study. However, calls were held more often when needed.

Conference call participants varied somewhat from month to month, depending on study timeframe and issues at hand, but generally they included the country's NPM, key Leading Survey Institute (LSI) staff (who speak English), and key Westat operations staff.

Conference calls followed a specific agenda guided by the data collection QC form and were documented using QC meeting minutes reports which summarized the items discussed, the decisions made and the pending action items.

Data collection QC form

Countries were required to complete monthly QC monitoring forms. These forms were used to guide the conference call meetings and focused on the topics covered in Chapters 2, 8, 9 and 10 of the PIAAC TSG (82 short answer questions). Topics included:

- field staffing and management
- plans for contacting households/respondents
- respondent outreach materials
- ways of dealing with nonresponse
- field management system
- response rates and production
- field QC measures
- plans to train staff on ethics and confidentiality

The completed electronic forms were posted each month on the project's SharePoint site, which is accessible by all participating countries and organizations.

Interviewer training QC form

To ascertain adherence to the interviewer training program designed by the Consortium, countries were required to complete an interviewer training QC form at the end of each interviewer training session in each country (28 questions). The form included questions about the:

- number of trainers and trainees
- experience profile of trainees
- training materials used
- topics covered at training

The completed electronic forms were also posted on the project's SharePoint site.

Interviewer debriefing questionnaire and report

Countries were required to administer a debriefing questionnaire to interviewers following the conclusion of data collection to ensure that interviewer feedback was obtained. The form included 47 questions covering:

- training
- the administration of the Background Questionnaire
- the administration of the computer-based exercise
- the administration of the paper exercise
- the interview in general
- the interviewer help line

Each country was required to summarize interviewer feedback for each question on the questionnaire and submit the report to the Consortium for review.

Ongoing Web communication

Through Web communication, countries could ask for and receive responses from Westat to ad hoc questions arising throughout the planning and implementation phases of PIAAC data collection.

11.3 Country compliance

As shown in Table 11.1, virtually all countries fulfilled the QC requirements for Main Study data collection. Some countries met the requirements with some delay but were proactive in notifying the Consortium in advance. A few calls had to be rescheduled, but this was usually done with advance notice.

Table 11.1: Compliance with the Main Study Survey Operations Quality Control (QC) Program

Required QC Activities	Percentage of Countries Complying (n=25)
Revised Main Study NDSPR (1 report)	96
QC calls	--
- at least once prior to data collection	96
- at least every other month during data collection	100
- once after data collection	100
Data Collection Form	--
- at least once prior to data collection	96
- monthly during data collection	84
- once after data collection	100
- once after data cleaning	88
Interviewer Training Form (1 form per training session)	100
Interviewer Debriefing Report (1 report)	88

Next, we report in detail how countries fulfilled the requirements.

Survey operations sections of the revised Main Study NSDPR

Twenty-four of the 25 participating countries (96%) submitted a final NSDPR for the Main Study, although few did so on time (by 1 February 2011). One country submitted only a draft Main Study NSDPR (see Table 11.2).

Table 11.2: Final Main Study NSDPR Submission Dates

Country	Submission Date
Australia	4 February 2011
Austria	15 March 2011
Canada	25 February 2011
Cyprus[1]	1 February 2011
Czech Republic	25 January 2011
Denmark	31 January 2011
England/N. Ireland (UK)	2 February 2011
Estonia	1 February 2011
Finland	31 January 2011
Flanders (Belgium)	26 January 2012
France	3 October 2012
Germany	1 February 2011
Ireland	9 August 2011
Italy	2 August 2011
Japan	31 January 2011 *(revised 9 March 2012)*
Korea	Draft Main Study only
Netherlands (The)	21 March 2012
Norway	1 February 2011 *(revised 12 August 2011)*
Poland	24 January 2011 *(revised 2 August 2011)*
Russian Federation[2]	1 February 2011
Slovakia	31 October 2011
Spain	11 February 2011 *(revised 27 April 2012)*
Sweden	10 February 2011
United States	1 February 2011

Source: PIAAC SharePoint site timestamps.

Data Collection Form submission and conference calls prior to data collection

Twenty-four countries (96%) submitted the required Data Collection Form and participated in a QC call at least once prior to the beginning of data collection, which is satisfactory. The requirement called for the submission of a Data Collection Form for each month leading up to the beginning of data collection. A few countries could not fulfill this requirement due to staff shortages during summer vacation. As in the Field Test, it appears that a few countries (n=9) may have misunderstood the requirement to submit a new form even if there were no changes.

[1] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[2] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

One country (Slovakia) did not submit a Data Collection Form and did not participate in a QC call prior to the beginning of data collection (see Table 11.3).

Table 11.3: Actual Schedule of Data Collection Form Submission and Associated QC Monitoring Calls Prior to Main Study Data Collection

	About Two Months Prior to Data Collection		About One Month Prior to Data Collection	
	Form	**Call**	**Form**	**Call**
Australia	5 August	16 August	12 Sept	20 Sept
Austria	7 June	7 June	Not submitted	Not required
Canada	8 Sept	20 Sept	Not required[2]	Not required
Cyprus[3]	18 July	26 July	16 August	Not required
Czech Republic	Not submitted	22 July	15 August	16 August
Denmark	1 July	6 July	27 July	3 August
England/Northern Ireland (UK)	26 May	6 June	29 June	Not required
Estonia	16 June	16 June	7 July	12 July
Finland	Not required[1]	Not required[1]	8 August	16 August
Flanders (Belgium)	23 June	27 June	23 August	Not required
France	Not required	Not required	13 July	27 July
Germany	Not required[1]	Not required[1]	4 July	13 July
Ireland	2 June	7 June	Not submitted	Not required
Italy	6 July	26 July	29 July	8 August
Japan	3 June	14 June	7 July	12 July
Korea	16 August	23 August	Not submitted	Not required
Netherlands (The)	20 July	21 July	Not submitted	24 August
Norway	20 June	22 June	Not submitted	Not required
Poland	22 June	22 June	25 July	25 July
Russian Federation[4]	5 August	10 August	Not submitted	Did not take place
Slovakia	Not submitted	Did not take place	Not submitted	Did not take place
Spain	8 July	15 July	5 August	23 August
Sweden	4 July	5 July	Not submitted	Not required
United States	20 July	Not required[3]	28 July	Not required[3]

Source: SharePoint and email timestamps, QC meeting minutes reports.

[1] A special agreement was reached in which it was agreed that the Data Collection Form submission and the QC call would take place in August only due to the difficulty of having staff available during summer vacations.
[2] It was agreed that a new submission was not necessary as the country certified that no change would be made to procedures.
[3] The Consortium's survey operations quality control manager attended weekly management meetings of the US PIAAC team.

Interviewer Training Forms

Twenty-five countries (100%) reported on their interviewer training sessions. The requirement was for countries to report on each training session held by submitting a separate report for each.

[3] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[4] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

Worldwide, 380 interviewer training sessions were held. Countries conducted between two and 72 training sessions each, a number that includes both the initial training sessions and any session held to compensate for interviewer attrition.

Data Collection Form submission and conference calls during data collection

The majority of countries (84%) submitted one Data Collection Form for each month of fieldwork as required. All countries complied with the requirement to participate in a QC call at least every other month. The standard quality control program during the data collection period called for monthly submissions of the Data Collection Form (eight submissions) and QC conference calls at least every other month (at least four calls).[5] However, depending on their respective data collection start date and the end date of QC activities set by the Consortium (3 April 2012, extended to 2 May 2012), a number of countries had fewer than the typical number of submissions/calls. They are nonetheless considered to have fully complied with the quality control program (see Table 11.4).

[5] Requirements were adapted to France's shorter data collection period.

Table 11.4: Actual Schedule of Data Collection Form Submission and Associated QC Monitoring Calls for Main Study

	Month 1		Month 2		Month 3		Month 4		Month 5		Month 6		Month 7		Month 8	
	Form	**Call**	**Form**	**Call**	**Form**	**Call**	**Form**	**Call**	**Form**	**Call**	**Form**	**Call**	**Form**	**Call**	**Form**	**Call**
Australia	19 Oct	NR	8 Nov	14 Nov	13 Dec	20 Dec	11 Jan	NR	14 Feb	21 Feb	13 Mar	NR	--	--	--	--
Austria	30 Aug	8 Sept	20 Oct	NR	14 Nov	17 Nov	2 Dec	NR	10 Jan	12 Jan	16 Feb	NR	8 Mar	8 Mar	--	--
Canada	25 Nov	29 Nov	NS	NR	20 Jan	30 Jan	NS	NR	26 Mar	29 Mar	30 Apr	NR	--	--	--	--
Cyprus[6]	21 Sept	27 Sept	18 Oct	NR	14 Nov	22 Nov	13 Dec	NR	18 Jan	24 Jan	21 Feb	NR	20 Mar	27 Mar	--	--
Czech Republic	20 Sept	23 Sept	26 Oct	31 Oct	28 Nov	29 Nov	NR[3]	NR	25 Jan	27 Jan	21 Feb	NR	20 Mar	27 Mar	--	--
Denmark	7 Sept	NR	29 Sept	5 Oct	1 Nov	NR	30 Nov	7 Dec	10 Jan	NR	23 Jan	1 Feb	1 Mar	7 Mar	--	--
England/N. Ireland (UK)	31 Aug	7 Sept	28 Sept	NR	27 Oct	2 Nov	1 Dec	NR	21 Dec	4 Jan	25 Jan	NR	29 Feb	7 Mar	--	--
Estonia	4 Aug	NR	20 Sept	22 Sept	4 Nov	NR	4 Nov	8 Nov	6 Dec	13 Dec	9 Jan	NR	8 Feb	14 Feb	12 Mar	NR
Finland	14 Sept	NR	12 Oct	18 Oct	10 Nov	NR	13 Dec	20 Dec	11 Jan	NR	15 Feb	21 Feb	12 Mar	NR	--	--
Flanders (Belgium)	22 Sept	28 Sept	21 Oct	NR	23 Nov	28 Nov	15 Dec	NR	16 Jan	23 Jan	16 Feb	NR	14 Mar	2 Apr	--	--
France	NS	NR	3 Oct	4 Oct	NS	NR	--	--	--	--	--	--	--	--	--	--
Germany	2 Aug	NR	7 Sept	14 Sept	5 Oct	NR	2 Nov	9 Nov	7 Dec	NR	4 Jan	11 Jan	1 Feb	NR	8 Mar	14 Mar
Ireland	29 Aug	14 Sept	26 Oct	NR	7 Nov	9 Nov	13 Dec	NR	11 Jan	11 Jan	17 Feb	NR	14 Mar	14 Mar	--	--
Italy	10 Oct	17 Oct	14 Nov	NR	6 Dec	12 Dec	NS	NR	11 Feb	15 Feb	19 Mar	NR	--	--	--	--
Japan	14 Sept	14 Sept	5 Oct	NR	1 Nov	8 Nov	15 Dec	NR	6 Jan	10 Jan	7 Feb	NR	--	--	--	--
Korea	22 Oct	25 Oct	27 Nov	NR	17 Dec	21 Dec	NR[2]	NR[2]	NR[2]	NR[2]	30 Mar	3 Apr	--	--	--	--
Netherlands (The)	28 Sept	NR	14 Oct	20 Oct	10 Nov	NR	8 Dec	15 Dec	20 Jan	NR	10 Feb	16 Feb	12 Mar	NR	--	--
Norway	6 Sept	9 Sept	18 Oct	25 Oct	1 Dec	NR	21 Dec	21 Dec	25 Jan	NR	9 Feb	22 Feb	29 Mar	NR	--	--
Poland	8 Aug	NR	5 Sept	19 Sept	31 Oct	NR	21 Nov	21 Nov	8 Dec	NR	17 Jan	17 Jan	21 Feb	NR	20 Mar	20 Mar
Russian Federation[7]	28 Nov	22 Dec	23 Jan	25 Jan	25 Feb	28 Feb	28 Mar	NR	15 Apr	18 Apr	--	--	--	--	--	--
Slovakia	18 Nov	25 Nov	11 Dec	16 Dec	16 Jan	NR	21 Feb	23 Feb	26 Mar	NR	--	--	--	--	--	--
Spain	20 Sept	NR	14 Oct	21 Oct	14 Nov	NR	9 Dec	16 Dec	16 Jan	NR	10 Feb	17 Feb	9 Mar	NR	--	--
Sweden	28 Sept	4 Oct	NS	NR	15 Nov	6 Dec	17 Jan	NR	7 Feb	14 Feb	28 Mar	NR	--	--	--	--
United States	30 Aug	NR[1]	28 Sept	NR[1]	28 Oct	NR[1]	22 Nov	NR[1]	3 Jan	NR[1]	24 Jan	NR[1]	21 Feb	NR[1]	20 Mar	NR[1]

Source: SharePoint and e-mail timestamps and QC meeting minutes reports.

[1] The Consortium's survey operations quality control manager attended weekly management meetings of the US PIAAC team.
[2] Data collection was suspended.
[3] Not required by special agreement due to holiday break.

[6] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[7] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

Data Collection Form submission and conference calls after data collection

Twenty-five countries (100%) submitted a Data Collection Form after completion of data collection. However, only 22 countries (88%) submitted a Data Collection Form after data cleaning was completed despite several reminders (see Table 11.5).

Table 11.5: Actual Schedule of Data Collection Form Submission and Associated QC Monitoring Calls

	After Data Collection		After Data Cleaning	
	Form	**Call**	**Form**	**Call**
Australia	13 Apr	17 Apr	21 June	Not required
Austria	30 Mar	12 Apr	11 June	Not required
Canada	27 July	Did not take place[1]	17 Sept	Not required
Cyprus[8]	17 Apr	24 Apr	25 May	Not required
Czech Republic	23 Apr	27 Apr	27 June	Not required
Denmark	2 Apr	4 Apr	6 July	Not required
England/N. Ireland (UK)	30 Mar	4 Apr	6 July	Not required
Estonia	4 Apr	10 Apr	19 June	Not required
Finland	11 Apr	17 Apr	20 June	Not required
Flanders (Belgium)	16 Apr	23 Apr	7 Aug	Not required
France	21 Dec	11 Jan	Not submitted	Not required
Germany	5 Apr	16 Apr	27 June	Not required
Ireland	17 Apr	18 Apr	31 Jan 2013	Not required
Italy	26 Apr	2 May	3 July	Not required
Japan	13 Mar	13 Mar	15 June	Not required
Korea	24 Apr	25 Apr	9 Aug	Not required
Netherlands (The)	22 Apr	26 Apr	18 June	Not required
Norway	25 Apr	25 Apr	20 June	Not required
Poland	12 Apr	16 Apr	Not submitted	Not required
Russian Federation[9]	6 June	Did not take place[1]	Not submitted	Not required
Slovakia	25 Apr	26 Apr	15 June	Not required
Spain	13 Apr	20 Apr	18 June	Not required
Sweden	28 Mar	3 Apr	23 Aug	Not required
United States	27 Apr	Not required[2]	15 June	Not required[2]

Source: SharePoint and e-mail timestamps, QC meeting minutes reports.

[1] Main Study quality control calls ended on 31 May 2012 for all countries but France.
[2] The Consortium's survey operations quality control manager attended weekly management meetings of the US PIAAC team.

[8] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[9] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

Interviewer Debriefing Report

Countries were required to debrief their interviewers on their Main Study experience and provide a report to the Consortium. Twenty-two countries (88%) submitted such a report (see Table 11.6).

Table 11.6: Main Study Interviewer Debriefing Report Submission Dates

Country	Date
Australia	4 June
Austria	20 July
Canada	6 September
Cyprus[10]	15 May
Czech Republic	20 June
Denmark	3 August
England/N. Ireland (UK)	8 October
Estonia	19 June
Finland	29 May
Flanders (Belgium)	30 May
France	Not submitted
Germany	4 July
Ireland	20 June
Italy	18 June
Japan	10 September
Korea	9 August
Netherlands	30 July
Norway	28 August
Poland	Not submitted
Russian Federation[11]	17 July
Slovakia	7 June
Spain	27 June
Sweden	Not submitted
United States	18 May

Source: SharePoint and e-mail timestamps.

[10] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[11] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

11.4 Conclusion

Overall, the PIAAC quality control program for survey operations met its goals. During the Main Study: (1) country compliance was high; (2) the OECD and the Consortium were kept informed about the progress of data collection; (3) countries were supported – their questions were answered and areas of concern were pointed out promptly throughout the critical months before and during data collection; (4) the program allowed for the sharing of status information with all countries and helped foster a sense of cooperation and "shared experience"; and (5) the program experience should assist counties and the OECD as they plan for future cycles of PIAAC.

Chapter 12: Scoring Reliability Studies

Claudia Tamassia, Mary Louise Lennon and Kentaro Yamamoto, ETS

While PIAAC was primarily a computer-delivered and computer-scored instrument, a paper-and-pencil version of the cognitive instruments was also an important component of the assessment. The Field Test design required all participating countries to administer paper-based versions of the literacy and numeracy items as part of the study to compare the performance of items that served to link the paper-and-pencil and computer-delivery formats.[1] In the Main Study, paper booklets were administered to study participants who were either unwilling to take the test on the computer or unable to do so because they lacked basic computer skills. Therefore scoring designs and operational procedures were developed for these human-scored items.

This chapter describes the scoring process and associated scoring reliability studies for the paper-and-pencil instruments. Without accurate and consistent scoring of paper-and-pencil items, all subsequent psychometric analyses of those items are severely jeopardized. Therefore PIAAC, like other large-scale assessments before it, defined a set of essential processes that all participating countries were required to implement to maintain scoring consistency within and across countries. These included having items scored independently by two different scorers and providing a common set of anchor booklets to be scored by all national teams. An important aspect related to scoring in PIAAC was the requirement that countries follow specified scoring designs to ensure that each booklet was scored twice and that scorers functioned in both the first- and second-scorer roles across all the booklets. These scoring designs, along with a specified set of procedures for training scorers and monitoring the scoring process, were designed to ensure that PIAAC would provide accurate and reliable data for policymakers, researchers, and other stakeholder groups interested in adult skills and their distribution in an international context.

12.1 The scoring process

The PIAAC paper instruments included four booklets:

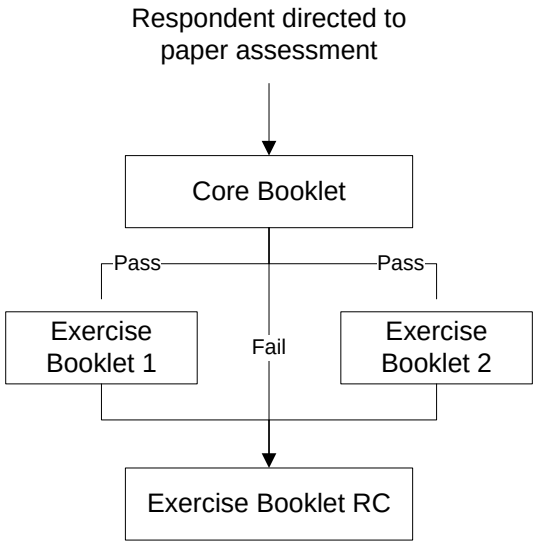
- the Core Booklet, which contained eight items (four literacy and four numeracy),
- Exercise Booklet 1, which contained 20 literacy items,
- Exercise Booklet 2, which contained 20 numeracy items, and

[1] Because the ICT component was an integral part of problem solving in technology-rich environments, there was no paper-based version of that domain.

- Exercise Booklet RC, which contained 109 reading components items.

In the Main Study, the paper-based assessment was administered to respondents who either reported they had no computer experience; failed the test of basic computer skills required to take the assessment; or refused to take the assessment on the computer. Within this design, the Core Booklet was presented first and included the easiest questions. If the respondent passed the Core Booklet, the interviewer administered either Exercise Booklet 1 or Exercise Booklet 2. Each respondent had a 50% chance of receiving one or the other booklet. In countries that opted to assess reading components, after the respondent completed Exercise Booklet 1 or 2, or in cases where a respondent failed the core, the interviewer administered Exercise Booklet RC. This Main Study design is illustrated in Figure 12.1.

Figure 12.1. Main Study paper booklet assessment design



The Core Booklet, Exercise Booklet 1 and Exercise Booklet 2 were scored by trained teams within each participating country. These same booklets were used to conduct within- and across-country reliability studies as described in section 12.2.

Responses for reading components (Exercise Booklet RC) were not scored. Instead, members of the scoring team recorded answers on response sheets that were then used for data entry and automated scoring. Therefore, the PIAAC scoring designs include only the Core Booklet and Exercise Booklets 1 and 2.

12.1.1 Preparing for scoring

A number of key activities were completed by the Consortium and National Centers prior to the assessment to prepare for scoring the paper-based instruments. The Consortium developed detailed scoring guides that included scoring rubrics as well as examples of correct and incorrect responses. For linking items, scoring information from previous assessments (IALS and ALL) was included in the scoring guides. For new items, scoring rubrics were defined for the Field Test, and information from Field Test scoring was then used to expand the scoring guides for the Main Study.

A two-day meeting with NPMs and chief scorers was conducted where scoring guides were presented and explained. Participants practiced scoring sample items, and the group discussed any ambiguous or problematic situations. By focusing on sample responses likely to provide the greatest challenge to scorers, meeting participants had the opportunity to ask questions and clarify the application of scoring rubrics. When the discussion revealed areas where rubrics could be improved, those changes were made and included in an updated version of the scoring guide documents provided after the meeting.

To support countries during the scoring process, the Consortium established a distribution list that allowed national teams to submit scoring questions and receive responses from the relevant domain experts. National teams were also able to review questions submitted by other countries along with the responses from the Consortium. A summary report of scoring issues was provided on a regular basis and all emails were archived on a discussion board on the PIAAC SharePoint site for reference by national scoring teams.

National Centers were responsible for assembling a team of scorers. The first task was to identify a lead scorer who would be part of the scoring team and additionally be responsible for the following tasks:

- Hiring and training scorers within the country
- Monitoring the scoring process. This included daily monitoring of the scores in the data entry software (Data Management Expert, or DME), reviewing scoring progress and outcomes, and taking action when scoring anomalies were identified. At the beginning of the scoring process, the lead scorer was required to manually inspect a portion of the scored booklets for scoring accuracy before scores were entered into the DME. This series of checks ensured that the initial booklets were scored according to the guidelines. When the lead scorer was comfortable and confident that all the scorers were consistently following the scoring guidelines, he or she then monitored outcomes through the DME software.
- Monitoring the inter-rater reliability and taking action when the scoring results were unacceptable and required further investigation
- Retraining or replacing scorers if necessary
- Subscribing to the PIAAC scoring distribution list, submitting any scoring questions for resolution by the PIAAC domain experts, and monitoring the weekly summary reports
- Reporting scoring results and providing status reports to the NPM and Consortium

The lead scorer was required to be proficient in English, as international training and interactions with the Consortium were in English only. It was also assumed that the lead scorer for the Field Test would retain that role for the Main Study. When this was not the case, it was the responsibility of the National Center to ensure that the new lead scorer received training equivalent to that provided at the international scoring training prior to the Field Test.

The guidelines for assembling the rest of the scoring team included the following requirements:

- All scorers were to have more than a high school qualification, with university graduates preferable.

- Scorers were to be trained based on a nationally developed training package that included an overview of the survey and training manuals based on the manuals and materials provided by the Consortium.

- The lead scorer and one other scorer were required to be bilingual, meaning they had to be proficient in English and the national language. Both scorers would serve as part of the scoring team and be responsible for scoring the anchor booklets. If countries followed a design that required only two scorers, both had to be bilingual.

- Scorers were expected to be organized in teams and to work on the same schedule and in the same location to facilitate discussion about scoring issues as they arose. Past experience showed that if scorers were able to discuss questions among themselves and with their lead scorer, many issues could be resolved in a way that resulted in more consistent scoring.

- Each scorer was assigned a unique scorer ID.

- Due to normal attrition rates and unforeseen absences, the Consortium strongly recommended that lead scorers train a backup for their scoring teams.

Additional information about the scoring staff was provided in standard 11.4 in the PIAAC Technical Standards and Guidelines.

12.2 Reliability studies

Comparability both within and across countries was an important design criterion in PIAAC. The equivalence of scoring was established by double scoring the Core Booklet, Exercise Booklet 1 and Exercise Booklet 2 by two independent scorers, as well as carefully monitoring and responding to the scoring results. These steps were required as quality-assurance measures to determine whether scorers were applying the scoring rubrics consistently. The purposes for rescoring were to: i) document the degree to which the same scores were given to items regardless of the scorer; and ii) identify items and scorers with low inter-rater agreement. To ensure that the first and second scores were truly independent, certain precautions were taken. For example, scores had to be assigned by two different scorers, and the second scorer was not allowed to see scores given by the first scorer.

12.2.1 Within-country scoring reliability study

The purpose of the within-country inter-rater scoring reliability study was to ensure scoring reliability within a country and identify scoring inconsistencies or problems early in the scoring process so they could be resolved as soon as possible. In general, inconsistencies or problems were due to scorer misunderstanding of general scoring guidelines and/or a rubric for a particular item.

The level of agreement between two scorers was represented by an inter-rater reliability index based on percent correct. In PIAAC, inter-rater reliability represented the extent to which any two scorers agreed on how a particular response should be scored, and thus how comparably the scoring rubric was being interpreted and applied. Inter-rater reliability varied from 0 (no reliability or 0% agreement) to 1 (maximum degree of reliability or 100% agreement). The goal for PIAAC was to reach a within-country inter-rater reliability of 0.95 (95% agreement) across all items, with at least 85% agreement for each item.

The IEA DME Tools Software was developed for calculating inter-rater reliability. As the name implies, these tools were used with data from the DME database. Once scores were entered into the database, the IEA DME Tools were used to produce output and reports needed for examining scoring reliability. Countries received training on the use of these tools to monitor scoring reliability.

12.2.2 Cross-country scoring reliability study

Accurate and consistent scoring within a country does not necessarily imply that all countries are applying the scoring guides in the same manner. Scoring bias may be introduced if one country scores a certain response differently from other countries. Therefore, in addition to within-country inter-rater reliability, it was also important to check the consistency of scorers across countries.

Guidelines 11.3.3A and 11.3.3B in the PIAAC Technical Standards and Guidelines described the cross-country reliability study using a set of anchor booklets. The anchor booklets were a set of 180 completed core, literacy and numeracy booklets. Item responses in these booklets were based on actual responses collected in the field as well as responses that reflected key points on which scorers were trained. They included the international cover page and were identified by an international booklet serial number (or ID) prefilled on that cover page. The anchor booklets were in English and scored by the two bilingual scorers.

The anchor booklets were required to be scored and rescored by every country as the main and reliability scores for these booklets were used by the Consortium to calculate inter-rater agreement across countries. There was no scoring resolution for these booklets. Thus, countries were to simply single score these booklets and enter the data into the DME. It was important that countries did not resolve any discrepancies with the anchor booklet items because the Consortium needed the original scores to examine the degree of consistency among the participating countries.

12.3 Scoring designs

Three different scoring designs were developed to meet the needs of countries with varying numbers of respondents taking the paper-based instruments. These designs ensured a balanced distribution of bundles, or groups of booklets, across the number of scorers in a country while also balancing the order in which the bundles were scored. The Consortium also worked with countries that needed to deviate from these standard scoring designs, developing a tailored design to meet the country's circumstances while still adhering to technical requirements.

Within each scoring design, of the following conditions had to be met:

- A minimum of 600 booklets sets (i.e., the set of booklets completed by a respondent) was required to be double scored using a balanced design to assess within-country scoring reliability. For some countries this meant that all booklets had to be double scored. Countries that collected more than 600 booklets had the option of single scoring booklets once the threshold of 600 was reached. For countries that collected fewer than 600 booklets, the guidelines required that 100% of the available booklets be double scored.

- Each scorer needed to score at least 125 items that were also scored by another scorer. This condition was necessary in order to generate enough data to evaluate the accuracy of the scorers.

- Two scorers were required to score the anchor booklets as specified in the scoring design to assess cross-country scoring reliability.

12.3.1 'Standard' three-scorer design

The standard three-scorer design was the default recommended design and applied to most participating countries. The design could be used in cases where countries collected *a total of around 600 booklet sets*. In this design, countries double scored all of their paper booklets, except for any extra bundles that were organized after this process was completed for the initial booklets. This design is presented in Table 12.1 below. Note that the numbers 1 and 2 shown in the table represent main (1) and reliability (2) scoring and not the scoring order. The design is summarized as follows:

- 18 bundles were assembled including:
    - C01 to C06 (Core Booklets)
    - L01 to L06 (Literacy Exercise Booklets 1), and
    - N01 to N06 (Numeracy Exercise Booklets 2).

  Within each booklet type, bundles included approximately equal numbers of booklets.

- Three bundles of anchor booklets were included, with 60 booklets in each bundle. As shown by the yellow highlighting, anchor bundle C00 included Core booklets, L00 included Exercise 1 booklets, and N00 Exercise 2 booklets. Each of these booklets was single scored.

- Three bundles (E01, E02 and E03) were reserved for any extra national paper booklets received after the initial booklet organization, bundling and dispersion took place. These booklets were single scored.

Table 12.1: Scoring design with three scorers

Bundle	Scorers		
	A	B	C
C01	1	2	
C02	2		1
C03		1	2
C00	1	2	
C04	2	1	
C05	1		2
C06		2	1
L01	1	2	
L02	2		1
L03		1	2
L00	1	2	
L04	2	1	
L05	1		2
L06		2	1
N01	1	2	
N02	2		1
N03		1	2
N00	1	2	
N04	2	1	
N05	1		2
N06		2	1
E01	1		
E02		1	
E03			1

As required, this design ensured that all scorers had a minimum of 125 scored items that could be matched to scores from other scorers.

The design required Scorers A and B to be bilingual as they scored the English language anchor booklets in bundles C00, L00 and N00.

12.3.2 Three-scorer design with single score bundles

If a country had *more than 600 booklet sets*, it could opt to use one of two scoring designs. It could use the standard three-scorer design described above and double score all of its Core Booklets, Exercise Booklets 1 and Exercise Booklets 2. It could also use the three-scorer design with single-score bundles presented in Table 12.2. In this design, 600 booklet sets were double scored to fulfill the requirements for the within-country reliability study, and the remaining were single scored. The three-scorer design with single score bundles is summarized as follows:

- As with the standard three-scorer design, 18 bundles were assembled including:
  - C01 to C06 (Core Booklets)
  - L01 to L06 (Literacy Exercise Booklets 1), and
  - N01 to N06 (Numeracy Exercise Booklets 2).

  These bundles included the 600 booklet sets to be double scored. Within each booklet type, bundles included approximately equal numbers of booklets.

- Additionally, nine bundles of national paper booklets were single scored. Bundles S01 to S03 were Core Booklets, S04 to S06 were Exercise Booklets 1, and S07 to S09 were Exercise Booklets 2. These bundles included the booklets remaining after the required 600 booklets were assembled for double scoring.

- Three bundles of anchor booklets were included, with 60 booklets in each bundle. As shown by the yellow highlighting, anchor bundle C00 included Core booklets, L00 included Exercise 1 booklets, and N00 included Exercise 2 booklets.

- Three bundles (E01, E02 and E03) were reserved for any extra national paper booklets received after the initial booklet organization, bundling and dispersion took place. Each of these booklets was single scored.

This design also ensured that all scorers had a minimum of 125 scored items that could be matched to scores from other scorers.

Table 12.2: Scoring design with three scorers and single score bundles

Bundle	Scorers		
	A	B	C
C01	1	2	
C02	2		1
C03		1	2
C00	1	2	
S01	1		
S02		1	
S03			1
C04	2	1	
C05	1		2
C06		2	1
L01	1	2	
L02	2		1
L03		1	2
L00	1	2	
S04	1		
S05		1	
S06			1
L04	2	1	
L05	1		2
L06		2	1
N01	1	2	
N02	2		1
N03		1	2
N00	1	2	
S07	1		
S08		1	
S09			1
N04	2	1	
N05	1		2
N06		2	1
E01	1		
E02		1	
E03			1

The design required Scorers A and B to be bilingual as they scored the English language anchor booklets in bundles C00, L00, and N00.

12.3.3 Two-scorer design

Although one of the three-scorer designs was appropriate for most countries, an alternative two-scorer design was also provided. This two-scorer design was used by countries that had 250 or fewer total booklet sets. The design ensured that each scorer would score at least 125 each of Exercise Booklet 1 and Exercise Booklet 2 as specified in guideline 11.3.2B in the *PIAAC Technical Standards and Guidelines*. The design is shown below in Table 12.3. As with the previous designs, note that the numbers 1 and 2 shown in the table represent main (1) and reliability (2) scoring and not the scoring order. The design was identical to the standard design for three scorers except that:

- Only one bundle, E01, was reserved for any extra national booklets received after the initial booklet organization, bundling and dispersion took place.

- Both scorers needed to be bilingual as they scored the English language anchor booklets in bundles C00, L00, and N00.

Table 12.3: Scoring design with two scorers

	A	B
C01	1	2
C02	2	1
C00	1	2
C03	2	1
C04	1	2
C05	1	2
C06	2	1
L01	1	2
L02	2	1
L00	1	2
L03	2	1
L04	1	2
L05	1	2
L06	2	1
N01	1	2
N02	2	1
N00	1	2
N03	2	1
N04	1	2
N05	1	2
N06	2	1
E01	1	2

In the Main Study, countries did not know, and could not control, how many respondents would take the paper instruments, as that was defined by the number of respondents who had no computer experience or failed the test of basic computer skills. Therefore, the Consortium recommended the following procedure:

1. Estimate the number of respondents who may go to the paper branch because they either did not have computer experience or failed the test of basic computer skills required for the assessment. This initial estimate was needed because countries selected the design they expected to use prior to scorer training.

2. Gather all returned national paper booklets, record their IDs in the appropriate tracking system, assemble and count the number of booklet sets.

    a) If the number of booklet sets is fewer than or equal to 250, use the two-scorer design.

    b) If the number of booklet sets is between 250 and 600, use the three-scorer design and double score every booklet set.

    c) If the number of booklet sets is greater than or equal to 600, choose one of these two options:

        - Option 1**:** double score all booklet sets using the three-scorer design.

- Option 2: use the three-scorer design with single score bundles, where a portion of the booklets are double scored for the reliability study and the remaining booklets are single scored.

Options 1 and 2 were contingent on following these two rules:

1) *Rule 1: A minimum of 600 booklet sets must be double scored and used in the within-country reliability study.*

2) *Rule 2: Each scorer must have a minimum of 125 scores that can be matched to scores from one other scorer.*

12.4 Outcomes of the scoring reliability studies

Within- and cross-country reliability studies were conducted in both the PIAAC Field Test and Main Study.

The Main Study data showed a high degree of agreement for within-country scoring reliability, averaging 99.1% and surpassing the goal of 95%. It should be noted that a few countries showed 100% agreement between the main score and reliability score for one or more domains. This level of agreement has not been seen in previous international surveys of adult skills such as IALS and ALL. The most likely explanation for this finding is that in a few cases, countries implemented a resolution process that eliminated any scoring discrepancies.

The Main Study data also showed that average scoring accuracy across countries was very high, averaging 96.7% agreement. The cross-country reliability measures obtained from the anchor booklet scoring ranged from 89.9% to 98.5% across participating countries. Only three countries were below 95%. Thus the use of the anchor booklets verified that overall agreement across countries was good and allowed us to achieve common item parameters across countries, with very few items being assigned unique item parameters.

These data for both the within- and cross-country reliability studies demonstrate the success of international scoring training and the national application of that training. Overall, the data support that the result of this work by the Consortium and participating countries resulted in accurate and comparable scoring of the PIAAC paper-based items.

Chapter 13: Data Management Procedures

Ralph Carstens and Tim Daniel, IEA Data Processing and Research Center

13.1 Overview

In PIAAC, as in any multinational survey, it is a challenge to minimize total survey error, part of which can be introduced during capture, coding and processing of data. Subsequent steps in a survey process depend on the quality of the data that was originally collected. Errors during data capture, coding and processing of the data are difficult if not impossible from which to recover.

PIAAC administered an assessment of adult skills in two modes (computer and/or paper) in addition to a computer-assisted administration of a BQ. Design, data structures and formats in PIAAC are quite complex. For example, rich auxiliary and behavioral data, such as response times and navigation information were collected and processed in addition to the raw responses to support instrument validation, analysis and reporting.

Given these complexities – the timeline under which PIAAC was carried out and the diversity of contexts in which it was administered – it was imperative to standardize, as much as practically possible, the procedures as they relate to the national and international data management. A comprehensive manual, training sessions, a range of other materials, and in particular, a mandatory data management software were designed to help NPMs and their National Data Managers (NDMs; more on this role later) to carry out their tasks, prevent introduction of errors, and reduce the amount of effort and time involved in resolving them. Approaches had to be generally strict yet flexible at the same time to accommodate for some idiosyncrasies and needs (e.g., with respect to data sharing constraints) as part of the country-by-country data management process. In order to prepare a high-quality database (i.e., one that is valid, reliable and comparable) with the highest possible analytical utility, a variety of quality control processes and procedures were implemented.

This chapter summarizes the collaborative efforts, strategies and processes resulting in the rich, standardized international master database supporting all PIAAC reporting. The final PIAAC international master database included more than 1,700 international variables. In addition, more than 1,200 national variables (relating to adaptations and extensions) were defined, collected and processed for the 24 countries participating in the first round of the first cycle of PIAAC.

13.1.1 Tasks and responsibilities at the international level

The design and implementation of PIAAC was the responsibility of an international consortium of institutions led by Educational Testing Service (ETS). In this Consortium, the International Association for the Evaluation of Educational Achievement (IEA) Data Processing and Research Center (DPC) in Hamburg, Germany, had primary responsibility for designing, facilitating and supporting the data management at the national level, as well as the overall data management at the international level. In particular, the IEA DPC:

- proposed standards, guidelines and recommendations for the data work in countries;
- created and provided software, codebooks and manuals to countries;
- organized and conducted data management trainings;
- supported countries during the national database building;
- managed, processed and cleaned data at the international level;
- prepared analysis and dissemination databases for use by the Consortium, the OECD and countries; and
- provided data analysis software (see Chapter 23).

Conducting a study like PIAAC would not be possible without close cooperation and consultation among all stakeholders. These were the roles fulfilled by each partner in achieving a quality data product:

- ETS: review, cleaning, quality control and support with respect to interview workflow as well as cognitive response and log data (aggregate and full), release of data products to the Consortium, countries and the OECD;
- ROA: review, cleaning, quality control and support with respect to BQ data, questionnaire flow, harmonization of information from national adaptations, and coding of occupation and industry;
- Westat: review, cleaning, quality control and support with respect to sampling, weighting and survey operations related data; and
- OECD: overall review, quality control and support with respect to the resulting data products.

13.1.2 Tasks and responsibilities at the national level

Each participating country appointed an NPM to take responsibility for implementing PIAAC at the national level. The NPM had overall responsibility for ensuring that all required tasks, especially those relating to the production of a quality national database, were carried out on schedule and in accordance with the specified international standards and quality targets. The NPM was responsible for supervising, organizing and delegating all data management work. By "data management," we refer to the collective set of activities and tasks that each country had to perform to produce the required national database. This included the adaptation of codebooks, integration of data from the national PIAAC interview systems, manual capture of data after scoring, export/import of data required for coding (e.g., occupation), data verification and validation, and eventually submission of the national PIAAC database to the Consortium.

Because data-related tasks tend to be highly technical and require special skills, the Consortium recommended that an NDM be appointed by each NPM. The NDM was responsible for the day-to-day data management tasks within the country, was expected to carefully review all provided information and instructions, participate in all applicable trainings, supervise local data work, and, most importantly, communicate on data cleaning with the IEA in a timely manner. The NPM and NDM were expected to be supported by staff or organizations for manual data capture, scoring and coding during

the applicable phases of the survey. The contribution that national technical personnel made was crucial to the survey's success and quality.

13.2 Key data management systems and integration processes at the National Center

13.2.1 Data management software, manuals and training

To standardize the national data work, countries were provided with a customized and extended version of the IEA Data Management Expert (DME) software originally designed and implemented for IEA work including Trends in International Mathematics and Science Study and Progress in International Reading Literacy Study. The IEA DME software supported data management at the National Center(s) after data collection. The IEA DME software was written in C# against the Microsoft .NET 4.0 framework and made use of a desktop, in-process variation of Microsoft SQL Server, more specifically, Microsoft SQL Server Compact 3.5 (SP2). Two versions of the software were created, one for the Field Test and one reflecting the revised assessment design and processes for the Main Study. The following list presents the key features of the IEA DME software and the customization to the PIAAC context:

- maintenance of a single, multi-table, robust and relational database for the integration of all sampling, response, workflow, log, scoring and coding data;
- documentation of the international as well as national record layout (codebook/code plan) and support for the addition and adaptation of national variables within constraints;
- extraction, transformation and storage of data from the various sources in PIAAC, most importantly the interview system;
- export and import to and from Excel; comma-separated and flat text files to interface with external processes, for example, the coding of occupation or the import of sample design data;
- manual data capture from scoring and response capture sheets as well as checks for double captured data;
- validation during import, manual entry and on demand by using pre-specified validation rules by variable, across variables, and across data sources using validity reports and statistics;
- supports for work on separate computers for data capture via file merging; and
- access control by using "roles" for managers and named data capture staff.

In concert with the IEA DME software, countries were provided with a comprehensive, 200-page data management manual detailing the processes, steps and checklists to be followed from the moment that the national interview systems, case management systems and paper instruments were readied for collection until the moment when national databases were submitted and initial data cleaning completed. Again, a Field Test manual and a revised/extended Main Study manual were provided.

Prior to the Field Test and again before the Main Study, NDMs or (in their absence) the NPM were expected to participate in comprehensive, data management trainings. Participation in these trainings was vital for the success of the project. These trainings focused on the setup and use of the provided

IEA DME software, the way it interacted with the assessment designs and interview system, the incorporation of national adaptations made in the BQ in codebooks, the integration testing between the national interview system and the data extraction logic, the import/export of relevant data stored in national case management systems or resulting from scoring processes, manual data capture from scoring sheets and the overall validation and verification of the database's completeness and consistency.

13.2.2 Codebook, database structure, record and value representation

Given the study's design and the technologies, the data structures and formats were relatively complex and somewhat different from those found in other school based large-scale international surveys. A variety of data sources were combined to build the national and international analysis and dissemination databases in PIAAC. The information in the database originated from the following assessment components, modules, sources and processes, mainly:

- sample design information (e.g., ID numbers, selection probabilities, stratification);
- screening and/or disposition information from countries' case management systems;
- interviewers' input into, or automatic import of, data into the case initialization module;
- interviewers' input into the BQ via the CAPI;
- behavioral/auxiliary information for the BQ (e.g., answers selection, timing, language changes, forward or backward navigation, consistency edits);
- interviewers' input and respondents' actions in the core modules;
- respondents' answers, detailed actions, timing and auto-assigned scores in the CBA;
- workflow information such as random numbers used in routing, automatically or interviewer assigned disposition codes, and timing information;
- respondents' original answers in the paper-based exercise and the reading components;
- countries' scoring and capture of scoring sheets for the paper-based exercise and the reading components (where used); and
- countries' coding of responses relating to the industry, occupation, language, country and region.

The PIAAC database included information from the sources above, and there was much more to consider. The interviews and exercises may have followed a variety of trajectories, data may have been generated for some respondents yet not others, and some data were captured during administration whereas other data were integrated after collection (for example, codes for occupation). Taking all this into account, the Consortium organized the data into a single relational database, though in multiple tables within this database. Each table corresponded to one or more modules in the survey. Later during the international data processing, most of these sources and tables were combined to form a more familiar "flat" analysis file.

The key concepts used in the PIAAC data management and database structure were *variables* (including their *value* and *missing schemes)*, *datasets* and *instruments*. The combinations of information in these entities form the PIAAC *codebooks* (elsewhere called *metadata* or *record layout*). In addition, data in PIAAC is stored by means of *data records* and eventually *data values.*

Variables correspond to fields (columns) in the resulting database. Each variable in PIAAC was defined by a set of attributes. The IEA DME software "reused" variable definitions in a number of ways. Variables were defined once, and only once, and then referenced in the corresponding *datasets* or *instruments* in which they were assembled. Secondly, *value* and *missing schemes* in the IEA DME software were defined only once and then referenced by the corresponding variables rather than being defined multiple times. This recycling of variables and schemes allowed efficient and consistent definition and adaptation of codebooks. *Variable attributes* were defined with the two most commonly used packages for statistical data, SAS and SPSS. Systematic and consistent *variable naming conventions* were applied for each component of PIAAC. Whereas variables of the BQ followed a naming convention derived from work at Statistics Canada, naming conventions for other assessment components followed a generic logic designed for PIAAC and took trend aspects into account (e.g., item naming found in IALs and ALL). Note, that variable names present in the exported interview system result files used a different naming convention and had to be renamed on import into the IEA DME database and for further analysis.

Each of the 33 datasets in PIAAC comprised the information for specific parts of the survey. A *dataset* is a logical collection of rows and columns where each column represents a clearly defined variable identified by its unique name and each row corresponds to a record of valid or missing values collected for a case or sampled person. Table 13.1 below describes the type of information they held along with the respective sources. Note that not all information was stored as part of the country database. Full cognitive log information was stored in its native format (XML) and provided to the Consortium at the time of data submission outside of the database maintained by the IEA DME software.

Instruments as used in the IEA DME software and database are logical sets of variables, i.e., a subset of variables selected in a particular sequence from a larger set of variables. Instruments were used for the manual data capture of paper scoring and response capture sheets.

Data records in the IEA DME software and database simply corresponded to a single row in a dataset, identified by one or more unique identifiers. Depending on a sampled person's path through the interview, data records for a single person existed in multiple but not all datasets. Each data record in a dataset had the same set of variables, and for each of these variables, either a valid value or a missing value was stored.

Table 13.1: Main Study datasets and sources

Dataset	Description	Specifics	Unique Identifiers	Source
SDIF	Sample Design International File	n/a	CASEID and/or PERSID (depending on sample design)	Imported from a country's study management system
BQR	BQ and global workflow	Results	PERSID	Extracted from BQ result files (XML)
BQL		Log	PERSID and SEQUENCE	Extracted from BQ log result files (XML)
BQC		Coded responses	PERSID	Imported from a country's coding process/system
CBR	Computer-based exercise	Results	PERSID	Extracted from cognitive result files (XML)
PCM1/ACM1	Paper Core Booklet (respondents or anchor)	**Main** scoring **First** capture	PERSID Secondary IDs: SCORERID_PPC, BOOKID_PPC, BUNDLEID_PPC, KEYOPID_PPC	Manually captured from core booklet scoring sheets
PCM2/ACM2		**Main** scoring **Second** capture		
PCR1/ACR1		**Reliability** scoring **First** capture		
PCR2/ACR2		**Reliability** scoring **Second** capture		
PLM1/ALM1	Paper Literacy Booklet (respondents or anchor)	**Main** scoring **First** capture	PERSID Secondary IDs: SCORERID_PP1, BOOKID_PP1, BUNDLEID_PP1, KEYOPID_PP1	Manually captured from literacy booklet scoring sheets
PLM2/ALM2		**Main** scoring **Second** capture		
PLR1/ALR1		**Reliability** scoring **First** capture		
PLR2/ALR2		**Reliability** scoring **Second** capture		
PNM1/ANM1	Paper Numeracy Booklet (respondents or anchor)	**Main** scoring **First** capture	PERSID Secondary IDs: SCORERID_PP2, BOOKID_PP2, BUNDLEID_PP2, KEYOPID_PP2	Manually captured from numeracy booklet scoring sheets
PNM2/ANM2		**Main** scoring **Second** capture		
PNR1/ANR1		**Reliability** scoring **First** capture		
PNR2/ANR2		**Reliability** scoring **Second** capture		
RCM1	Paper Reading Components Booklet	**Main** scoring **First** capture	PERSID Secondary IDs: SCORERID_PRC, BOOKID_PRC, BUNDLEID_PRC, KEYOPID_PRC	Manually captured from reading components response capture sheets
RCM2		**Main** scoring **Second** capture		
RCR1		**Reliability** scoring **First** capture		
RCR2		**Reliability** scoring **Second** capture		

Each logical dataset corresponded to a physical table in the relational database managed by the IEA DME software and had one or more identification variables in its first positions. Identification variables corresponded to units, entities or people in the survey or those that participated in its conduct. The identification variables used in the PIAAC Main Study are described below.

- CNTRYID: The country ID holds a 3-digit numeric code that follows the ISO 3166/UN M49 standard.

- CASEID: This is the household operational ID. It was assigned at the sampling stage for countries using a household sampling design.

- PERSID: This is the sampled person's operational identification number that uniquely identifies him or her. The PERSID variable appeared in all datasets as assigned at the sampling stage. In the case of household sampling, the PERSID was only assigned when within-household screening was completed and persons were sampled. The PERSID included a mandatory check digit based on approaches for universal product codes (UPC). The check digit proved to be highly efficient and effective in avoiding or identifying the vast majority of key entry mistakes.

- SCORERID_xxx: This ID identified the persons who scored paper-based exercise booklets on the corresponding sheets.

- KEYOPID_xxx: This ID identified the persons entering the values from scoring and/or response capture sheets, the key operators.

- BOOKID_xxx: PIAAC required countries to assign a unique booklet ID (serial number) to each printed paper-based exercise and reading component booklet.

- BUNDLEID_xxx: The bundle ID identified the bundles and their contained paper-based exercise booklets as defined by the international scoring design.

The following list provides a brief description of these datasets and the types of information they held:

- SDIF – Sample Design International File

    - The SDIF dataset held the required and optional variables as defined by the international sampling standards and included unique identifiers, sampling IDs, selection probabilities, stratification information, screening information, demographic information, disposition codes, information for variance estimation, raking dimensions and nonresponse adjustments variables.

- BQR – BQ and global workflow – Data

    - The dataset comprised explicit, implicit or derived variables captured as part of the general workflow, more specifically from the following case initialization module, the BQ (the bulk of the BQR dataset, hence the name), the CBA Core Stage 1 (ICT tasks), the administration of paper-based booklets (core, literacy, numeracy and reading components) and the observation module. The BQ variables in this dataset were subject to adaptation and extension, and any deviations from the international codebooks had to be reflected prior to production use.

- BQL – BQ and global workflow – Log

    - The interview system maintained a log file of actions and events relating to the same modules as described above for the BQR dataset. This log/audit dataset held information about the interviewer's actions during the CAPI, that is, any actions or variables that were explicitly shown on screen. This dataset contained multiple records per person. Each data record included information about the type of event, a timestamp, the item ID where the event occurred, and, where applicable, a value associated with the event depending on the type.

- BQC – BQ – Coded responses

    - Some of the answers to the BQ that were captured during the interview were subject to coding according to schemes for occupation (International Standard Classification of Occupations, or ISCO, 2008), industry (ISIC rev 4), language (ISO 639-2 alpha-3), country (UN M49 numeric) and region (TL2 OECD classification of geographical regions).

- CBR – Computer-based exercise – Results

    - The variables in this dataset represented the different pieces of information directly captured or derived from the computer-based exercise. It held all variables that were related to the computer-based literacy, numeracy and problem-solving items, more specifically the actual response; interim, and/or final scored responses; the number of defined action; the time elapsed before the respondent's first action; and the total time taken for the item.

- PCM1, PCM2, PCR1 and PCR2 – Paper Core Booklet

    - The PCM1 and the three related reliability ("R") and double capture ("2") datasets PCM2, PCR1 and PCR2 contained data for all items in the Paper Core Booklet. The responses to this booklet were scored on Core Booklet Scoring Sheets, and scored responses were captured and stored rather than the actual responses.

- PLM1, PLM2, PLR1, PLR2, PNM1, PNM2, PNR1 and PNR2 – Paper Literacy/Numeracy Booklet

    - The PLM1/PNM1 and the three related reliability ("R") and double-punching ("2") datasets PLM2/PNM2, PLR1/PNR1 and PLR2/PNR2 contained variables for all items in the Paper Literacy Booklet. The responses to this booklet were scored on Literacy Booklet Scoring Sheets and scored responses were stored rather than the actual responses.

- RCM1, RCM2, RCR1 and RCR2 – Paper Reading Components Booklet

    - The RCM1 and the three related reliability ("R") and double-punching ("2") datasets RCM2, RCR1 and RCR2 contained variables for all items in the Paper Reading Components Booklet. The responses to this booklet were captured on Reading Components Response Capture Sheets and, in contrast to the other paper-based booklets, actual responses were stored rather than the scored responses.

- AxM1, AxM2, AxR1 and AxR2 – Anchor booklets

    - These datasets held data originating from the anchor booklets scoring process in the cross-country scoring reliability study.

Each of the above datasets included records per person or case depending on the trajectory through the assessment. Each intersection of a variable and a record in the above datasets either held a valid or a missing value. Valid values were the individual pieces of collected information conforming to the corresponding variable specification, that is, the defined lengths, value schemes or ranges. The majority of variables in PIAAC were numeric and had a value scheme assigned to them (e.g., "1" corresponded to "Yes," "2" corresponded to "No"). Wherever possible, value schemes limited the possible values that a variable could take. Missing data/values in a survey may occur when there are no data whatsoever for a respondent (unit nonresponse) or when some variables for a respondent are unknown, cannot be known, refused or otherwise not useful (item nonresponse). Missing data were distinguished semantically in essentially two broad groups: i) data that were missing by design, and ii) data that were supposed to be there but were not provided, or omitted. While missing data are inevitable in survey research, it is important to describe it properly and use it as information in itself to evaluate procedures, refine instruments or make assumptions about the mechanisms responsible as well as the likely consequences for the validity and possible bias of estimates. Analysis of item nonresponse is an important part of quality control, and consistent use of missing values ensured that the PIAAC data files contain detailed enough information on unit and item nonresponse (see also Chapter 16 on item-level nonresponse bias analysis).

The schemes to describe missing data in PIAAC during the time of data capture and building the national database were relatively simple and distinguished only a few types of missing data. In the following, the key missing value schemes used in PIAAC at the time of data integration are listed. A description of the missing values in the resulting public-use data products is presented in Chapter 23.

- Default missing scheme

    - This scheme was used for a large number of variables in PIAAC for which either a valid value was expected to exist for each and every data record or where there was no need to distinguish reasons for missing data during capture and database building.

- BQ missing scheme (numeric variables only)

    - All questions directed to the sampled person in the BQ explicitly included the options "refused" and "don't know." This missing scheme therefore distinguished the nature of the missing data and retained the information captured during the interview. The scheme applied to all BQ variables as well as most coded variables:

        - *Don't know*: The sampled person was responsible for this type of item nonresponse by indicating "I don't know" or similarly.
        - *Refused*: The sampled person was responsible for this type of item nonresponse by refusing to answer the question.
        - *Not stated /inferred*: This is a systemic, catch-all nonresponse and was assigned if a variable was expected to hold a valid value but the value was missing, out of range, otherwise useless, and could not be reconciled or fixed.

- Free-text entry (FTE) missing scheme

    - PIAAC used a number of free text entry responses for occupation, field of industry, country, language, foreign qualifications and some other fields in the BQ. In order to retain the information provided by the respondents and/or the interviewer for later analysis and disambiguation, the IEA DME software imported missing values for any free text entries in the CAPI system as string constants, that is, either "[REFUSED]" or "[DON'T KNOW]".

13.2.3 National adaptations and extensions

Along with the IEA DME setup, countries were provided with an international codebook template. The international codebook for PIAAC included each and every variable and dataset known to the survey, including all variables relating to two international options (problem solving in technology-rich environments and reading components). Regardless of a country's participation in these options, the codebook and databases always included and displayed the variables for all components, but they simply had missing data if an option was not used.

The general approach to national variables was to include all international as well as all national data in a country's database in order to harmonize and map data post-collection. To do so, the international master codebook had to be adapted to reflect the national BQ in which countries adapted certain variables to their national and cultural settings as well as introducing additional national questions or adaptations/additions to the internationally ones. All adaptations and extensions applied in the national BQ had to be reflected in the codebooks as well in order to parse and store the information stored by the interview system. These adaptations related to the creation and specification of national variables, associated value schemes, as well as the adaptation of valid ranges for international variables as applicable (e.g., for currency units).

The adaptation of the international codebooks to reflect the national BQ was the responsibility of the NDM and performed according to instructions and guidelines provided by the Consortium. The international codebook template was used by NDMs as the starting point to which adaptations and extensions for national use were applied through controlled interfaces in the IEA DME software. The key input for this work was the national BQ itself as well as the agreed-upon Background Questionnaire Adaptation Spreadsheet (BQAS). As a key strategy, any adaptations to BQ questions had to be reflected under a new national variable name in order to clearly identify the likely need to harmonize, map or recode national to international variables after collection. A naming convention was applied that uniquely identified each national variable within and across countries. For example, a national variable for Germany that was based on item ABC would receive a name such as "ABC_DEU". In the case of extensions, that is, questions and variables unrelated to the domains and contents of the international BQ, a further variation using the suffix "X" would have been used, resulting for example in a name such as "ABC_DEUX". After all national adaptations were reflected in the codebooks, NDMs were responsible for thoroughly testing the import and correct mapping of data from the interview system, then submitting these codebook to the Consortium for further review, verification and for preparing international processing.

13.2.4 Data extraction from the computer-based delivery system

All data collected for PIAAC was integrated into a single national database managed by the IEA DME software. The primary means of integrating the database were by i) importing data from the national interview system, ii) manually entering the data via the data capture interfaces, or iii) importing data from national systems or processes. The bulk of the data in PIAAC naturally originated from the

interview sessions and was stored in per-respondent result files in .zip format, each including a sizable set of XML format files for the various components of the assessment (BQ, core cognitive modules, main cognitive modules and observation module).

The contents of the per-respondent result file archives were generally stored as single records and mapped to the variables defined in the BQR, BQL and CBR datasets introduced earlier. In doing so, data were extracted from the individual XML files stored by the interview system, transformed as necessary, and then loaded into the respective target datasets (tables). Result data for the BQ was stored in datasets BQR and log data in BQL; cognitive result file information were combined from multiple XML files to form a single record in dataset CBR. The transformation comprised the mostly one-to-one mapping of values yet changed the data type from the generally used string types in the interview system to numeric values in the target database. For example, originally stored string literals such as "01" were stored as a numeric value "1". Missing values were mapped as well, from string literal "DK" for "Don't know" to a numeric value depending on the length of the variables (code 7, 97, 997 and so on). A refused response ("RF" in the result files) was mapped to numeric code 8, 98, 998 and so on in the database.

Additional transformation logic was applied in the following contexts:

- For multiple-choice items allowing more than one response in the BQ, values stored under the same name in the result files were mapped to individual variables.
- For currency values in the BQ data, any currency symbols were stripped.
- For numeric values with decimal places, thousand separators were stripped.
- For the BQ and workflow log data, string literals for event types were mapped to a numeric value scheme. For example, the event type "INTERVIEW_START" was mapped to the labeled value "1" in the target dataset BQL.
- Relating to workflow information, timer values for the reading components were transformed from string values formatted as "minutes:seconds:tenths" (e.g., "1:59:9") to tenths of seconds.
- For cognitive results, a name-mapping table matched long result variable names that were idiosyncratic to the interview system or sometimes not fully compliant with the naming conventions to shorter names used throughout all subsequent data products and analysis, such as names limited to eight characters in length.

As far as possible, the extraction and transformation logic checked for the integrity of the result file archive. However, given that some respondents broke off the interview and technical problems occasionally occurred, result files were parsed in a positivistic way, meaning that contents of the archives were parsed, provided that the main BQ result file existed along with any other materials found in the archive. As described before, NDMs were responsible for testing the integration between the interview system and the IEA DME maintained national database to make sure that i) all files, variables and values were mapped as expected, and ii) all nationally adapted or extended variables in the interview system were also reflected in the national codebooks. Certain values in the result files were only of interest at the time of collection and were not parsed and stored in the national database. For the most part, this related to strings for dynamic texts or interim values stored for some routing logic.

The full information, native CBA log files holding information on respondents' work on the cognitive assessment items were not parsed and loaded into the database. Instead, these were merely extracted from the result file archives and stored in separate folders. Countries were requested to provide these log files to the Consortium for further processing together with their initial data submission.

13.2.5 Data capture from scoring sheets and double capture reliability

Data capture is the process by which data collected on paper (e.g., on questionnaires, scoring sheets, or administrative records) are put in a machine-readable form. This section provides a description of the default process in PIAAC, that is, the recording of scored responses on scoring and response capture sheets and the subsequent capture of this information by means of the IEA DME's data capture interfaces.

According to the PIAAC technical standards and guidelines, the scoring of the paper-based exercise booklets had to be done twice by two different scorers following a scoring design recommended by the Consortium. Further, the manual data capture of each scoring (literacy/numeracy) or response capture (reading components) sheet had to be done twice by different key operators. This approach, although labor-intensive, allowed for separate investigation of error or disagreement introduced by the scoring and the data capture processes. This requirement also held for the international scoring bundle (anchor booklets) used in the cross-country reliability study.

This general data capture process was documented in detail in the data management manual along with advice on how to recruit, train and supervise key operators as well as operational recommendations for logistics, forming batches of materials for data capture and batch header examples. The manual entry of data in the IEA DME software was restricted to valid and missing values as defined by the respective scoring guides for literacy, numeracy and reading component items, and these permissible definitions were reflected in codebooks. The header of each scoring or response capture sheet included: the respondent's ID, the booklet ID, the scorer ID, the bundle ID, the score run (main or reliability) and the date of scoring. The information on the response capture sheets was simple and straightforward, allowing for efficient capture of data from sheets using numeric key pads. Respondent IDs were validated on capture. Similarly, out-of-range values or undefined codes were flagged and data capture was not allowed to proceed without correction. Partial entry was not supported. Each sheet had to be captured completely or not at all.

The set of rules provided to NDMs and their key operator staff included the following key advice:

- All scoring and response capture sheets must be fully completed before data entry can start. This included the header information on each sheet. In case there was missing, conflicting, otherwise or inexplicit information on any sheet handed to a key operator, these sheets must be returned to scorers (or the scoring process in more general terms) in order to be scored and filled correctly. Scorers were advised to revisit the original paper material in doing so.

- Data must be entered exactly as values appeared on the sheet, that is, without any corrections, unjustified interpretations or imputation.

- Checks for data capture accuracy and reliability must be checked on a regular basis, that is, at least once a week during the data capture process. This allowed the NDM to detect general misconceptions about the data capture rules or systematic misconceptions, underperformance or high incidental error rates of individual staff members. In addition, the Consortium recommended that the NDM monitor the accuracy of data entry on a more frequent, preferably daily, basis during the beginning of the manual data capture work.

The IEA DPC required countries to double enter all scoring/response capture sheets twice followed by a full reconciliation of any identified discrepancies by consulting the original paper materials. Checks for the accuracy and reliability of this double capture were built into the IEA DME software. This component listed all instances of disagreeing data and further provided an overall percentage of discrepancies. This procedure allowed the NDM to resolve data discrepancies before submission and the Consortium to estimate the agreement between key operators as well as the overall reliability of the manual data capture.

No margins were set for the acceptable levels of disagreeing data as a result of double capture. The Consortiums expected the manual key data capture to be 100% accurate and NDMs to resolve all identified discrepancies by revisiting the original scoring or response capture sheets and correcting the concerned values. All countries complied with this requirement and the evidence of data capture reliability provided by countries suggested that data were virtually free of data-capture error.

A number of countries requested permission to use alternative data capture means and processes. For example, some countries used scanning, followed by on-screen scoring processes, essentially collapsing the scoring and data capture processes into a single process. The Consortium carefully reviewed such plans and accepted deviations from the standard provided that countries were able to demonstrate similar or better quality. In these cases, the data resulting from these alternative processes were imported directly into the respective datasets.

13.2.6 Import of sample design data from study management systems

The SDIF was a mandatory deliverable from countries to the Consortium; the standard mode of transfer was as part of the national database. Countries were required to make use of one of the three supported import file formats (comma-separated, fixed length or Excel) to load SDIF-related data into the respective dataset. The actual import of the sample design information data into the SDIF dataset, using the import feature, was straightforward. Most variables in the SDIF were stored in a country's study management system. To import the sample design information countries had to:

- consult with Westat on the applicable variables in the SDIF to fill given the national sample design, plans for post-stratification and the like;
- export the applicable variables from the national study management system (or compiled/merged them from multiple data sources if applicable) into a single file in one of the import formats supported by the IEA DME software;
- ensure that the data contained were complete, accurate and formatted as defined by the respective codebook;
- ensure that variables not applicable to the national sample design were either represented by blanks in fixed-length import files and empty cells in CSV and Excel, or not included at all in the import file;
- ensure that all records in the import file were uniquely identified by a valid person ID and/or case ID as applicable; and
- ensure that any numerical variables used no more than the specified number of decimals.

Whereas the above stated prerequisites as well as file structure and variable definitions were automatically validated on import, no checks for completeness of SDIF data could be run given the

varying sample designs across countries. Sampling- and weighting-related data were reviewed by Westat following the submission of national databases, and numerous corrections and additions were processed for a large number of countries until a complete, valid and accurate SDIF could be finalized and receive signoff prior to weighting.

13.2.7 Import of coding data from external processes

A number of free text entry variables in the BQ were not only captured during the interview but were subject to coding according to schemes for:

- Education: International Standard Classification of Education, or ISCED, 1997 long, ISCED 1997 broad fields of education and training, ISCED 1997 short
- Occupation: ISCO 2008 at the four-digit unit group level
- Industry: International Standard Industrial Classification of All Economic Activities, or ISIC, Revision 4 at the four-digit class level
- Language: ISO 639-2/T (alpha-3/terminologic)
- Country: UN M49 numeric
- Region: TL2 level of the OECD classification of geographical regions

The BQ variables that served as inputs for coding, as stipulated by the BQ framework as well as the corresponding validation and analysis plans, were documented as part of the data management manual. Related advice and training was given to countries as part of data management trainings. Separate expert trainings were held for the coding of occupation against the ISCO standard and industry against the ISIC standard. The respective coding schemes were included in the codebooks to facilitate validation at the time of database integration.

More specifically, the following instances of coding were required from countries:

- Coding/mapping general ISCED responses
    - All countries posed education-related questions in a closed format using national classification. In that sense, no actual coding was carried out (except in the case of "foreign qualifications" that had to be coded; see below). Countries either converted these national codes into ISCED 1997 themselves or provided conversion rules. Countries were required to deliver both the code in the national classification and the corresponding international code.
    - Countries were required to code the highest foreign qualification for all respondents who reported a foreign qualification using responses to B_S01a1, the name of the "foreign" highest qualification (write-in), and B_Q01a3, the nationally corresponding level of the "foreign" highest qualification (a nationally adapted list).
    - ISCED codes for respondents' highest foreign qualification were stored in variable ISCED_HF in the BQC dataset.

- The missing scheme for the variable ISCED_HF was the standard numeric scheme for the BQ. Because ISECD_HF was of length 2, the missing codes were also of length 2:

    - *Don't know* was used if the two raw responses were marked as "don't know."
    - *Refused* was used if at least one raw response was marked as "refused."
    - *Not stated* was used if at least one raw response was given but not interpretable or otherwise useless and it could not be reconciled or fixed.

- Coding of occupation to ISCO 2008 and coding of industry to ISIC Revision 4

    - Four-digit codes from the 2008 ISCO-08 were used to code the occupation of the respondent (current and last job as applicable). The corresponding target variables in the BQC dataset were: ISCO08_C (current job) and ISCO08_L (last job).

    - Countries that opted to initially code in ISCO 1988 were made aware that no automatic conversion from the ISCO 1988 to ISCO 2008 existed: certain codes in ISCO 1988 were split up into multiple codes in ISCO 2008, while other codes were merged. Therefore a manual verification of the correspondence was required for these codes.

    - If a country had coded in ISCO 1988, this coding had to be provided as well as the coding in ISCO 2008. The corresponding target variables in the BQC dataset were: ISCO88_C (current job) and ISCO88_L (last job).

    - Four-digit codes from ISIC, Revision 41, were used to directly code the sector in which the respondent was working (current and last job as applicable). The corresponding target variables in the BQC dataset were: ISIC4_C (current job) and ISIC4_L (last job).

    - The missing scheme for the variables ISCO08_C, ISCO08_L, ISCO88_C, ISCO88_L, ISIC4_C and ISIC4_L was a special numeric scheme. Because the ISCO/ISIC variables were strings of length 4, the missing codes were also strings of length 4:

        - Don't know (code "9997") was used if all of the raw responses were marked as "don't know."
        - Refused (code "9998") was used if one or all of the raw responses were marked as "refused."
        - Not stated (code "9999") was used if at least one raw response was given but not interpretable or otherwise useless and it could not be reconciled or fixed.

    - The coding of occupation and industry to ISCO/ISIC was subject to quality control implemented by ROA. As part of the data submission, countries were required to provide corresponding evidence and reports comparing the unweighted and weighted distributions of occupational groups at the two-digit level to external information from, for example, the most recent national labor-force survey.

    - Responses that could not be coded at the four-digit level, that is, codes at the one-, two-, or three-digit level, were subjected to review by a coding expert.

[1] http://unstats.un.org/unsd/cr/registry/isic-4.asp

- Some countries were not legally able to disclose ISCO/ISIC data at the four-digit level and submitted data only at the permissible level of detail.

- Coding of language to ISO 639-2/T

  - For language-related free-text entries, the ISO 639-2/T alpha3 (terminologic) scheme was used.

  - The corresponding target variables in the BQC dataset were LNG_L1 (first language learned at home and still understood), LNG_L2 (second language learned at home and still understood) and LNG_HOME (language most often spoken at home). By their very nature, ISO 639-2 three-digit alphanumeric codes for languages were defined as strings of length 3 in the BQC dataset.

  - The coding of languages involved two steps:

    - Mapping the numeric responses to the national closed format language questions in the BQ to the codes in ISO 639-2.
    - Coding the write-in responses to the "other" languages questions in the BQ to the codes in ISO 639-2.

  - The missing scheme for the variables LNG_L1, LNG_L2 and LNG_HOME was a special numeric scheme. Because the ISO 639-2 variables were strings of length 3, the missing codes were also strings of length 3:

    - Don't know (code "997") was used if the raw response was marked as "don't know."
    - Refused (code "998") was used if the raw response was marked as "refused."
    - Not stated (code "999") was used if a raw response was given but not interpretable, otherwise useless, not covered by the scheme and it could not be reconciled or fixed.

- Coding of country to UN M49

  - Countries coded the country names in various questions of the BQ using the numerical codes of UN M49. In most cases, a country-specific list of countries was used that covered the most relevant countries plus a category "other." Both the "listed" countries as well as the "other" category were converted by the countries into UN M49.

  - The name of the country reflected the CURRENT name of the country in which the highest qualification was attained or in which the respondent was born, not the name of the country in the past (regardless of whether the question related to the past, e.g., country of birth).

  - The corresponding target variables in the BQC dataset were CNT_H (country of highest qualification) and CNT_BRTH (country of birth). UN M49 country codes were defined as integers of length 3 in the BQC dataset.

  - The coding of countries involved two steps:

        - Mapping the numeric responses to the national closed format country questions in the BQ to the codes in UN M49.
        - Coding the write-in responses to the "other" country questions in the BQ to the codes in UN M49.
    - The missing scheme for the variables CNT_H and CNT_BRTH was the standard, numeric BQ missing scheme. For the coding of country, the missing codes were used as follows:
        - Don't know was used if the raw response was marked as "don't know."
        - Refused was used if the raw response was marked as "refused."
        - Not stated was used if a raw response was given but not interpretable, otherwise useless, not covered by the scheme and it could not be reconciled or fixed.
- Coding of region to OECD TL2
    - Countries were required to code the geographical region corresponding to the respondent's address at the TL2 level using the OECD classification of geographical regions, for example, "DE6" for a respondent in Hamburg, Germany. The corresponding target variable in the BQC dataset was REG_TL2, and this variable was defined as a string of length 5 in the BQC dataset.
    - The variable REG_TL2 was not derived from BQ responses but from sampling/survey control data. Therefore, the missing scheme for the variables REG_TL2 was the default missing scheme that only permitted "blank" as a missing value as data were expected to be available for all sampled persons.

13.3 Data verification and editing at the National Center

13.3.1 Validation, verification and record consistency checking prior to data submission

Each country was required to perform verification of the national database to identify and, if necessary, resolve errors and inconsistencies in the data. For carrying out this important part of the quality control work, tools to apply the minimally required checks as well as policies regarding the within-country editing of data were provided to countries.

Automated validation checks to detect values outside of the defined range for a variable, duplicate IDs and double data capture checks to detect and resolve data capture errors were made available as part of the IEA DME software. These checks were designed as an initial inspection of severe gaps or mismatches in the data and not intended to replace the more thorough data-cleaning process at the international level that was done centrally. Countries were required to run these on a regular basis. Further, record consistency checks were included in the software. The record consistency checks included 45 checks that identified possible inconsistent records across datasets. The checks were consecutively numbered and grouped by content:

- Checks 1 to 24 flagged possible linkage problems between datasets, that is, they listed respondent IDs that were expected to be found in a dataset, given their existence in another one and the interview flow;

- Checks 25 to 30 flagged possible problems in the scoring datasets, for example, an insufficient number of anchor booklets contained;

- Checks 31 to 43 flagged possible problems related to sampling information, for example, indication that two persons were sampled in the household but only one record for this household existed; and

- Checks 44 and 45 reported problems of general nature, especially related to technical problems and "out of design" cases.

In addition to the automated and consistency checks, the IEA DME software contained facilities to review descriptive statistics, including minimum, maximum, mean, median, variance, percentiles and sample standard deviations, as well as to cross tabulate variables for quality control purposes. NDMs were strongly urged to review frequency distributions of their data for plausibility and/or agreement with expectations. It was also important to verify the completeness and integrity of the database with respect to the included data records. Sampled persons in PIAAC followed a variety of paths through the interview, each generating records in one or more datasets yet not in others. In addition, the existence of data records also depended on whether the sampled person completed the entire interview, or broke off before its end and consequently didn't work on all of the applicable components. NDMs were advised and trained on the importance of checking the number and IDs of data records existing in the various tables of the database against the known and therefore expected numbers from survey records and study management systems.

13.3.2 Permissible and prohibited data editing and cleaning

Countries were requested to run the checks described so far in this chapter to ensure, as much as possible, that the within-country data capture and integration accurately and authentically reflected the values given by the sampled persons and/or the interviewers.

Countries were asked to refrain from implementing any type of general data-cleaning or data-flow editing on their own prior to the submission of the data. The Consortium partners requested original access to the types and the magnitude of, for example, outliers, implausible values or implausible combinations of these in order to refine the instruments and/or to identify problems with the translation of questionnaire items. However, countries were encouraged to make corrections to the data that were clearly attributable to the survey process, data-capture mistakes or similar misunderstandings made by, for example, the interviewer. Common examples of these edits included the correction of incorrectly recorded disposition codes or incorrect secondary IDs (e.g., booklet IDs). This was considered to be a part of the normal and mandated data verification and checking. Also, exceptions applied to instances of technical problems in the virtual machine (VM) where a disposition code "90" may have had to be assigned after data collection in those cases where on-site recovery was impossible and only partial data (or none at all) was extracted from the VM. Other exceptions related to reproducible and verified error sources, for example, residual BQ routing errors, recoding errors and so on which could be corrected using logical and verified correction procedures.

The Consortium received a number of requests to change/edit the data in order to make it more consistent across variables or more consistent with other data collections. The Consortium's consistent position communicated to countries was that data collected during the interview took precedence over wholesale interpretations or assumption without concrete verification or evidence indicating that originally corrected data were unreliable or invalid. Where no additional data collection was conducted, or counter information was available to override the original information, no change was implemented

or allowed. Exceptions were related to reproducible errors (e.g., routing, recoding, etc.). A small number of verifiable exceptions were made but required written documentation and pre-approval by the Consortium.

13.3.3 Confidentiality review, editing and variable suppression

Some countries had regulations and laws in place that restricted the sharing of data, as originally collected, with the Consortium and/or the OECD. The key goal of such disclosure control is usually to prevent the spontaneous or intentional identification of individuals in the full-information microdata. On the other hand, suppression of information or reduction of detail clearly has an impact on the analytical utility of the data. Therefore, both goals had to be carefully balanced. As a general directive, the OECD requested all countries to make available the largest permissible set of information at the highest level of disaggregation possible.

A small number of directly identifying variables that were collected during the case initialization were suppressed by default in any database exported for submission to the Consortium. This included the respondent's name, address, and telephone number. According to the technical standards, each country had to provide the Consortium with early notification of any rules affecting the disclosure and sharing of PIAAC sampling, operational or response data. Furthermore, each country was responsible for implementing any additional confidentiality measures in the database before delivery to the Consortium. Countries especially reviewed the sample design information (dataset SDIF) and the variables collected through the BQ (dataset BQR) with respect to indirectly identifying variables or otherwise sensitive information. Most importantly, any confidentiality edits changing the response values had to be applied prior to submitting data to the Consortium in order to work with identical values during processing, cleaning and analysis. The IEA DME software only supported the suppression of entire variables. All other measures had to be implemented under the responsibility of the country via the export/import functionality or by editing individual data cells.

The Consortium asked for complete and detailed documentation about any implemented measures to evaluate the impact on the analytical utility of the dataset, especially with respect to the introduction of bias, attenuation of within-variable variance, or between-variable correlations as a result of data suppression or perturbation. The majority of countries suppressed data at the variable level and submitted a database excluding certain types of information such as birth countries, original free text entries, full log information or detailed earnings values. These suppressions were carried forward throughout all subsequent data processing and analysis stages and into the public-use data products. Perturbation of original values according to the documentation known to the Consortium applied in two instances:

- Austria used statistical coarsening for the original, detailed earnings values (micro-aggregation).
- The United States perturbed data prior to submission following local standard operating procedures for large-scale surveys. Within-record consistency was maintained. The Consortium received no detailed account of these perturbations and consequently was unable to review, validate or assess the impact of these edits on the data or any inferences based on it.

A general procedure for the suppression of information from the for public-use databases was implemented after processing. These additional suppressions were handled by the Consortium in a standardized way. Exceptions to the general rule of suppressing an entire variable apply in these cases:

- Austria and Estonia suppressed single values given small frequencies for some language and country variables.

- Canada applied a small number of case-level suppressions that held values or combinations believed to identify sample and or population uniques.

13.3.4 Data submission and required documentation

After the collection, integration and verification of data, each country was responsible for submitting the required materials to the Consortium. The materials to be submitted electronically to the Consortium after the Main Study were the following:

- A single, integrated, verified, confidential and exported database per country in the IEA DME's format using the adapted national codebooks, that is, including all national variables and values (except for suppressions).

- A single zip archive including all original cognitive log files extracted and stored as part of the data parsing from the interview system.

- A free-format documentation with double-coding reliability evidence and explanations for QC purposes according to the technical standards and guidelines. The information requested comprised tables in which countries compared data collected in PIAAC with the most recent labor force survey (or equivalent) on the distribution of i) highest level of education, ii) labor force status, ii) occupation at the one- and two-digit level (ISCO 2008), as well as iv) sector of industry in 21 sections (ISIC, A-U).

- A comprehensive and detailed free-format documentation of implemented confidentiality edits, if any, and the effect of these edits on univariate and multivariate properties.

- A comprehensive and detailed free-format documentation of any other issues or notes that required attention by the Consortium during data processing and analysis. The document was expected to include notes for example pertaining to out-of-design cases, that is, respondents that did not follow the assessment design as prescribed or technical problems.

On export from the IEA DME software, a copy of the current national database was created. All values for all occurrences of a variable marked as "suppressed" in the codebook were set to blank values in the exported database. The national database exported was marked as non-productive and read-only.

Any data submission to the Consortium had to be made through secure channels. For this purpose, a SSL/TLS secured FTP site and a corresponding Web interface were set up. Document exchange folders were created for each country. Access to such a country exchange folder was limited to authorized staff members of the Consortium and the national center.

13.4 Data processing and editing at the international level

This section describes the process from the moment that national databases were received from countries until the moment that a preliminary international database, consisting of each national database, was produced. The main international data processing phase stretched from June to October 2012. The initial phase (June-July 2012) was used to clean data at the case level and with respect to all relevant fields in order to prepare and flag cases for weighting that are valid and comply with the PIAAC definition of "complete." The following months (August through October 2012) were used for any residual data cleaning and/or for the processing of additional, revised or erroneous data. Exceptions to this general timeline apply given the slightly differing schedules in countries' data submissions.

In general, the data processing for PIAAC was straightforward, carried out separately for each country, yet based on a common framework of standardized procedures and edits applicable to all countries. The bulk of the data processing was implemented using SAS version 9.2. All data processing was run in Unicode mode, thereby preserving all national strings in free text entry variables. Programs for initiating and controlling SAS or other processing programs were based on generalized processing systems used across all IEA and third-party surveys managed by the IEA DPC. All processing systems were set up so that the different steps, from import to exporting data products, could be run again to include and reflect all changes and edits. Missing values were represented using SAS standard (".") or special missing values (".A"-".Z").

13.4.1 Data import and structural reorganization

The import and merge of data essentially followed the below sequence of steps. As a first step, data capture accuracy was checked using the submitted IEA DME database and recorded. As noted before, data capture accuracy was found to be satisfactory for all participating countries in the Main Study. Data from the double capture process were set aside and not processed further.

Next, each national database in the DME's native Microsoft SQL Server Compact format were loaded into a temporary Microsoft SQL Server 2008 R2 server database "as is," that is, without any transformations or changes. Using these SQL server data as the input, a SAS-based program read all data from the national databases, merged tables as necessary and checked for structural integrity and deviations from the international variable layout. This step produced four SAS formatted files.

Original national database tables were consecutively merged using PERSID to form a single flat file named PRG (for PIAAC Response General) encompassing all variables for a single case from the following source datasets (see Section 13.2.2 above for details):

- SDIF – Sample Design International File
- BQR – BQ results and workflow
- BQC – Coded responses
- CBR – Computer-based exercise results
- PCM1 – Paper core booklet results
- PLM1 – Paper literacy booklet results
- PNM1 – Paper numeracy booklet results
- RCM1 – Paper reading components results

Cases or respondents present in neither the SDIF nor BQR dataset were dropped at this stage. The PRG file was inclusive of all national adaptations and extensions introduced by countries.

The dataset in the national database holding reliability scoring/capture data were merged using PERSID to form a flat file named PRR (for PIAAC Response Reliability), encompassing all variables for a single case from the following source datasets:

- PCR1 – Paper core booklet results

- PLR1 – Paper literacy booklet results
- PNR1 – Paper numeracy booklet results
- RCR1 – Paper reading components capture results

The IEA DME dataset holding reliability scoring/capture data were merged using PERSID to form a flat file named PAG (for PIAAC Anchor General) encompassing all variables for a single case from the following source datasets:

- ACM1 – Anchor core booklet results
- ALM1 – Anchor literacy booklet results
- ANM1 – Anchor numeracy booklet results

The IEA DME dataset holding reliability scoring/capture data were merged using PERSID to form a flat file named PAR (for PIAAC Anchor Reliability) encompassing all variables for a single case from the following source datasets:

- ACR1 – Anchor core booklet results
- ALR1 – Anchor literacy booklet results
- ANR1 – Anchor numeracy booklet results

One additional file named PRL (for PIAAC Response Log) was produced from the information parsed in the national database's BQL dataset. This file was not subject to cleaning or editing as it mainly included timing information for validation purposes.

For each component and source table, a flag was created regarding whether data relating to the case existed in the source dataset with only missing values, with some valid values, or with a complete set of values.

13.4.2 Structure check and recoding of national adaptations

The structure check stage performed several checks that related to file and variable structure integrity. It checked for changes in international variable definitions, availability of mandatory variables applicable to all sample designs and contexts, as well as the validity of national variable definitions with respect to naming conventions and in light of agreed-upon adaptations in the BQAS. All original missing values in national databases were programmatically mapped to SAS missing values on import. At this stage, validation checks for all numerical variables ran and ascertained that no unconfirmed out-of-range values remained in the data. NDMs received standardized reports on any flagged inconsistencies for either confirmation or resolution.

Questions in the PIAAC master BQ were designed to have the same meaning for respondents in all participating countries irrespective of differences in language and culture. However, two sets of adaptations or extensions had to be applied by countries in the process of translation/adaptation: i) mandatory adaptations in the case of ISCED levels, country name placeholders, and the like, and ii) idiosyncratic adaptations and extensions that reflected national research interest or were used to align questions with other data collections. These national adaptations and extensions had to be processed

along with data for not adapted questions. While national extensions were processed, returned to countries for their own use, and also referenced in the psychometric analysis, data collected from national adapted questions had to be harmonized by means of recoding for it to be internationally comparable.

For this purpose, the IEA DPC processed and reviewed all final BQAS and created Excel documents that only included national extensions and those structurally adapted (e.g., added response options). The result from this process was documentation of country adaptations requiring attention during the international data processing phase by recoding national responses back to the international response schemes and variables where needed. Additionally, it was recorded for each adaptation whether a recoding was needed and, if yes, whether the IEA DPC or the country was responsible for implementing it. These "reverse" BQAS sheets were discussed with the concerned country and finally reviewed by ROA, the Consortium partner initially responsible for reviewing and approving national adaptations.

The recodings due to national adaptations were applied by default during the course of processing countries' data according to agreement found in the process described above. National variables affected by these adaptations retained their original values through the whole cleaning process and provided to countries unchanged after data processing. Many countries, though, decided to perform several, if not all, necessary recodings themselves prior to data submission. This was supported and approved by the Consortium in cases where countries also provided the constituent national variables referenced in the recodings. In some cases, countries used complex adaptations in the BQ, and this in turn resulted in very complex recodings that had to be harmonized under country responsibility and local validation and verification. In some other cases, countries were responsible for recoding data prior to submission given confidentiality reasons, that is, situations where countries were not able to release certain variables to the Consortium due to national legislation.

The Consortium reviewed the appropriateness of all applied recodings with respect to international comparability of data by means of cross tabulations using a single or multiple source and target variables. This also applied to cases where countries applied recodings prior to data submission and the source national variables were provided to the Consortium. For recodings where the original national variables were not disclosed to the Consortium, no detailed validation of the recoding process was possible and the Consortium informed the concerned countries that any error as a result of these recodings was entirely the responsibility of the country. Nonetheless, the Consortium applied coarse and technical plausibility checks of the resulting data. Countries were provided with the same frequency distributions in the resulting data and were asked to check and verify them. Table 13.2 provides an overview where recodings were applied and whether the national variables referenced were available to the Consortium. Following from the process descriptions provided by those countries which applied recodings prior to submission, the Consortium was not aware of any indication that particular recodings applied by countries were invalid or flawed in other ways. However, the volume of national questions and variables (in excess of 1,200 variables across the 24 countries), the complexity of some adaptations and extensions, a somewhat different response process, and differential missing data in cases where multiple questions were referenced to yield an international value made it quite likely that some minor errors remained undetected in the data.

Table 13.2: Responsibility for, and time of, mapping national to international variables

Country name	All mappings applied by country prior to submission	Some mappings applied by country prior to submission	All mappings applied by Consortium after submission	Consortium had no access to some or all original national data
Australia	X			X
Austria		X		
Canada	X			
Cyprus[2]			X	
Czech Republic	X			
Denmark			X	
England/N. Ireland (UK)	X			
Estonia		X		
Finland	X			
Flanders (Belgium)			X	
France			X	
Germany		X		
Ireland			X	
Italy	X			
Japan			X	
Korea			X	
Netherlands	X			
Norway	X			
Poland	X			
Russian Federation[3]			X	
Slovak Republic			X	
Spain			X	
Sweden	X			
United States			X	

13.4.3 Data cleaning process, systems, communication and reports

For the PIAAC Main Study, a comprehensive set of checks was implemented that allowed for a broad as well as deep inspection and cleaning of data files. As stated initially, this process of cleaning involved the Consortium partners directly involved in the database building, the OECD as the primary data consumer at the international level, and last but not least the NDMs and NPMs in each country.

[2] Please refer to notes A and B regarding Cyprus in the *Note to Readers* section of this report.

[3] Please refer to the note regarding the Russian Federation in the *Note to Readers* section of this report.

As part of the data cleaning process, records and variables were checked for consistency, that is, that no duplicate IDs existed, no unaccounted for wild codes existed, and the expected data pattern (given for example a case's booklet assignment, disposition codes or ICT core pass status) matched the observed data patterns. Additional checks focused on the consistency of records and variables within a dataset, the linkage of records between datasets, as well as repeating soft validation checks already run during the interview. Any flagged issues had to be reviewed, verified, resolved or, where this was not possible, at least commented on by countries. Extensive and detailed communication between the IEA DPC and any participating country on data inconsistencies and their resolution took place and detailed reports were provided to NPMs and their NDMs on any such issues, and they were asked for confirmation or advice.

The overall observed quality of the submitted data was usually very good. There were no substantial structural errors in databases and almost all cases matched between data sources. In total, only about a dozen or so cases out of more than 150,000 had to be removed or corrected because they were out of scope – for example, if they included both computer- and paper-based data where only one of the two was expected given the respondent's trajectory. The high degree of internal consistency of the data can probably be attributed to three main factors: i) the fact that the PIAAC assessment was highly standardized and computer-controlled and, technical problems aside, provided no possibility to follow an incorrect path, ii) the use of strict ID validation in all components of the survey, and iii) the diligent work of NDMs in identifying the few mismatching cases and allocating data as appropriate. Where data were not matching the expected design, narrative reports from countries indicated that this was due to interviewers not following the intended workflow. For example, some interviewers administered paper booklets in instances where there was a technical problem with the CBA portion of the assessment. Data values for components not applicable to a respondent were, after careful inspection, reset to their respective missing codes.

Other potential issues were mostly related to incidental, variable-level errors. These were too diverse and too sparse to be reported here in any detail. Recurring issues across countries included, but were not limited, to:

- incorrect or inconsistent disposition code assignment, often in cases of technical problems;
- incomplete data for anchor booklets used for cross-country reliability analysis;
- incomplete or incorrect mapping of national adaptations where this was the country's responsibility;
- missing indication of suppressed variables at the time of data submission;
- discrepancies between age and gender as recorded in sampling frames, collected via screeners, entered or loaded during case initialization, or reported by respondents in the field;
- incomplete loading of sample design data for noninterviews, that is, individuals who refused to take the interview, language-related nonresponse, or absences;
- incorrect loading of other sample design data given a country's plans for nonresponse adjustment and raking; and

- incorrect, incomplete or unreliable coding – for example, concurrent mapping and coding of country of birth responses, agreement of occupational distributions with external data sources, or handling of occupation/industry codes as numerical.

In many cases, such issues could be resolved by reviewing the original instruments and reviewing registry information and/or feedback from field operation, scoring or data entry staff.

13.4.4 General edits and the iterative integration of derived variables

After the cleaning phase, the state of the data reflected the data collection as accurately as possible, and any individual or structural inconsistencies were removed to the extent possible and known to the Consortium. During this post-cleaning phase, general structural changes were made to the data for all countries based on the now cleaned original values.

Most importantly, the processing systems reimplemented the routing logic included in the international master BQ and assigned "valid skip" missing values to any variable that was not in the respondent's path. It should be noted that this was done in a comparable way for all countries and only considered original or recoded responses to any international BQ variables. "Valid skip" missing codes were generally only assigned for respondents who started the BQ. In the case of breakoff during the BQ, the "valid skip" recoding was only implemented up until the last known valid value. All subsequent values were coded as "not stated." Further, "valid skip" codes were carried forward to any coded variable (e.g., second language learned) if the referenced original variables were previously coded as "valid skip."

In a few cases, countries not only adapted questions but were given permission to also adapt the routing rules implemented in the international master BQ. This resulted in a few instances where too little or too much information was collected in comparison to a route that a "standard" respondent would have taken through the BQ. Excess data collected due to national routing was overwritten with "valid skip" codes in the process described above for reasons of international comparability because respondents affected were not supposed to have data observed according to the international routing rules. In cases where too little information was collected, and thus missing data were present yet not expected, there was usually no way to recover from this. Such data cells were coded as "not stated." The overall number of affected cases was very small (a maximum of 207 cases with excess data and 38 cases with missing data for a few variables were present in one country) but nevertheless shows the risk and possible impact of excessive national adaptations to already complex international collection instruments.

Further at this stage, "not reached" codes were assigned to cognitive assessment items in the paper path. For this, items with value 0 = "no response" were recoded to "not reached/not attempted" according to a generic algorithm that checked for "no response" values from the end of each item block (and individually for each item block) and assigned value "not reached" until a valid code was encountered. "Not reached" codes were also assigned to item responses in the computer-based path. These adjustments were done at ETS, delivered to the IEA DPC as a set of mergeable files with revised data, and integrated into the master databases on each run.

This processing phase was further used to derive or merge reporting variables, weights and scale scores. The process of deriving variables was highly iterative and depended on the progress of the weighting and analysis. The derivation and integration observed the necessary sequencing conditions. For example, scripts for the coarsening of variables had to be based on revised original and/or derived variables.

- Derivation of variables from sample design and case initialization data

    - A number of sample-design related variables were derived from sample design, case initialization and BQ information. With the exception of some special settings in some countries, the derivation of these variables was done according to standardized scripts that were consistently applied with each pass of the data for all countries. These sampling-related variables were independently computed by IEA DPC and Westat and compared as well as reconciled as necessary. The most important derived variables in this segment were:

        - Three final, combined disposition codes for the case initialization and BQ phase (DISP_CIBQ), for the main assessment (DISP_MAIN) and including reading components for those countries participating in the option (DISP_MAINWRC).
        - Resolved age (AGE_R) and gender (GENDER_R) taking into account frame information but giving precedence to observed data during the interview, further incorporating collected age and gender in the case of literacy-related nonresponse.
        - A completion flag (COMPLETEFLG) set according to technical standards definitions in relation to assessment components and/or key items.
        - A weighting flag (WEIGHTFLG) computed from the disposition codes and/or literacy-related nonresponse information.
        - An interim code (SCENARIO) derived according to a set of rules intended to identify cases earmarked for weighting yet with insufficient information or vice versa.

    - The key Consortium partners responsible for identifying valid cases reviewed the outcomes of the above assignment in regular online meetings and revised the weighting and completion flags as well as aggregate disposition codes in a small number of cases depending on whether sufficient information was available to assign a weight and/or analyze the cases.

- Integration of weighting and variance estimation variables

    - Once valid cases where flagged for weighting and analysis, weights and scale scores were merged to countries' data files as they became available. Weights were computed by either the concerned countries or Westat and were merged to the files.

- Derivation of variables from the BQ data

    - A vast amount of variables were derived from original responses to the BQ. These variables relate to a set of broad groups, namely the respondent's background, education/training, earnings and skill use.

        - The majority of these variables were computed automatically during each pass over the data. These were based on definitions provided by the OECD and other partners of the Consortium.
        - Derived earnings variables were directly derived in the case of detailed responses or imputed from broad categories, and merged to the files.

        - Skill use derived variables were based on IRT estimation procedures, computed, and merged to the files.

    - A set of coarsened variables was scripted at the IEA DPC to cater for countries' needs to protect the confidentiality of respondents' information in the database. For these variables (suffix "_C"), one of three types of coarsening was applied: i) top coding, ii) categorization, or iii) collapsing of existing categories into a smaller set.

    - Finally, a set of "trend" variables was derived by ETS and provided to the IEA DPC as mergeable files (suffix "_T"). These trend variables relate to variables collected in the same or similar way as the ALL and IALS surveys; PIAAC variables were recoded to match the metric or coding schemes used in ALL and IALS in order to be comparable across surveys.

    - A small number of the derived BQ, trend and coarsened variables were computed under the responsibility of countries because the Consortium was not given access to the full source information required for the derivation. These variables were provided as mergeable files, validated and merged at the IEA DPC.

- Derivation of variables from the actual responses to the reading components items

    - At the time of data collection, three different types of response value schemes were used on the response capture sheets for print vocabulary, sentence processing and passage comprehension. During the data processing a response key was programmatically applied and used to assign actual responses (variables ending in "A") to scored responses (ending in "S") for all reading component items by mapping the correct distractor to code 1 = "correct" and other distractors to 7 = "incorrect."

- Derivation of variables from problem-solving unit responses

    - The PIAAC CBA system stored rich auxiliary information that provided indicators of respondents' actions during the cognitive assessment. At the time of collection, a large number of aggregate variables and interim scores were exported and processed. Following the data collection, "total correct scores" were derived and integrated into the master databases.

- Derivation of scale scores

    - PIAAC cognitive item responses were calibrated, analyzed and scaled. This process resulted in a set of 10 plausible values for each domain (literacy, numeracy and problem solving) plus one additional variable indicating the availability of plausible values for a particular respondent given the design and path.

13.4.5 Production of the preliminary national and international databases

The data finalization phase transitioned data from the internal IEA DPC processing systems to data products ready for further use by the Consortium, the OECD or the participating countries. The final processing phase further repeated many of the checks implemented at earlier stages to ensure that automated or incidental data editing did not introduce any inconsistencies, for example out-of-range codes, into the data. In addition, a set of additional checks was conducted that ensured data integrity after all cleaning steps had been run through and before export to the different final formats took place.

For example, checks ensured that the variable widths and types in the codebooks were defined wide enough to actually hold the data in the national master database.

At this stage, a single international codebook was used to describe and document the data. Widening conversions were applied consistently across all countries in case one or more countries extended the width of a variable in their national database's codebook (e.g., with respect to currency values). The final international master database held 1,712 international variables for each participating country. Codebook information for nationally adapted or extended variables was taken from the national databases originally submitted by countries.

In all, the 33 datasets present in the IEA DME software and database at the time of data capture were processed and eventually resulted in the following six export file types, each produced in both SPSS as well as SAS format:

- *PRGxxxMS.sav/.sas7bdat*: The main analysis file with all originally collected and derived variables, international as well as national.

- *PRRxxxMS.sav/.sas7bdat*: An auxiliary file holding reliability scores for the core and literacy/numeracy booklets as well as responses captured for reading components. The PRR file includes a true subset of the variables in PRG but with values from the reliability scoring process.

- *PAGxxxMS.sav/.sas7bdat*: A flat file with scores from the cross-country reliability study, main scoring.

- *PARxxxMS.sav/.sas7bdat*: A flat file with scores from the cross-country reliability study, reliability scoring.

- *PSDxxxMS.sav/.sas7bdat*: A flat file encompassing sample design variables. This file included a true subset of variables as well as all records from the PRG file and was mainly used by Westat or countries in the process of weighting.

- *PRLxxxMS.sav/.sas7bdat*: A flat file for the CAPI event log.

Data files were exported separately by country. This allowed for the provision of files to the Consortium as well as to individual countries on a rolling basis. The placeholder "xxx" used in the file names above corresponds to operational identifiers based on ISO 3166.

SPSS data files were standard, Windows-based *.sav* files and encoded in Unicode (UTF-8). SPSS data files included full dictionary information from the applicable metadata maintained in the codebooks including variable types and formats, variable labels, value labels (including any labels for missing values), missing value definitions and variable measurement levels. SAS-formatted files were standard, compressed *.sas7bdat* data files for Windows environments and encoded in Unicode (UTF-8). Variable types, widths, decimals and labels were assigned to all variables according to the labels defined in the metadata. SAS does not provide for a way to permanently store value labels on the file. Therefore, each file in SAS format was accompanied by an equivalently named *.sas* file which could be used to assign formats (value labels) to working files. Missing values represented as SAS missing values were programmatically mapped to either numerical missing values in the case of SPSS or a reduced set of special missing values in the case of SAS.

To allow for the export of data products for the various data users and stakeholder, data files could be produced according to three export profiles:

- Profile 1 for international analysis, weighting and archiving
    - This export profile retained all international and national variables originally submitted or derived on the data file.
    - These full information files were made available only to the Consortium partners who required access to the data as well as the OECD. These files were kept strictly confidential and were not shared beyond the group of organizations and individuals involved in the analysis and weighting.
    - This profile included all records originally submitted by a country.
- Profile 2 for the release of national databases to countries
    - This export profile maintained the vast majority of international and national variables. It excluded a small set of internal, interim or redundant variables produced as part of the scaling and analysis process and only relevant for the purpose or archiving.
    - This profile was provided only to the concerned countries.
    - This profile included all records originally submitted by a country.
- Profile 3 for public use
    - This export profile, by default, maintained all international variables approved for release by the BPC as part of the public-use file.
    - Any and all national variables were dropped.
    - For this profile, all international variables earmarked for suppression by a country were blanked (i.e., set to the appropriate missing value for all cases).
    - This profile only included records with the PIAAC in-sample flag (INPIAAC) equal to 1.

Each data exported was uniquely identified by an export data and an export version variable in the data files. These two variables allowed analysts to compare the data version underpinning the current work. In terms of data flow, the IEA DPC, as a subcontractor, provided all data products exclusively to ETS followed by quality control there. Subsequent data releases to other Consortium partners, the OECD, and participating countries were managed by ETS. An alternative data exchange protocol was used in the case of Australia to account for special regulations pertaining data security.

13.5 Data review and finalization

Following the initial data cleaning process described above, an iterative process of data review and correction began within the Consortium and later involved the participating countries as well as the OECD Secretariat. Integrating, verifying and, where necessary, updating the above stated groups of variables as well as the implementation of countries' feedback on their national databases all occurred

under a tight timeline and included multiple data sendouts and review rounds. The general principle followed was that data collected, cleaned or derived by one party (e.g., the participating country or a Consortium partner) was reviewed by at least one other partner as well as the concerned country. Building and verifying the national and international databases was a collaborative process involving the specific expertise, knowledge and experience of the surveys designers, stakeholders and national project teams.

The list below presents the key data products and times in the process of reviewing and finalizing national and international databases for the majority of countries.

- Preliminary international database (July 2012)
    - The IEA DPC provided a preliminary international database including data from 20 countries to the Consortium for internal review and to ensure that all processes and procedures for analyzing Main Study data were in place.
    - This database included originally submitted, initially cleaned, and where applicable, perturbed data. Further, this database contained the design weights provided by countries.
    - A series of country-by-country updates to the preliminary international database and initial versions for two late-submitting countries were issued between July and November 2012 in parallel to data cleaning and initial weighting efforts.
- First international database (December 2012)
    - The IEA DPC provided a first international database including 22 countries' data for analysis to the Consortium.
    - This database included weights, replicate weights, and a basic set of scripted derived variables.
- Second international database (January 2013)
    - The Consortium completed the initial data analysis and generated the majority of derived variables and plausible values for 22 countries.
    - This database was shared with the OECD in order to prepare international reporting.
- Release of preliminary national databases (January 2013)
    - At the end of January 2013, the Consortium released cleaned, weighted and analyzed national data to countries for review and approval. The microdata files were accompanied by summary data tables.
- Review of preliminary national databases (February to June 2013)
    - This period, originally scheduled until the end of May 2013, was intended for countries to review records and variables included in their cleaned, weighted and analyzed national databases.

- As a result of countries' review of their respective national database, the Consortium's own observations, and the initial reporting work at the OECD, the Consortium and the OECD agreed on data changes and error corrections to be applied commonly for all or just individual countries in order to improve the validity and quality of the data. Such changes related to:

    - repeated or corrected coding of occupational information with initially insufficient reliability or agreement with external data sources (e.g., labor force surveys);
    - minor corrections to the mapping of national educational attainment variables to international ISCED levels in light of discrepancies with other data collections (such as the OECD's Education at a Glance) in some countries;
    - correction of outliers in earning variables for some countries;
    - assignment of valid skip codes for skill use, earnings, reading components outcome variables, and a few other variables given that the original variables were not applicable to the entire survey population;
    - corrections to country-specific or general scripted derived variables; and
    - numerous label changes to better describe and reflect the content and scope of variables.

- The correction of data in some cases required the reanalysis of the cognitive data; resulting updates to scale scores and other measures were reflected in the concerned national databases.

- Countries were further asked to identify variables for suppressions and coarsening in any public-use data file releases on the basis of a preliminary list of variables earmarked for inclusion in such files. By mid-February, countries provided the Consortium with lists of variables coarsened. From this, the OECD selected a set of coarsened variables to be included for all countries. By the end of March, countries provided the Consortium with a list of variables to be suppressed from the now complete set of variables intended for the public-use data.

- Release of restricted international database through the Data Explorer (April 2013)

    - Following the initial batch of corrections and updates, the IEA DPC finalized a third international database including 22 countries.

    - The third international database was shared with the OECD in order to continue the preparation of international reporting.

    - The database was exposed to participating countries via an initial, secure version of the PIAAC Data Explorer at the time of a training workshop delivered to NPMs. Access to this version of the Data Explorer was restricted to countries, the OECD and the Consortium partners. Countries were identified by codes rather than clear text names.

- Release of draft national public-use files (June 2013)

- Following the earlier corrections, the IEA DPC produced a draft of the public-use file for each country that reflected the respective national suppressions.

- Countries were asked to verify the contents and accurate suppression.

- Finalization of the international database (June/July 2013)

  - The international database for 22 countries was finalized at the IEA DPC. The Consortium applied last-minute tweaks to variable and value labels, as well as to missing value schemes.

  - This database was shared with the OECD in order to produce an updated draft of the international report.

- Release of unrestricted Data Explorer and public-use data to all countries (July 2013)

  - Countries received unrestricted but embargoed access to the PIAAC Data Explorer, that is, with country names unmasked, and the public-use data files for all other countries in order to advance and finalize work on national reports.

  - In August, data for France was additionally released to countries.

- Release of the international report and a public-use international database (October 2013).

  - The public-use version of the international database was scheduled to be released in parallel to the initial international report for PIAAC.